2021

Change detection and landscape similarity comparison using computer vision methods

Karim Malik Wilfrid Laurier University
mali3080@mylaurier.ca

Change detection and landscape similarity comparison using computer vision methods

by

Karim Malik

MSc., King Fahd University of Petroleum and Minerals, Saudi Arabi, 2015

BSc., University of Cape Coast, Ghana, 2006

A Dissertation Submitted in Partial Fulfillment for the Requirements of the Degree of

DOCTOR OF PHILOSOPHY

in the Department of Geography and Environmental Studies

Supervisory Committee

Change detection and landscape similarity comparison using computer vision methods

by

Karim Malik

MSc., King Fahd University of Petroleum and Minerals, Saudi Arabi, 2015

BSc., University of Cape Coast, Ghana, 2006

Supervisory Committee

Dr. Colin Robertson (Supervisor)
(Department of Geography & Environmental Studies)

Dr. Steven A. Roberts (Departmental Member)
(Department of Geography and Environmental Studies)

Dr. Tarmo K. Remmel (Additional Member)

(Faculty of Environmental & Urban Change, York University)

Dr. Jed Long (Additional Member)

(Department of Geography & Environment Western University)

Abstract

Supervisory Committee

Dr. Colin Robertson (Supervisor)

(Department of Geography & Environmental Studies)

Dr. Steven A. Roberts (Departmental Member)

(Department of Geography & Environmental Studies)

Dr. Tarmo K. Remmel (Additional Member)

(Faculty of Environmental & Urban Change, York University)

Dr. Jed Long (Additional Member)

(Department of Geography & Environment, Western University)

ABSTRACT

Human-induced disturbances of terrestrial and aquatic ecosystems continue at alarming rates. With the advent of both raw sensor and analysis-ready datasets, the need to monitor ecosystem disturbances is now more imperative than ever; yet the task is becoming increasingly complex with increasing sources and varieties of earth observation data. In this research, computer vision methods and tools are interrogated to understand their capability for comparing spatial patterns. A critical survey of literature provides evidence that computer vision methods are relatively

robust to scale and highlights issues involved in parameterization of computer vision models for characterizing significant pattern information in a geographic context. Utilizing two widely used pattern indices to compare spatial patterns in simulated and real-world datasets revealed their potential to detect subtle changes in spatial patterns which would not otherwise be feasible using traditional pixel-level techniques. A texture-based CNN model was developed to extract spatially relevant information for landscape similarity comparison; the CNN feature maps proved to be effective in distinguishing agriculture landscapes from other landscape types (e.g., forest and mountainous landscapes). For real-world human disturbance monitoring, a U-Net CNN was developed and compared with a random forest model. Both modeling frameworks exhibit promising potential to map placer mining disturbance; however, random forests proved simple to train and deploy for placer mapping, while the U-Net may be used to augment RF as it is capable of reducing misclassification errors and will benefit from increasing availability of detailed training data.

Table of Contents

List of Tables

List of Figures

Acknowledgements

I want to first express my deepest appreciation for Dr. Colin Robertson's support throughout the four years of my PhD. The knowledge and skills he imparted to me, especially via his initiatives such as allowing me to be the technical lead in important projects has enabled me to acquire invaluable analytical skills. I further express my profound gratitude to Dr. Colin Robertson's for the mentorship as well as financial support he provided me as his first PhD student; my PhD would not have been possible without these essential elements.

I want to also explicitly thank Dr. Steven Roberts. Dr. Steven had not only been supportive in rendering constructive critique and alternative analytical approaches, but as part of the Spatiallab team he has been very welcoming and is always willing to respond to calls, especially when Dr. Colin is on not immediately available.

Dr. Jed Long and Dr. Tarmo K. Remmel, both members of my supervisory committee, have play instrumental roles throughout the trajectory of my PhD. I have learned how to clearly articulate the content of my research and writing to the wider audience in geography and related disciplines their ideas, critiques, and suggestions.

I am thankful to Wilfrid Laurier University for making awarding me International Doctoral Scholarship and as well as TAship from which I had the opportunity to develop and improve my pedagogical approach by teaching undergraduate students in diverse courses.

I would also like to thank the Spatiallab Team, both alumni and present members for the company, social support and encouragement I received while pursuing the PhD. Clara Greig, Lauren Yee, Nick Wilson, George Heath and Hyden Lawrence were welcoming members of the Spatiallab I first met while struggling to acclimatize to the Winter weather. Dr. Chiranjib (Post-Doc), Majid Hojati and Ben Bondaruk are also helpful members of the Spatiallab whose company I enjoyed.

Finally, I want to express my profound gratitude to my family, my Wife and Kids for all the daily support I received. Without their continuous cheers and encouragement, it would have been a daunting challenge to complete a research-intensive PhD.

Karim Malik
Wilfrid Laurier University
August, 2021

Co-Authorship Statement

Chapter two through to chapter 5 were co-authored and the contribution of each author, including the doctoral candidate – Karim Malik, is outlined below:

Chapter 2: Karim Malik designed and performed the literature review. Colin Robertson, Steven A. Roberts, Tarmo K. Remmel and Jed Long provided advice and comments on the manuscript content.
Article is under review in *International Journal of Geographical Information Science.*

Chapter 3: Karim Malik designed the research, performed data analysis and prepared the manuscript. Colin Robertson assisted in preparing the manuscript through comments and edits.
Article is published in *Geographical Analysis.*

Chapter 4: Karim Malik designed the research, performed modelling and data analysis, and prepared the manuscript. Colin Robertson provided assistance during modelling phase and helped with comments, edits and advice on the overall manuscript content.
Article is published in *Remote Sensing*

Chapter 5: Karim Malik designed the research, performed modelling and data analysis, and prepared the manuscript. Colin Robertson provided advice on modelling design as well as helped with comments and edits to prepare the manuscript. Douglas Braun aided in the manuscript

preparation by providing comments and edits. Clara Greig assisted in data preparation and editing the manuscript.

Article is published in *International Journal of Applied Earth Observations and Geoinformation*.

Chapter 1: Introduction to change detection and landscape similarity comparison

1.1 Introduction

Ecosystems and environmental resources are continuously impacted at local and global scales by human-driven perturbations. The direct effects of anthropogenic processes such as urbanization and resource extraction are having direct and lasting effects on ecosystem form and function (Jenerette and Potere, 2010; Hamilton and Friess, 2018); whereas indirect effects of human activities are changing the trajectory of climate-driven processes leading to profound shifts in sensitive ecosystems around the globe (Townshend et al., 2012; Wulder et al., 2012; Greig et al., 2018). With growing and universal calls for sustainable development and climate change mitigation and adaptation (Neil Adger et al., 2005), the need for in-depth and up-to-date information on processes driving land surface changes is becoming more imperative (Mairota et al., 2013). Environmental data at varying spatial, temporal and thematic resolutions are increasingly available and being deployed within new land and resource monitoring systems. This data explosion originates from a multitude of sources: advances in airborne and space-borne sensors, Internet-driven data access and redistribution, the growth of professional and amateur drones, and several other commercial sources. This increased volume of data provides exceptional opportunity to understand how Earth systems are changing at a variety of scales (Wulder et al., 2018).

The quest to quantify and compare patterns expressed in spatial data has been a long-standing objective in quantitative geographical research (Csillag and Boots, 2005; Boots and Csillag,

2006; Foody, 2007). Given the recent increases in data acquisition technology and publicly available sensor data (e.g., Landsat and Sentinel datasets), there is renewed interest towards understanding multitemporal and/or multi-location datasets in the context of detecting and characterizing change across time and space (Zhu and Woodcock, 2014a). Consequently, there has been increasing interest in exploring tools capable of extracting and comparing spatial-temporal structures from Earth observation data to enable us characterize and quantify process-driven landscape change (Townshend et al., 2012; Wulder et al., 2012). Understanding the links between observed spatial patterns and unseen stochastic spatial processes remains the overarching goal of spatial analysis (Fotheringham and Rogerson, 2008; Wagner and Fortin, 2005). However, while known processes may create defined patterns, different processes may result in similar patterns. Moreover, patterns may act on underlying processes, and create complex interactions between pattern and process (Turner, 1989a). Despite these challenges, spatial pattern analysis remains a frequent approach to uncovering the link between patterns and processes at various scales (McIntire and Fajardo, 2009; Szilassi et al., 2017). For example, landscape pattern comparison may be undertaken either to compare the same landscape over time in order to identify and characterize change induced by land change processes (Coppin et al., 2004; Frate et al., 2014), or to compare different landscapes at the same (or similar) time to quantify similarities in spatial processes (e.g., deforestation) across space (Niesterowicz and Stepinski, 2016).

Change detection has become a wide-spread application of spatial pattern comparison, in the context of environmental resource monitoring (Coppin et al., 2004; Kennedy et al., 2009). Researchers in GIScience and remote sensing have proposed a variety of methods aimed at detecting and isolating significant changes in patterns from less significant changes. Pixel-based

and object-based image analysis are frequently utilized to quantify and characterize spatiotemporal changes, especially in remotely sensed data. In terms of analyzing spatial patterns, pixels provide no valuable contextual information (Hussain et al., 2013). However, following pixel-based operations (i.e., post-classification), aggregating and comparing blocks of pixels belonging to identical landcover types can provide insight into spatial pattern change in multitemporal imagery (Mas, 1999; Hussain et al., 2013; Tewkesbury et al., 2015). With the availability of high and medium resolution satellite datasets, object-based image analysis is becoming increasingly common (Gartner et al., 2008). Compared to pixel-based techniques, object-based methods enable the extraction of contextual features required to develop robust environmental monitoring tools (Blaschke, 2010; Chen et al., 2012). For instance, Im et al. (2008) found that object-based methods that incorporate object correlation and neighborhood correlation features showed superior performance against pixel-based methods and object-based methods without contextual information. Given the availability of high spatial and temporal resolution datasets, opportunities to monitor ecosystem disturbances are now greater than ever; yet methods and tools required to achieve this ultimate goal remain fractured across diverse domains, making the exploration for novel and generic tools for extracting pattern information a pressing research need. Figure 1.1 depicts change detection in which a given landscape is compared over time (e.g., between the years 2010 – 2020) to quantify the loss and/or gain of certain landcover types. With landscape similarity assessment as illustrated in Figure 1.1, the focus centers on comparison and/or search for similar landscapes over variety of geographical locations of interest (i.e., $l_1 \ldots l_n$ locations) in order to understand similarities in processes driving pattern formation and/or change over space.

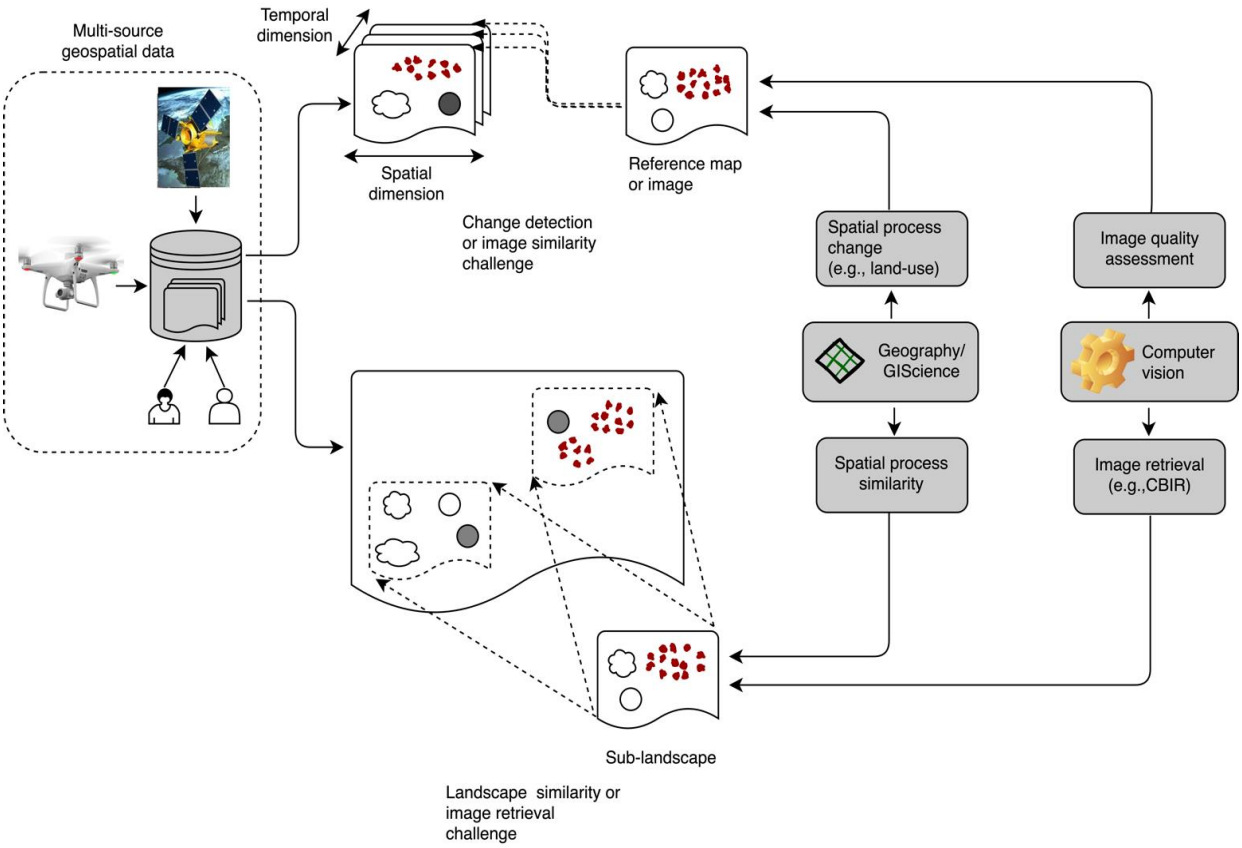


Figure 1.1 Change detection versus landscape similarity comparison.

With the advent of so called “analysis ready data” (Dwyer et al., 2018), and huge archives of raw geospatial sensor data, landscape similarity search is becoming a task of increasing importance, and complexity, especially in the context of monitoring landscape processes across space and time. Questions that often need to be answered through landscape similarity analysis include, but are not limited to: where are degraded landscapes prevalent? Are the underling patterns of land degradation common amongst candidate sites? Spatially explicit techniques are crucial to answering these questions. A possible approach is to derive landscape pattern metrics and compare one or several metrics between landscapes of interest (Frate et al., 2013; Remmel

and Fortin, 2013). For instance, the Euclidean distance between vectors of pattern metrics representing two landscapes could be computed to help answer questions pertaining to landscape similarity (Niesterowicz and Stepinski, 2016). Also, frequency distributions can be derived from landscape configuration components to provide a generic model of landscape structure for comparing similarity and differences between landscapes (Remmel, 2020).

Categorical maps and continuous-valued maps present two fundamental problems in pattern comparison research. While categorical maps assume that patterns are homogeneous with abrupt transitions at adjacent areas, continuous-valued data represents spatially continuous phenomena (Gustafson, 1998), and this dictates the range of tools required to extract spatial information for change detection and/or similarity analysis task. Although conventional methods for quantifying change, and detecting similarity have been used with some successes in categorical maps (Monserud and Leemans, 1992; Visser and Nijs, 2006), the problem of detecting significant patterns while disregarding spurious changes due to artifacts such as data collection methods or data errors persist to date, and requires careful consideration of the processes and patterns of interest (Kennedy et al., 2009; Long and Robertson, 2018). For example, using the Euclidean distance between landscape metrics as measure of similarity is prone to errors resulting from underlying data artifacts. Additionally, the technique requires metric normalization as well as weighting the compositional and configurational aspect of patterns, which often lacks any unified approach (Niesterowicz and Stepinski, 2016). To confound pattern analysis, the link between patterns and processes may change with scale; hence, the scale at which processes operate to generate patterns may be coarser or finer than the scale at which patterns interact to affect spatial processes (McIntire and Fajardo, 2009). Therefore, tools that can potentially handle slight changes and shifts in scale would be crucial for characterizing and quantifying patterns.

There is increasing scope and need for pattern comparison as a key component in geographical change detection and landscape similarity assessment in environmental monitoring systems. Fortunately, however, these dual challenges are related to image analysis tasks in computer vision; typically, image compression algorithm assessment which is often formulated as an image quality assessment problem, and image retrieval (e.g., content-based image retrieval (CBIR)). As depicted in Figure 1.1, image degradation behaviour of an image compression algorithm bears some resemblance with underlying spatial processes that for example, may have resulted in observed landcover changes. Similarly, content-based image retrieval in which the task focuses on finding a set of images similar to a query image is analogous to landscape similarity comparison wherein the search for similar landscapes could potentially lead to identifying the underlying drivers or processes of pattern similarity. Drawing from computer vision applications in other disciplines, there is potential to adopt computer vision algorithms to compare geographic spatial patterns. For example, Structure from Motion algorithms have been widely adopted for drone image processing in environmental applications, for example being deployed in monitoring canopy height as a function of ecosystem disturbance intensity (Dandois and Ellis, 2010; Dandois and Ellis, 2013). A landscape similarity search algorithm proposed by Jasiewicz et al. (2014) effectively identifies landforms whose topographic features are similar. Recently, Nowosad and Stepinski (2021) proposed deriving an integrated co-occurrence matrix for multi-thematic categorical patterns in local landscapes and calculating dissimilarities between landcover types or landforms using Jensen-Shannon Divergence followed by clustering or segmentation to identify landscape types.

The need to evaluate image compression algorithms quality is a commonly encountered problem in computer vision, often requiring the deployment of structurally sensitive indices. To

this end, the structural similarity (SSIM) index (Wang et al., 2004), and the complex wavelet structural similarity (CWSSIM) index (Sampat et al., 2009), were developed to address the challenge of comparing patterns of deformation present in image-reference pairs. The SSIM detects differences between image pairs while ignoring spurious sources of variation attributable to illumination and contrast change, and have been proven to effectively detect subtle changes in spatially patterned (geographic) phenomena (Robertson et al., 2014; Jones et al., 2016).

Recently, the SSIM was applied to evaluate the ecological performance of a variety of hydrologic restoration methods (Wiederholt et al., 2019). Despite these successes and potential, structurally-sensitive metrics have found very limited application in geographical research, and in resource monitoring and decision-making workflows. Moreover, aside from image quality assessment in computer vision applications, the CWSSIM which is more robust to multi-scale pattern characteristics, is yet to be tested in geographical context. Until recently, conventional methods dominated change detection and landscape similarity assessment systems (Hernandez and Ramsey, 2013; Zhu and Woodcock, 2014; Tewkesbury et al., 2015). While such techniques have yielded some acceptable range of outcomes, tracking incremental and progressive changes (e.g., in continuous-valued data) requires spatially explicit tools that incorporate structure, texture and contextual information pertaining to underlying patterns and processes. Such tools are potential candidates for bridging the limitations inherent in traditional techniques.

The need to employ data-driven computer vision approaches to understand underlying data structures is likely to become inevitable (Reichstein et al., 2019). Analysis and integration of "big data" now accessible via a variety of geospatial databases, for instance, could help uncover a wealth of information about the Earth and resources not previously known (Miller and Goodchild, 2015; Reichstein et al., 2019). Artificial neural networks (ANN), for example, are

typical data-driven modelling approaches with promising potential for extracting significant pattern information, while discarding noise and related data artifacts (Basheer and Hajmeer, 2000). Some of the early demonstrations of ANN potential in environmental monitoring include: multitemporal land use change analysis (Dai and Khorram, 1999), and change detection along the Nile River delta (Carpenter et al., 2001). Notwithstanding these applications, remotely sensed data which frequently come in several to hundreds of bands of data distributed over large pixel arrays have the tendency to increase the parameters required to train ANN. To reduce model parameters as well as exploit spatial dependence in remote sensing data, convolutional neural networks (CNN) are endorsed by many researchers focusing on landcover mapping (Liu et al., 2018a). CNNs are typical spatially explicit and contextual-aware deep learning algorithms capable of filling the gaps in ANN models (Janowicz et al., 2020). Recent deployment of CNNs in a variety of tasks involving segmentation and classification confirms their superior performance (Buscombe and Ritchie, 2018). For example, forest change detection using CNN is demonstrated in Khan et al. (2017) research wherein the change detection challenge is cast as patch-based classification problem. Albert et al. (2017) employed deep features extracted from satellite imagery to compare urban neighborhood similarity across European cities, while Sylvain et al. (2019) and Boulent et al. (2019) demonstrate the potential of CNNs for plant disease identification and tree health monitoring. Given the data-hungry nature and training requirements to build deep-learning models, off-the-shelf pre-trained models are gaining increasing attention (Cao et al., 2019). For instance, a pre-trained CNN could be employed to spatially filter and classify photographs to aid in land cover and land use characterization (Tracewski et al., 2017), and thus providing valuable insight into processes driving patterns of landscape similarity. An emerging research domain which is promising and could potentially

improve the performance of data-driven frameworks is the utilization of features extracted from CNNs layers. CNN features are discriminative and hold potential for deployment in both change detection and landscape similarity assessment systems (El Amin et al., 2016).

As computer vision methods are gradually being adopted in geographical research, there are several lingering questions that need to be addressed to pave the way for a more wide-spread, adoption, refinement, and deployment in environmental monitoring systems. To that end, a fundamental question this dissertation attempts to answer is: Are computer vision tools sufficiently robust to detect and quantify "*significant patterns*", and to characterize change and/or similarity in patterns across temporal and spatial scales? Some of the more specific set of questions to be addressed include: In the context of characterizing change and/or similarity, how do emerging pattern comparison indices such as structural similarity (SSIM) index and complex wavelet structural similarity (CWSSIM) index perform against common classical metrics (e.g., the mean squared error)? How do we parameterize computer vision methods to quantify significant changes over varying spatial scales? How do we incorporate texture learning in computer vision models such as CNN models when used in a geographic application? And how can we exploit discriminative features extracted by deep-learning models to answer spatial and temporal questions? This dissertation aims to approach and address these questions with practical examples involving the application of computer vision methods for pattern comparison in simulation-based data (e.g., pattern simulated via Gaussian Markov Random fields), coarse-scale remotely sensed imagery (i.e., GlobeSnow snow water equivalent data), and fine-scale remotely sensed imagery (i.e., aerial photography, SPOT imagery).

1.2 Research objectives

Given the challenges and questions outlined above, this dissertation focuses on the following specific objectives:

(a) Evaluate the use of computer vision metrics for spatial pattern comparison and change detection.

(b) Develop a new change monitoring framework using texture-encoded features learned within a CNN setting.

(c) Apply new change monitoring methods in case studies exploring different spatial scale/pattern configurations.

A summary of how the research objectives will be organized into chapters and analysis is provided in Table 1.1

Table 1.1 How this dissertation approaches the objectives of the proposed research.

Chapter	Major issues addressed	Approach utilized
2	*Theoretical:* Understanding computer vision and scale issues with respect to model parameters	A review of scale issues and how CNN could be parameterized for GIScience and geographical research
3	*Methodological:* Emerging structurally sensitive metrics for pattern comparison	Structural similarity metrics are explored, highlighting their potentials to characterize patterns in continuous valued spatial data
4 & 5	*Application oriented:* Developing a deep-learning CNN for landscape similarity assessment	Texture learning incorporated to develop a CNN model whose feature maps are subsequently employed to map landscape similarity
	Developing a U-Net CNN for placer mining disturbance monitoring	A contextually aware U-Net model is developed and compared with Random Forest in the context placer disturbance monitoring

1.3 Overview of chapter contents

In Chapter 2, a thorough literature review is conducted to illuminate scale challenges which may be encountered when adopting computer vision methods in geography. More importantly, linkages between CNN model parameters and classical tools and concepts are highlighted. As CNNs are gaining increasing attention in both change detection and landscape/image retrieval workflows, it is essential to understand how such models could be parameterized to effectively discriminate patterns of interest, and to enable the extraction of discriminative features for environmental monitoring and decision-making systems. A review of CNNs and synthesis of scale terminologies provides valuable information for adoption and deployment of computer vision methods across multiple disciplines. The findings reveal that scale impacts both classical techniques and computer vision methods; but the latter is relatively robust to small-scale variations, though parameterization challenges remain.

Chapter 3 examines the use of structurally sensitive metrics to compare patterns in continuous-valued data. Comparing patterns which vary gradually and incrementally over space is a daunting challenge and requires indices that are insensitive to scale, contrast and illumination. To this end, Chapter 3 demonstrates the potential of two computer vision metrics – SSIM and CWSSIM to characterize and quantify differences between images simulated using Gaussian Markov Random fields. To further evaluate the performance of these metrics in real-world dataset, Snow Water Equivalent maps are used to test the metrics potential to capture patterns of snow variability over the Northwest Territories, Canada. The analysis in Chapter 3 point to the complementarity of SSIM and CWSSIM indices. While SSIM appears to be sensitive to local patterns of snow variability, CWSSIM tends to capture changes in snow distribution over large scales. One notable advantage of CWSSIM over SSIM, however, is its robustness to shift in

scale. Analysis of sub-images which are shifted versions of the original image off the coast of Gabriola Island, British Columbia, Canada shows that the CWSSIM yields high similarity scores, thus demonstrating the metric's geometric invariance property.

The development and deployment of deep-learning models, forms the focus in Chapters 4 and 5. In Chapter 4, fine-scale landscape similarity analysis is presented using the Aerial Imagery Data (AID) and Sentinel datasets. A texture encoded CNN model is developed to extract discriminative feature maps. The histogram of oriented gradients (HoG) vector is extracted from the resulting feature maps after PCA has been performed; next, the Earth Mover's distance algorithm is used to quantify landscape similarity. The finding provides insight into the potential to develop novel landscape similarity mapping metrics using CNN feature maps for.

In Chapter 5, a U-Net CNN model is developed and utilized to map placer mining disturbance in Yukon, Canada, using SPOT-6/7 imagery. To compare context-aware deep-learning models with pixel-based machine learning techniques, a Random Forest (RF) model is developed in parallel. The analysis provides evidence that both modelling frameworks can detect placer mining disturbances, and are potential tools for monitoring placer disturbance. While RF excels at classifying known sites – high Producer's accuracy, U-Net performs better at unknow sites – high User's accuracy. Thus, U-Net is likely to substantially mitigate misclassification errors but is more challenging to parameterize and train. RF on the other hand requires less computational time and could easily be trained and deployed to monitor placer disturbance. Chapter 6 presents and discusses the main contributions of this dissertation and concludes with limitations of the study as wells as future research directions. Figure1.2 depicts a schematic of the dissertation chapters and the key focus of each of the nested research areas. Sequentially, as can be seen in Figure 1.2, Chapter 2 provides a fundamental background on scale issue in both geography and

computer vision and how the computer vision discipline addresses scale problems. Next, structurally-sensitive computer vision metrics are presented in Chapter 3 to demonstrate spatial pattern comparison in continuous-valued data. Chapters 4 and 5 then present advanced computer vision methods – deep learning modelling to illustrate landscape similarity comparison and human-induced disturbance monitoring, respectively.

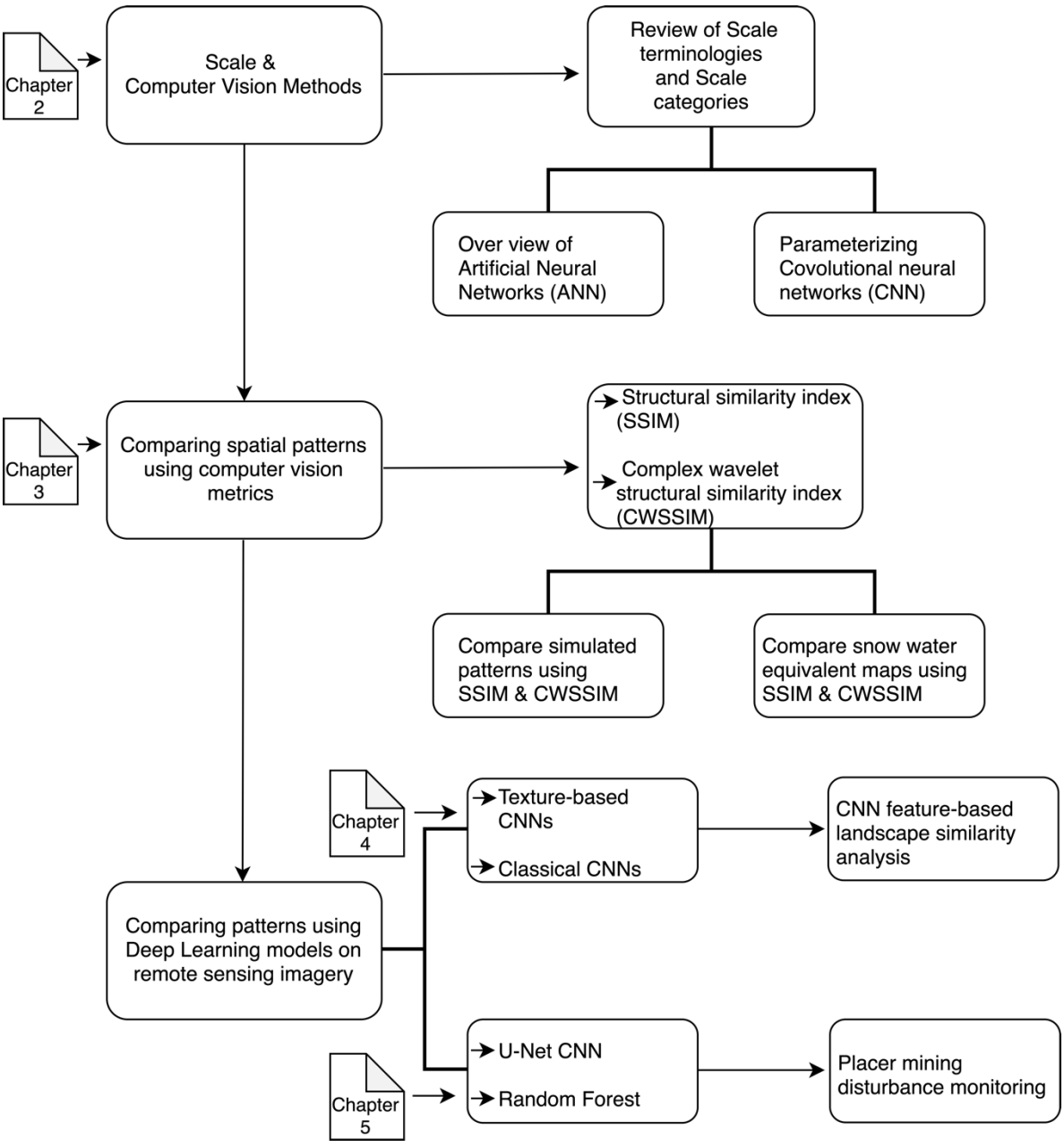


Figure 1.2 A flow chart of the approach to achieve research objectives.

References

Albert, A., Kaur, J., & Gonzalez, M. C. (2017). Using convolutional networks and satellite imagery to identify patterns in urban environments at a large scale. In *Proceedings of the 23rd ACM SIGKDD international conference on knowledge discovery and data mining*, 1357-1366.

Basheer, I. A., & Hajmeer, M. (2000). Artificial neural networks: Fundamentals, computing, design, and application. *Journal of Microbiological Methods*, *43*(1), 3–31.

Blaschke, T. (2010). Object based image analysis for remote sensing. *ISPRS Journal of Photogrammetry and Remote Sensing*, *65*(1), 2–16.

Boots, B., & Csillag, F. (2006). Categorical maps, comparisons, and confidence. *Journal of Geographical Systems*, *8*(2), 109–118.

Boulent, J., Foucher, S., Théau, J., & St-Charles, P. L. (2019). Convolutional Neural Networks for the Automatic Identification of Plant Diseases. *Frontiers in Plant Science*, *10*.

Buscombe, D., & Ritchie, A. C. (2018). Landscape classification with deep neural networks. *Geosciences (Switzerland)*, *8*(7), 1–23.

Cao, C., Dragicevic, S., & Li, S. (2019). Land-Use Change Detection with Convolutional Neural Network Methods. *Environments*, *6*(2), 25.

Carpenter, G. A., Gopal, S., Shock, B. M., & Woodcock, C. E. (2001). A Neural Network Method for Land Use Change Classification , with Application to the Nile River Delta. *BU/CNS Technical Report TR-2001-010, July 2015*.

Chen, G., Hay, G. J., Carvalho, L. M. T., & Wulder, M. A. (2012). Object-based change detection. *International Journal of Remote Sensing*, *33*(14), 4434–4457.

Coppin, P., Jonckheere, I., Nackaerts, K., Muys, B., & Lambin, E. (2004). Digital change detection methods in ecosystem monitoring: A review. *International Journal of Remote Sensing*, *25*(9), 1565–1596.

Coppin, Pol, Jonckheere, I., Nackaerts, K., & Muys, B. (2004). Digital change detection methods in ecosystem monitoring : a review. *International Journal of Remote Sensing*, *25*(9), 1565–1596.

Cressie, N. 1991. *Statistics for Spatial Data*. Vol. 2. Toronto: John Wiley & Sons, Inc.

Csillag, F., & Boots, B. (2005). Toward comparing maps as spatial processes. *Developments in Spatial Data Handling*, *24*(2), 641–652.

Dai, X., & Khorram, S. (1999). Remotely sensed change detection based on artificial neural networks. *Photogrammetric Engineering & Remote Sensing*, *65*(10), 1187–1194.

Dandois, J. P., & Ellis, E. C. (2010). Remote sensing of vegetation structure using computer vision. *Remote Sensing*, *2*(4), 1157–1176.

Dandois, J. P., & Ellis, E. C. (2013). High spatial resolution three-dimensional mapping of vegetation spectral dynamics using computer vision. *Remote Sensing of Environment*, *136*, 259–276.

Dwyer, J. L., Roy, D. P., Sauer, B., Jenkerson, C. B., Zhang, H. K., & Lymburner, L. (2018). Analysis ready data: Enabling analysis of the landsat archive. *Remote Sensing*, *10*(9), 1–19.

El Amin, A. M., Liu, Q., & Wang, Y. (2016). Convolutional neural network features based change detection in satellite images. *First International Workshop on Pattern Recognition. International Society for Optics and Photonics*, *10011*, 100110W.

Foody, G. M. (2007). Map comparison in GIS. *Progress in Physical Geography*, *31*(4), 439–445.

Frate, L., Carranza, M. L., & Bit, T. (2013). Quantifying Landscape-Scale Patterns of Temperate

Forests over Time by Means of Neutral Simulation Models. *ISPRS International Journal of Geo-Information*, *2*, 94–109.

Gartner, G., Meng, L., & Peterson, M. P. (2008). *Object-Based Image Analysis*. https://doi.org/10.1007/978-3-540-77058-9

Gustafson, E. J. (1998). Quantifying Landscape Spatial Pattern: What Is the State of the Art? *Ecosystems*, *1*, 143–156.

Hamilton, S. E., & Friess, D. A. (2018). Global carbon stocks and potential emissions due to mangrove deforestation from 2000 to 2012. *Nature Climate Change*, *8*(3), 240–244.

Hernandez, A. J., & Ramsey, R. D. (2013). A landscape similarity index: Multitemporal remote sensing to track changes in big sagebrush ecological sites. *Rangeland Ecology and Management*, *66*(1), 71–81.

Hussain, M., Chen, D., Cheng, A., Wei, H., & Stanley, D. (2013). Change detection from remotely sensed images: From pixel-based to object-based approaches. *ISPRS Journal of Photogrammetry and Remote Sensing*, *80*, 91–106.

Im, J., Jensen, J. R., & Tullis, J. A. (2008). Object-based change detection using correlation image analysis and image segmentation. *International Journal of Remote Sensing*, *29*(2), 399–423.

Janowicz, K., Gao, S., McKenzie, G., Hu, Y., & Bhaduri, B. (2020). GeoAI: spatially explicit artificial intelligence techniques for geographic knowledge discovery and beyond. *International Journal of Geographical Information Science*, *34*(4), 625–636.

Jasiewicz, J., Netzel, P., & Stepinski, T. F. (2014). Landscape similarity, retrieval, and machine mapping of physiographic units. *Geomorphology*, *221*, 104–112.

Jenerette, G. D., & Potere, D. (2010). Global analysis and simulation of land-use change

associated with urbanization. *Landscape Ecology*, *25*(5), 657–670.

Jones, E. L., Rendell, L., Pirotta, E., & Long, J. A. (2016). Novel application of a quantitative spatial comparison tool to species distribution data. *Ecological Indicators*, *70*, 67–76.

Kennedy, R. E., Townsend, P. A., Gross, J. E., Cohen, W. B., Bolstad, P., Wang, Y. Q., & Adams, P. (2009). Remote sensing change detection tools for natural resource managers: Understanding concepts and tradeoffs in the design of landscape monitoring projects. *Remote Sensing of Environment*, *113*(7), 1382–1396.

Khan, S. H., He, X., Porikli, F., & Bennamoun, M. (2017). Forest Change Detection in Incomplete Satellite Images with Deep Neural Networks. *IEEE Transactions on Geoscience and Remote Sensing*, *55*(9), 5407–5423.

Liu, T., Abd-elrahman, A., Morton, J., Wilhelm, V. L., Liu, T., Abd-elrahman, A., Morton, J., & Wilhelm, V. L. (2018a). Comparing fully convolutional networks , random forest , support vector machine, and patch-based deep convolutional neural networks for object- based wetland mapping using images from small unmanned aircraft system. *GIScience & Remote Sensing*, *55*(2), 243–264.

Mairota, P., Cafarelli, B., Boccaccio, L., Leronni, V., Labadessa, R., Kosmidou, V., & Nagendra, H. (2013). Using landscape structure to develop quantitative baselines for protected area monitoring. *Ecological Indicators*, *33*, 82–95. https://doi.org/10.1016/j.ecolind.2012.08.017

Mas, J. F. (1999). Monitoring land-cover changes : A comparison of change detection techniques. *International Journal of Remote Sensing*, *20*(1), 139–152.

McIntire, E. J. B., & Fajardo, A. (2009). Beyond description: The active and effective way to infer processes from spatial patterns. *Ecology*, *90*(1), 46–56.

Miller, H. J., & Goodchild, M. F. (2015). Data-driven geography. *GeoJournal*, *80*(4), 449–461.

Monserud, R. A., & Leemans, R. (1992). Comparing global vegetation maps with the Kappa statistic. *Ecological Modelling*, *62*(4), 275–293.

Neil Adger, W., Arnell, N. W., & Tompkins, E. L. (2005). Successful adaptation to climate change across scales. *Global Environmental Change*, *15*(2), 77–86.

Niesterowicz, J., & Stepinski, T. F. (2016). On using landscape metrics for landscape similarity search. *Ecological Indicators*, *64*, 20–30.

Nowosad, J., & Stepinski, T. F. (2021). Pattern-based identification and mapping of landscape types using multi-thematic data. *International Journal of Geographical Information Science*, *00*(00), 1–16.

Reichstein, M., Camps-Valls, G., Stevens, B., Jung, M., Denzler, J., Carvalhais, N., & Prabhat. (2019). Deep learning and process understanding for data-driven Earth system science. *Nature*, *566*(7743), 195–204.

Remmel, T. K., & Fortin, M. J. (2013). Categorical, class-focused map patterns: Characterization and comparison. *Landscape Ecology*, *28*(8), 1587–1599.

Robertson, C., Long, J. A., Nathoo, F. S., Nelson, T. A., & Plouffe, C. C. F. (2014). Assessing Quality of Spatial Models Using the Structural Similarity Index and Posterior Predictive Checks. *Geographical Analysis*, *46*, 53–74.

Sampat, M. P., Wang, Z., Gupta, S., Bovik, A. C., Markey, M. K., & Member, S. (2009). Complex Wavelet Structural Similarity : A New Image Similarity Index. *IEEE Transactions on Image Processing*, *18*(11), 2385–2401.

Sylvain, J. D., Drolet, G., & Brown, N. (2019). Mapping dead forest cover using a deep convolutional neural network and digital aerial photography. *ISPRS Journal of Photogrammetry and Remote Sensing*, *156*, 14–26.

Szilassi, P., Bata, T., Szabo, S., Czucz, B., Molnar, Z., & Mezosi, G. (2017). The link between landscape pattern and vegetation naturalness on a regional scale. *Ecological Indicators*, *81*(April), 252–259.

Remmel, T. K. (2020). Distributions of Hyper-Local Configuration Elements to Characterize, Compare, and Assess Landscape-Level Spatial Patterns. *Entropy*, 22(4), 420.

Tewkesbury, A. P., Comber, A. J., Tate, N. J., Lamb, A., & Fisher, P. F. (2015). A critical synthesis of remotely sensed optical image change detection techniques. *Remote Sensing of Environment*, *160*, 1–14.

Townshend, J. R., Masek, J. G., Huang, C., Vermote, E. F., Gao, F., Channan, S., Sexton, J. O., Feng, M., Narasimhan, R., Kim, D., Song, K., Song, D., Song, X. P., Noojipady, P., Tan, B., Hansen, M. C., Li, M., & Wolfe, R. E. (2012). Global characterization and monitoring of forest cover using Landsat data: Opportunities and challenges. *International Journal of Digital Earth*, *5*(5), 373–397.

Tracewski, L., Bastin, L., & Fonte, C. C. (2017). Repurposing a deep learning network to filter and classify volunteered photographs for land cover and land use characterization. *Geo-Spatial Information Science*, *20*(3), 252–268.

Turner, M. G. (1989a). Landscape Ecology : The Effect of Pattern on Process. *Annual Review of Ecology and Systematics*, *20*, 171–197.

Visser, H., & Nijs, T. De. (2006). The Map Comparison Kit. *Environmental Modelling & Software*, *21*, 346–358.

Wagner, H. H., & Fortin, M. J. (2005). Spatial analysis of landscapes: Concepts and statistics. *Ecology*, *86*(8), 1975–1987.

Wang, Z., Bovik, A. C., Sheikh, H. R., & Simmoncelli, E. P. (2004). Image quality assessment:

from error visibility to structural similarity. *IEEE Transactions on Image Processing*, *13*(4), 600–612.

Wiederholt, R., Paudel, R., Khare, Y., Davis, S. E., Melodie Naja, G., Romañach, S., Pearlstine, L., & Van Lent, T. (2019). A multi-indicator spatial similarity approach for evaluating ecological restoration scenarios. *Landscape Ecology*, *34*(11), 2557–2574.

Willis, K. S. (2015). Remote sensing change detection for ecological monitoring in United States protected areas. *Biological Conservation*, *182*, 233–242.

Wulder, M. A., Coops, N. C., Roy, D. P., White, J. C., & Hermosilla, T. (2018). Land cover 2.0. *International Journal of Remote Sensing*, *39*(12), 4254–4284.

Wulder, M. A., Masek, J. G., Cohen, W. B., Loveland, T. R., & Woodcock, C. E. (2012). Opening the archive: How free data has enabled the science and monitoring promise of Landsat. *Remote Sensing of Environment*, *122*, 2–10.

Zhu, Z., & Woodcock, C. E. (2014a). Continuous change detection and classi fi cation of land cover using all available Landsat data. *Remote Sensing of Environment*, *144*, 152–171.

Zhu, Z., & Woodcock, C. E. (2014b). Continuous change detection and classification of land cover using all available Landsat data. *Remote Sensing of Environment*, *144*, 152–171.

Chapter 2: Using computer vision to compare spatial patterns: understanding scale parameters

2.1 Abstract

Comparison of landscapes and patterns is a long-standing challenge in spatial analysis research. Recently, new models and tools developed for non-geographic image data are being used to study geographic problems involving classification or prediction. Specifically, computer vision models and artificial neural networks have been deployed in an ever-growing number of geographical analysis. In this paper, we review the use of these models in geographical analysis, focusing on the representation and comparison of spatial patterns. We review artificial neural networks and provide a semantic linking across domains using similar model constructs through the lens of scale. We note that scale, a contextual element in geographical research, is typically considered as a model parameter in computer vision. Scale impacts both computer vision techniques and traditional pixel-based or object-oriented analysis; yet computer vision methods are relatively robust to small-scale variations, though parameterization remains ad hoc. A typology of scales therefore provides a framework for mapping model constructs with the aim of developing guidelines for parameterizing and evaluating computer vision models in a geographic context.

Keywords: Computer vision, pattern comparison, scale, convolutional neural networks.

2.2 Introduction

Two current trends in geographical analysis are simultaneously producing an increased demand for methods that facilitate the comparison of landscapes and their characteristic spatial patterns over time. First, the intensification of both natural and anthropogenic landscape change requires large scale monitoring of the environment in order to better understand, mitigate, and adapt to changes in climatic, economic, and social conditions (Townshend *et al.* 2012). Second, the availability of Earth observation sensor data at varying spectral, spatial, and temporal resolutions has enhanced the possibility to develop systems and approaches to rapidly detect and characterize change across broad landscapes (Townshend *et al*. 2012, Wulder *et al.* 2018). Initiatives such as Arctic Observation Networks exemplify the trend for ongoing monitoring of natural systems; developing networked sensors and autonomous systems to support modelling and analysis of environmental system change across the biome (National Science Foundation 2018). Embedded within such initiatives are questions about change: where is arctic tundra vegetation change intensifying? What is the distribution pattern of insect pests causing forest defoliation? Only through analysis of spatial patterns, typically derived from Earth observation data, can such questions begin to be addressed across broad geographic areas (Franklin 1995, Potter *et al.* 2003, Bajocco *et al.* 2011, Hansen and Loveland 2012). Despite this need, methods for comparing the patterns observable in landscapes over time and space have seen limited unified treatment and remain sub-discipline or even study-dependent, with limited work towards a generic set of tools for understanding spatial pattern change (Long and Robertson 2018).

Spatial pattern comparison (SPC) can be defined as the analysis of two or more spatial patterns to characterize their degree of similarity or difference. This definition encompasses a wide array of applications, such as evaluating differences between maps while noting potential

pattern forming processes (Csillag and Boots 2005); detecting spatial and temporal changes (Visser and Nijs 2006); assessing the rate of change or fragmentation of forests (Wickham and Riitters 2019), and so on. For example, Jasiewicz *et al*. (2014) provide an example focused on pattern characteristics of landform elements to quantify similarity in landscape patterns. Such an application is useful for monitoring or locating degraded landscapes for restoration. Monitoring ecosystems productivity, disturbance, topography, and land cover through time can provide early warning information on locations of potential changes to biodiversity (Duro *et al.* 2007). Mairota *et al*. (2013) illustrate how landscape structure analysis can be used to identify areas facing fragmentation risks to enable the implementation of environmental monitoring and intervention programs. The analysis of changes to spatial pattern and forest loss can be used to inform strategies to mitigate landscape fragmentation (Carranza *et al*. 2015).

Geographers and GIScience have long recognized the need to consider scale in pattern comparison workflows, and have provided an extensive discussion on the effects of scale under concepts such as ecological fallacy (Openshaw, 1984) and, the modifiable aerial unit problem (MAUP) (Marceau and Hay, 1999; Dark and Bram 2007). *Grain* and *extent* are two essential components of scale. The finest spatial or temporal resolution available for a given dataset is referred to as grain, while the total size of an area under study or the duration of a study is termed as extent (Turner et al., 1989b). Landscape patterns tend to exhibit varying relations when measured over a range of scales (i.e., grain and extent) (Wu 2004). This comes with several implications for pattern analysis; for example, Townsend et al. (2009) demonstrated that landscape metrics for maps derived from remotely sensed imagery at varying grain size differed significantly, with varying map extent frequently resulting in substantial differences. Generally, the values of descriptive statistics derived from classification are profoundly affected when

measurement scale and aggregation level are altered (Marceau et al., 1994). Additionally, results derived from pattern metrics are more variable at the per class level than the landscape level but show more consistency and predictability with changing grain size than changing extent (Wu 2004). These challenges partly stem from the fact that different spatial processes can act across scales, creating multiscale structural variation (Kulha et al., 2019; Kulha et al., 2020). To date, these issues persist, and with remotely sensed data now available and ranging from sub-meter to kilometres in spatial resolution, scale issues in pattern comparison will continue to gain more attention.

Recent research has shown the capability of computer vision methods to quantify and characterize spatial patterns and as tools for assessing spatial models (Robertson *et al*. 2014, Jones *et al.* 2016). Computer vision methods comprise the design and application of algorithms that simulate the functionality of information processing principles of the human visual system (HVS) (Kruger *et al*. 2013). Although computer vision methods are increasingly being adopted for a wide variety of uses, these tools – developed for image data – have yet to be unified and integrated with traditional methods for spatial analysis that were specifically developed for geographic data. Figure 2.1 shows the relationships between computer vision methods, SPC tools, and common information needs in geographical analysis. We illustrate that when searching for tools capable of integrating and facilitating spatial pattern analysis, computer vision methods are potential candidates. For example, spatial accuracy assessment is a common analytical process which is used to quantify discrepancies between a reference map (i.e., labelled as accurate representation) and another map (i.e., denoted as derived map) (Foody 2004, Boots and Csillag 2006). Spatial measures of model goodness-of-fit can also be assessed through comparing observed maps and maps derived from spatial models (Csillag and Boots 2005, Visser

and Nijs 2006). Spatial pattern change detection has been utilized to study spatial process change over space and time (Long and Robertson 2018), while landscape similarity assessment paves a way for spatial process classification (Keane *et al.* 2004, He 2008). Landscape comparison has also been widely utilized to address information needs in content-based image retrieval problems (Jasiewicz *et al.* 2014). Unlike the aforementioned classical approaches, computer vision methods with neural networks as the underlying framework can be used to discover spatial patterns and identify features given some input data (Law et al. 2020, Li and Hsu 2020).

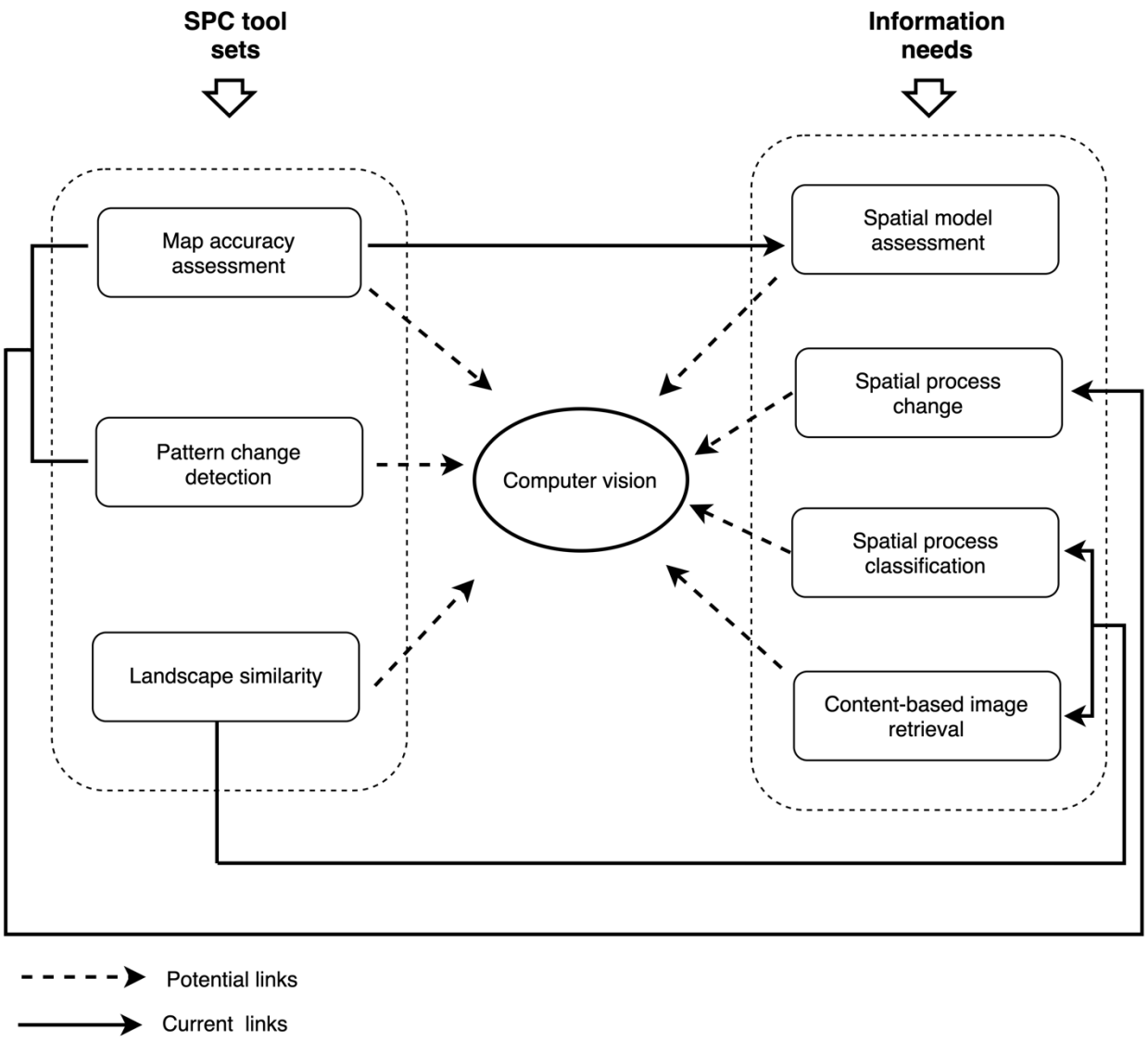


Figure 2.1 Relationships between spatial pattern comparison tools and information needs.

This paper provides a critical review and synthesis of: (1) scale representation in geography/GIScience and its relation to emerging computer vision methods, (2) key scale issues in process-pattern information extraction and SPC using computer vision, and (3) how computer vision methods can be applied to SPC. The paper is structured as follows: computer vision methods – artificial neural networks are introduced briefly, followed by a discussion on scale

issues in geography and computer vision in relation to pattern analysis. Finally, spatial patterns comparison using computer vision methods is presented. We conclude with a summary of our review and potential areas for future research in this emerging research area.

2.3 Overview of Artificial Neural Networks

Artificial neural networks (ANN) are abstractions of biological neurons and are designed to mimic the complex information processing and problem-solving capacity of the human brain (Prieto *et al*. 2016). ANNs are characterized by nonlinearity in input-output mapping, high parallel processing, robustness, fault and failure tolerance, learning, ability to handle fuzzy information, and generalization capacity across domains (Jain and Mao 1996). Researchers in many disciplines continue to develop ANN models to solve variety of problems. Some common problems in which ANN have been utilized include pattern classification, change detection, and forecasting/prediction (Gopal 2016).

There are several discussions in the literature on the theoretical underpinnings and inner workings of ANNs. Fausett (1994) and Gershenfeld (1999) provide extensive details on the theory and applications of ANNs. Figure 2.2 depicts a typical ANN with one hidden layer. Gershenfeld (1999) notes that an ANN with one hidden layer can describe any continuous function and one with two hidden layers (by an obvious extension of the architecture given below) can describe any function at all).

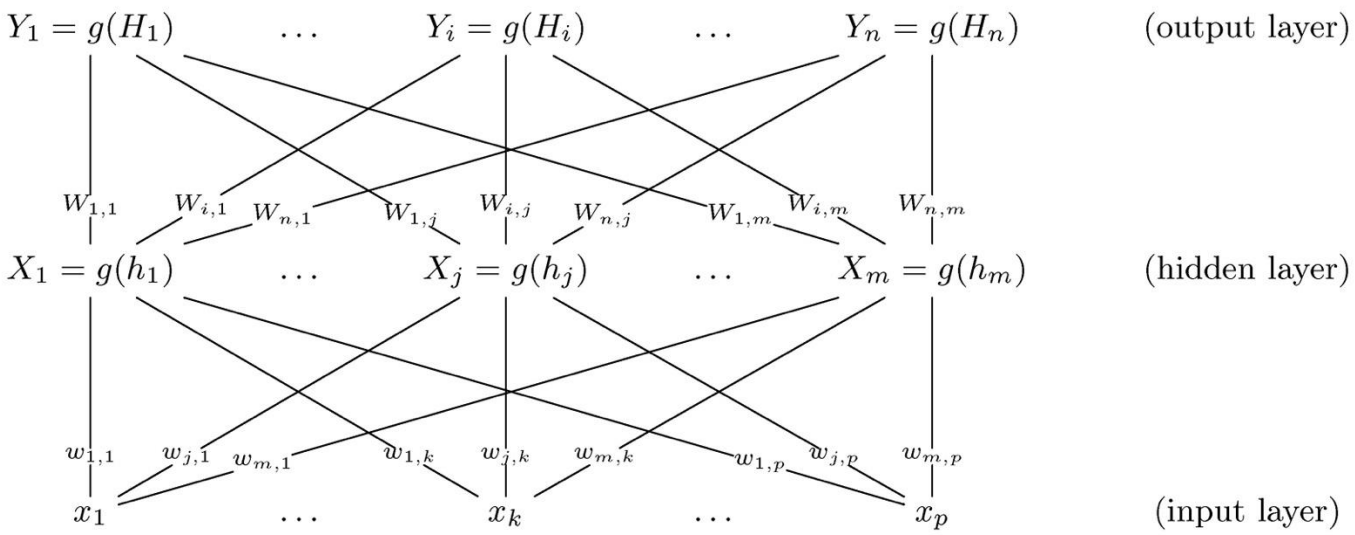


Figure 2.2 ANN architecture for a back propagation single hidden layer neural network.

The inputs x_k are combined linearly with a set of weights $w_{j,k}$. The network weights, $w_{j,k}$ and $W_{i,j}$, are usually determined by applying *n* training sets of data (e.g., a vector quantity), $y_i(\underline{x}_n)$, with known inputs, $\underline{x}_n$'s, and outputs, y_i's to the network that is initialized to small randomly generated weights and optimizing the weight values by minimizing the least squares error function of the difference of $y_i(\underline{x}_n)$ and $Y_i(\underline{x}_n)$. Once trained, novel inputs x_k^* are applied and the ANN generates outputs Y_i^* based on the previous learning encoded in the ANN weights.

2.4 Convolutional Neural Networks

Convolutional neural networks (CNN) are variants of ANNs that incorporate spatial, contextual and textural information in their processing of inputs through the utilization of spatial neighborhoods (e.g., convolutional filters or kernels) (Lecun et al., 2015). Convolutional filters/kernels are n × n windows that are employed to extract pattern information captured in image pixels. In figure 2.3, each convolutional layer (e.g., conv1, conv2, conv3 and conv4) applies filters/kernels of arbitrary dimension to extract features. Features may be edges, lines, corners, object parts or even whole objects contained in images. A feature map denotes a 2D image extracted from a CNN layer using a filter. A fully convolutional CNN model contains no fully connected layers (i.e., layers with 1D feature vectors) (Springenberg et al. 2014), thus retains spatial information inherent in feature maps. Feature pooling (i.e., down-sampling) may be applied to reduce the spatial dimension of 2D maps. The complexity of pattern representation increases with depth of the network. Lower layers (i.e., conv1) learn to detect fine grain features, edges, lines and corners, while higher layers encode global information (i.e., object parts, whole objects and shapes) (Nogueira *et al.* 2017).

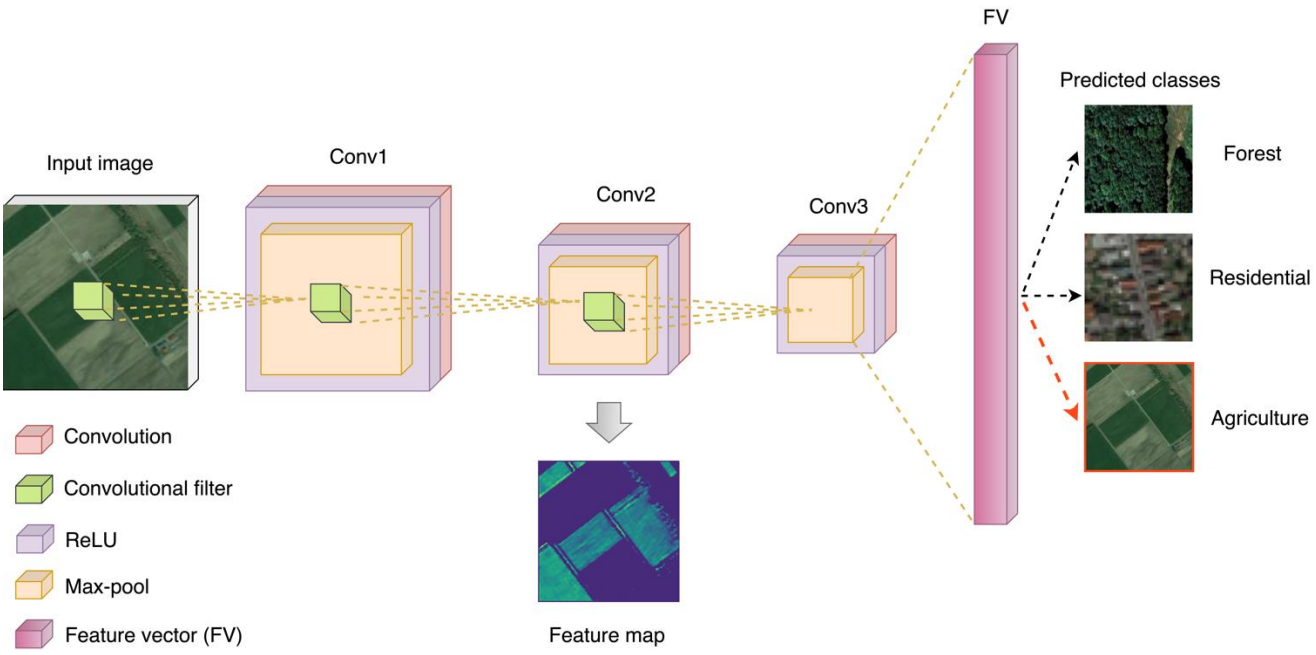


Figure 2.3 Scene segmentation and classification CNN model. Fully convolutional layers (conv1, conv2 and conv3) process pattern information hierarchically.

Earlier techniques to detect objects and edges bear some functional link with window-based operators which have been used in remote sensing for decades. For example, the Sobel filter detects edge features in images by using a simple pre-specified weighting scheme to find sharp gradients in image data. This is similar in spirit to features detected in the lower layers of a CNN model, except in CNNs the filter weights are unknown and 'learned' by relating image data in the filters to class labels provided to the model. The weights that get learned by the model are then 'hard-coded' in a sense, and are therefore very fast to execute on new data once they are known (i.e., the model is trained). Having enough training data to learn a representation of the weights is a key issue in applying to geographic data, one that is somewhat alleviated by growing

archives of satellite image data, however labelled (i.e., annotated ground-truth data) geographic image data required for training a CNN is still somewhat lacking.

Patterns detected in lower layers of a CNN provide crucial information for higher CNN layers. In Figure 2.4, we show agriculture landscape and forest with corresponding patterns extracted using Sobel filter and CNN filters. In Figure 2.4 (row a), it can be seen that as the Sobel filter detects agriculture landscape boundaries, the CNN filter learned to detect whole agricultural land with vegetation. Also, notice that the Sobel filter detects tree canopy edges as well as edges of shadows, while the CNN feature map ignores shadow regions but shows finer patterns of the entire forest landscape. The Sobel filter picks up any sharp gradient (e.g., edge pixels) in image data, while the CNN filter *learns significant patterns* based on class labels so has an advantage of a priori information as opposed to the unsupervised Sobel filter.

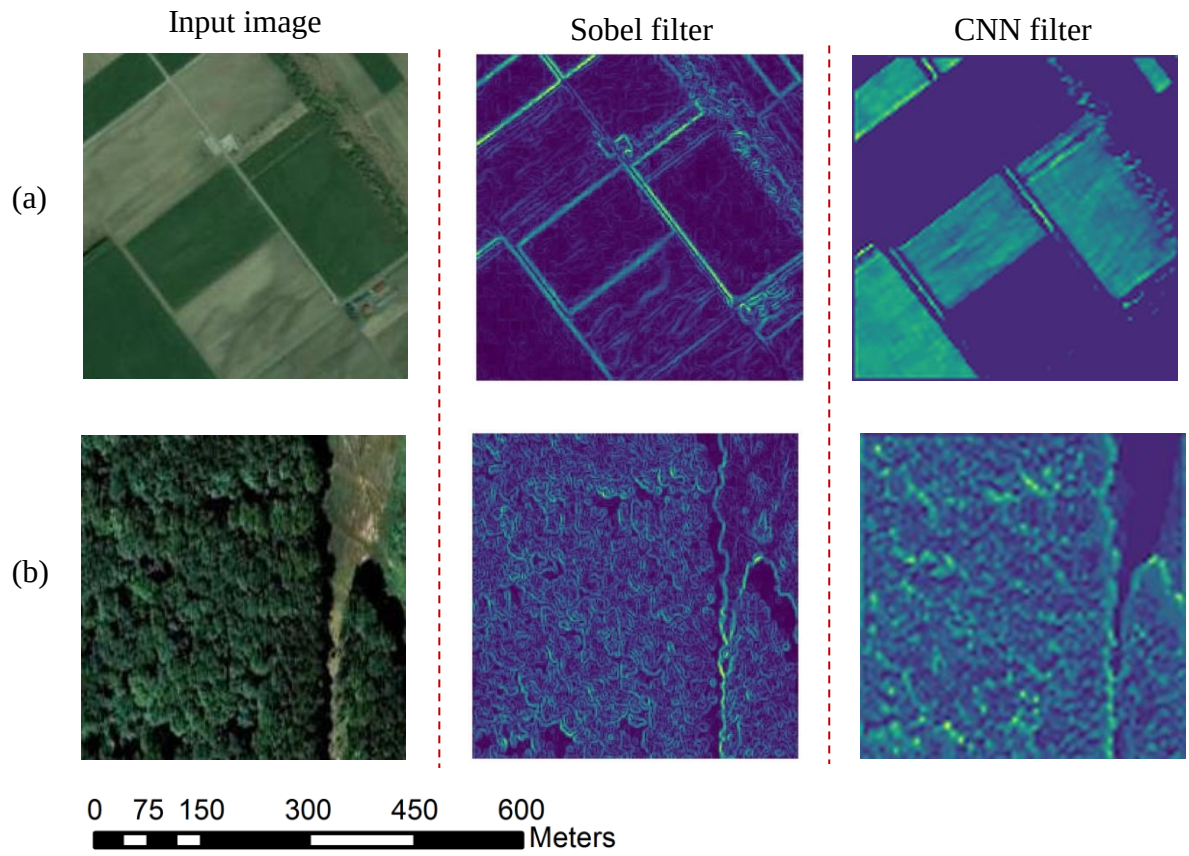


Figure 2.4 Edge detectors versus CNN higher layer filters. Row (a) shows an agricultural landscape with land parcels detected utilizing a Sobel filter and CNN filters. Row (b) depicts a forest stand with edge detection using a Sobel filter and the forest pattern extraction using a CNN filter.

2.5 Spatial Patterns and Scale

A spatial pattern is a scale-dependent outcome of one or multiple interacting spatial processes. Process-pattern interactions refer to the complex interrelationships between spatial patterns and pattern generating processes (Turner 1989a; McIntire and Fajardo 2009). The extent over which a pattern is defined might be the global extent of the dataset, or some sub-unit defined by tessellation, gridding, or geographically meaningful localized criteria – often termed a spatial

neighborhood (Nelson and Robertson 2012). Similarities and/or differences observed between maps when making spatial pattern comparisons may be a function of the size of the analytical unit adopted and the extent over which the patterns are defined (Boots and Csillag 2006).

Spatial pattern analysis involves quantifying and describing patterns of variability in space and time to understand how patterns vary with scale (Jelinski and Wu 1996), as well as the interactions among patterns and underlying spatial processes (Levin 1992). It is widely accepted that the scale at which patterns are analyzed influences results (Wu and Li 2006). When observation scales are altered, the resulting statistical outcomes, such as mean patch size and variance, spatial autocorrelation, and multivariate relationships, may change (Dungan *et al.* 2002, Dark and Bram 2007; Wheatley 2010). As pointed out earlier, scale may further impact analytical units selected for pattern comparison. Choosing the right scale however remains a daunting challenge. The dynamic and self-organized nature of processes and the hierarchical structure of patterns they generate leads to scale and scaling problems (Garry 2000). Garry (2000), for instance, highlights the complications involved in translating models and data across scales, and emphasize the need to use key processes that are relatively stable across scales to determine appropriate scaling methods. Wu and Qi (2000) on the other hand suggest understanding patterns and processes at broad scales (e.g., regional or global) and mapping them to their corresponding fine scales patterns and processes in order to transfer information between scales.

Beyond the analytical challenges associated with choosing analytical scale in geographical analysis, there has been a growing expansion of the terminology used to refer to three underlying concepts of scale: 1) spatial/temporal resolution (often termed grain), 2) spatial/temporal extent, and 3) spatial neighborhood (e.g., spatial window, local-sub extent). In Table 2.1, we seek to

harmonize scale related terminologies from computer vision with those used in geographical analysis. To quantify patterns, analytical units as well as scales over which patterns are to be characterized must be explicitly established. Here, we suggest possible observational scales and how they could be selected and implemented when analyzing spatial patterns.

Table 2.1 A summary of scale terminologies in computer vision and across other domains

	Observational scale hierarchy →				
Domain	**Singleton level** (This scale is the basic/atomic unit of any spatial data)	**Local level** (Local scale may consists of an arbitrary aggregation of pixels or objects)	**Landscape-level/Regional** (An area with a collection of features of defined attributes, administrative boundaries etc. Sizes of landscapes /regions may vary considerably)	**Global level** (The entirety of an area, spatial units, data, features etc. being studied)	**References**
GIS/GIScience	Pixel (for raster data), and object (vector data), a patch (e.g., fire-burned area)	Kernel, moving window, local neighborhood, focal operator	Spatial grid, tile, scene, fishnet, group/set of polygons, population or communities (e.g., forested area), administrative areas	A collection of all pixels/objects under a system being studied, A collection of global features (e.g., global landcover, lakes and wetlands)	(Wu 1999, Dungan *et al.* 2002, Nowosad and Stepinski 2018a, Dabiri and Blaschke 2019)
Computer Vision	Pixel, image object	Neuron, kernel, filter, feature detector, detection window	Contextual window, image patch, input image/tile,	A collection of all image pixels / objects, Contiguous set of contextual windows	(Sharma *et al.* 2017, Nogueira *et al.* 2017)
Remote Sensing	Pixel	Kernel, filter, moving window	Sampling grid, scene, spatial extent	Sequences of data/scenes (e.g., images covering a satellite's temporal resolution), all images/scenes sampled over an area and features/landcover types of interest	(Marceau and Hay 1999, Wu and Li 2009)
Medical Imaging	Voxel (3D pixel), pixel (2D)	Local neighborhood, filter (2D/3D), kernel (2D/3D), part of an organ (e.g., lower lungs, part of brain)	Image patch, anatomical region, organ (e.g., heart, kidney)	A collection of all image pixels/ objects, whole image object (i.e., for an image containing one object)	(Shin *et al.* 2016, Altaf *et al.* 2019, Bernal *et al.* 2019)

2.5.1 Singleton level

Image pixels and objects are the basic analytical units for raster and vector datasets, respectively (Dungan 2006). These geographic primitives are also the atomic unit in which any geographic phenomena may be captured and represented, thus form the bottom of the observational scale hierarchy.

2.5.1.1 Pixels

Pixel values reflect fundamental local properties of lattice data (Dungan 2006). While individual pixels cannot represent patterns, pixel resolution dictates the scale of the information about the underlying pattern generating process that manifests in spatial patterns that have been captured in data. Figure 2.5 shows possible scenarios for representation of spatial processes in raster data. For scenario 1, processes are operating at a finer resolution while data pixel spatial resolution tends to be coarser. This implies critical information pertaining to the underlying spatial processes may not be effectively sampled. In both scenarios 2 and 3, there could be a good alignment between underlying spatial processes and pixels spatial resolution, making the data useful for analysis (Comber and Wulder 2019).

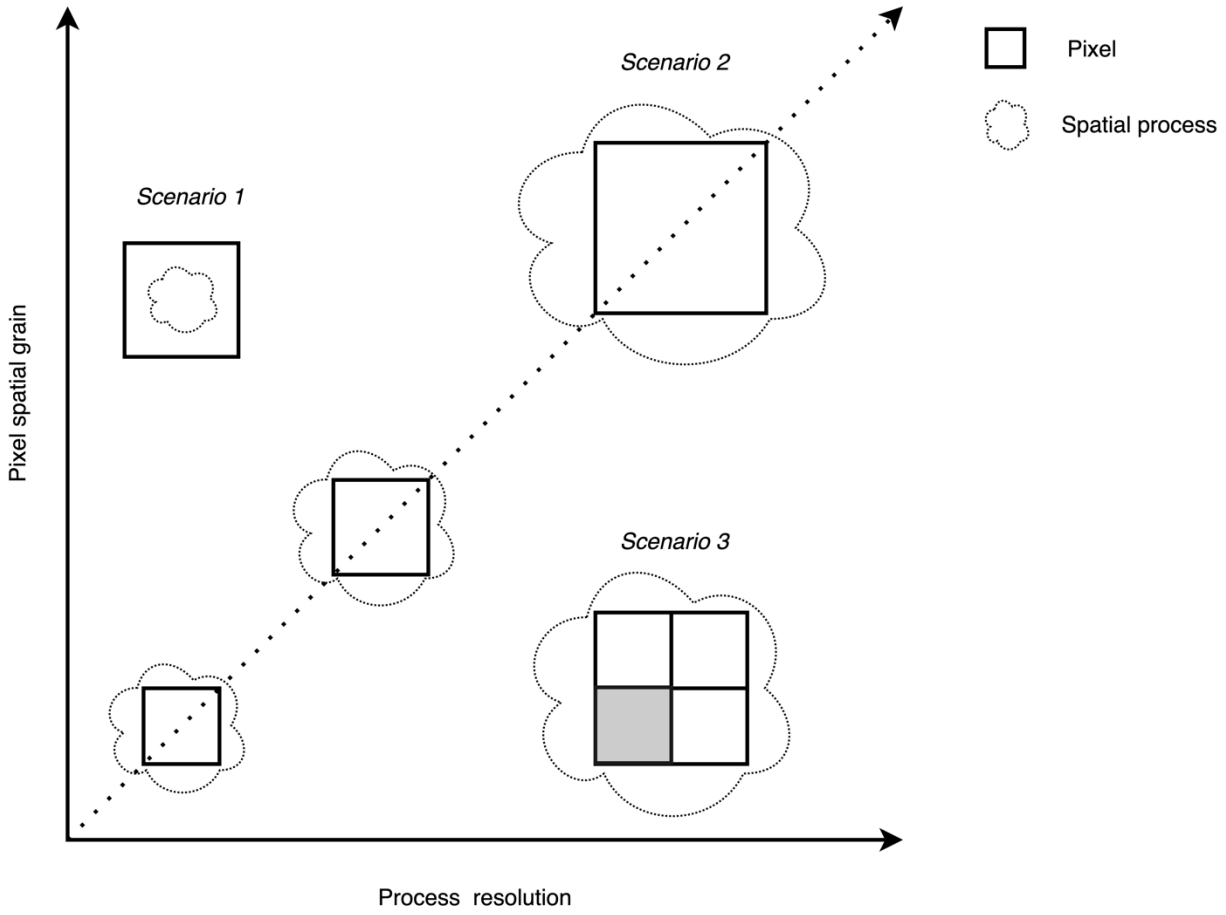


Figure 2.5 Process resolution and pixel spatial resolution. Processes are denoted by irregularly shaped areas confined by dotted lines, while pixels are depicted as square boxes whose boundaries are delineated by solid lines. Process resolution is the extent of an area over which a given spatial process occurs. For Scenario 1, process is finer compared to pixel resolution. In Scenario 2, along the diagonal, processes and pixels spatial resolution are approximately equal. In Scenario 3, pixels have finer spatial resolution (i.e., smaller grain sizes) than the underlying spatial process.

2.5.1.2 Image Objects

In image analysis, image-objects can be generated using a technique termed *segmentation* (Basaeed *et al*. 2016). Segmentation is typically followed by subsequent comparison of

corresponding image-objects (Chen *et al.* 2012). Vector-based data whose primary units are feature objects can be compared through identifying and mapping object locations between georeferenced datasets. Geographical objects can be categorized into area-based objects (e.g., polygons and image segments), and non-area-based objects (e.g., lines and points). Each of these analytical units present challenges in pattern analysis such as edge correction (Li and Zhang 2007). When an image-object shape and size remain fairly constant across two time points, it can be unambiguously tracked. Object-based analysis may be thought of as an application of an "irregular kernel" or an "irregular neighborhood", where the boundaries of the object represent an irregularly shaped and sized spatial kernel. Object-based approaches deal with individual objects as discrete entities rather than defining arbitrary scales for analysis and can form the basis for upscaling; thus, the MAUP and its attendant consequences can be reduced (Openshaw, 1984). Given the hierarchical nature of landscapes (Urban *et al*. 1987), objects often manifest themselves differently over a range of scales, necessitating multiscale analysis. For example, Hay et al. (2001) proposed a multiscale object-specific framework and argue that their approach may help define critical landscape thresholds, scale domains, boundaries of ecotones, and the grain and extent for developing and applying scale-dependent models, hence reducing MAUP effects. In related study, Hay et al. (2003) showed that Fractal Net Evolution Approach, Linear Scale-Space and Blob-Feature Detection, and Multiscale Object-Specific Analysis facilitate multiscale pattern analysis as they can be used to hierarchically link image-objects and derive spatially explicit multiscale contextual information. Object-based methods can also minimize errors originating from the effects of geo-referencing, higher spectral variability, and data acquisition tools artefacts (Dungan 2006; Hussain *et al.* 2013). Object-based methods make it possible to delineate usable objects from imagery while utilizing spectral and contextual

information in a more integrated way (Blaschke 2010). For instance, corresponding image-objects that have been successfully delineated can be matched across space and time allowing for comparison (Tewkesbury *et al.* 2015).

Object-based approaches are typically associated with very-high spatial resolution data (Hussain *et al*. 2013). High resolution data and segmentation parameters requirements for generating consistent image objects pose challenges for object-based approaches (Carleer *et al.* 2005); for example, inconsistent object sizes, shapes and numbers between image/map pairs can potentially render comparison problematic (Ye *et al.* 2018a). Additionally, textural information which is crucial in object-oriented analysis is largely associated with fine resolution data (Boyd and Danson 2005, Falkowski *et al.* 2009, Hussain *et al*. 2013, Wickham and Riitters 2019). Like many analytical units, scale of analysis can impact objects used to represent patterns. At coarser scales or extents, objects may be reduced to point features, thus their pattern representation signal diminishes, limiting the analytical power and sensitivity of pattern comparison tools that rely on objects. Complicating spatial patterns represented by objects is the fact that the spatial location of objects can potentially change through time (Robertson *et al.* 2007), this phenomenon limits the ability to accurately match and compare objects between maps. Moreover, vector-based objects may come with complex topological attributes that pose challenges in pattern comparison. For example, the presence of multiple objects (i.e., slivers), irregular boundaries, shapes, and unequal sizes, complicate the task of pattern comparison.

2.5.2 Local Scale

Two-dimensional spatial filters (or kernels) are now frequently utilized for extracting pattern information (Warner 2011). Filters are used to capture contextual information (Camps-valls *et al*. 2008). By having a defined local neighborhood, changes to a pixel's value can be interpreted statistically while filtering noise and identifying changes related to patterns. This approach is also useful for representing local texture and contextual relationships (Tewkesbury *et al*. 2015). Kernels also capture the spatial arrangement of differences in pixel values describing a scene, and therefore increase the ability to detect changes in spatial patterns of landscape features (Gillanders *et al*. 2008). Furthermore, contextual information incorporated through kernels is an effective way to isolate spurious change. More importantly, given that objects or pixels in a local spatial neighborhood may exhibit some level of dependency (i.e., spatial correlation), analysis based on kernels is likely to aggregate spatial information and signals most related to the underlying pattern generating processes (Bruzzone and Prieto 2000, Volpi *et al*. 2013). Adequate prior knowledge about the spatial resolution of the underlying process and the imagery can provide information on the choice of filter dimensions (Comber and Wulder 2019). The deployment of filters is however associated with scale and parameterization issues. Kernel-based methods require setting parameters prior to analysis, which is challenging as there is often no well-established theoretical and practical frameworks for texture parameters selection (Warner 2011). Additionally, the approach is highly sensitive to window size; an appropriate window is thus essential to avoiding smoothing-over pattern details (Shi *et al*. 2016).

Several methods have been proposed to characterize and compare patterns of local variability to understand the effects of localized spatial processes. While processes may operate across a wide range of scales, it is argued that observed patterns may be an aggregation of small units upon

which processes have operated on (Levin 1992). Thus, localized analysis represents a starting point into investigating short range processes that potentially generate patterns and to link outcomes to long range phenomena such as climate variables. In Figure 2.6, the notion of local sub-extent is depicted in case 9. A 3 × 3 square unit moving window is used to capture local patterns consisting of four feature types. We note that in computer vision, a 3 × 3 moving window is the equivalent of a convolutional filter. A given unit for local pattern analysis should be chosen based on its potential to represent the scale at which processes operate, as well as the scale that process-pattern interactions occur (Comber and Wulder 2019).

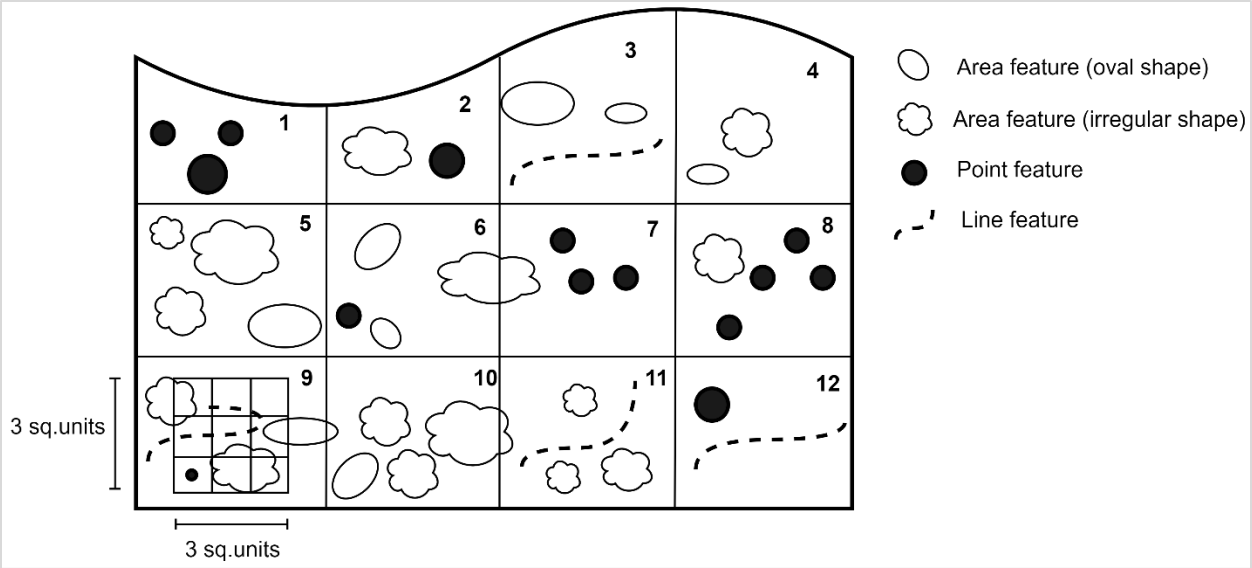


Figure 2.6 Local scale and landscape scale analysis. A 3 × 3 local spatial window is shown in case 9. Local scale analysis can be thought analogously to be a moving window of area 3 × 3 square units. Case 5 – 12 illustrate the notion of landscape level analysis using a larger pre-defined grid (commonly termed a fishnet). Case 1 – 4 highlight the challenges of using regularly shaped sub-units for analysis across regions/study areas that do not conform to those shapes.

2.5.3 Landscape/Regional Scale

Medium scale pattern analysis focuses on landscapes often containing several hundreds of observations (i.e., pixels). Landscape-level analysis is now commonly implemented using contiguous grids (i.e., fishnets) across defined areas. Niesterowicz *et al.* (2016) provide an example in which the size of a grid defines the spatial scale of local landscape. Cells containing differing land cover type in each grid are then represented by a histogram of co-occurrence features which provides insight into the nature of spatial configuration (i.e., spatial arrangement) of patterns. Such grid-based delineations of homogenous areas offer the potential to ignore

spurious local signatures of pattern and to allow for characterization and description of patterns. The selection of a grid size should however be informed by the spatial resolution and the spatial structure of the underlying data (Tisseyre *et al*. 2018). For example, Joseph and Possingham (2008) argue that uncorrelated patterns of species extinction may not be effectively detected by current grid-based methods. To circumvent this challenge, Hengl (2006) suggests adopting the finest grid resolution that encompasses 95% of spatial objects of interest and the coarsest resolution that considers the operational scale and data properties. While landscape and regional scales may seem arbitrary, and sometimes difficult to distinguish, regional scales denote geographically meaningful subdivisions of space whose extent may span administrative boundaries and, in their union, comprise the global study extent. Figure 2.7 illustrates a delineation of an area into four landscape types or regional sub-units with homogenous regions.

The concept of regional scale analysis is, in many ways, analogous to the process of regionalization (or spatial partitioning) (Long *et al*. (2010). Regional sub-extents are delineated via segmentation of the entire area of interest into a set of geographically meaningful single-connected units. Natural regional scale units can be defined according to geographically proximal areas with similar hydrological, geological, ecological, and/or physiographical characteristics. Olson *et al.* (2001) applied a synthetic approach to delineate terrestrial ecoregions of the world. Zhou *et al*. (2003) proposed a model to generate spatially contiguous regions by merging the most similar pair of neighboring polygonal land units. The authors utilized a region growing technique to develop a three hierarchical-level eco-region map of Nebraska. Regionalization is another effective way of generating homogeneous zones for pattern comparison, but the balance between the number of clusters of regions as well as their spatial configuration is of fundamental importance for meaningful pattern analysis. To this end,

(Nowosad and Stepinski 2018b) propose using an index of spatial association called a V-measure to select optimal number of clusters of regions for spatial variance analysis. Klapka *et al.* (2016) emphasize regionalization of administrative areas into functional regions to capture variability of geographical information suitable for pattern analysis. Furthermore, region delineation serves to define ecosystem recovery criteria, extrapolate site-level management, and monitor global change.

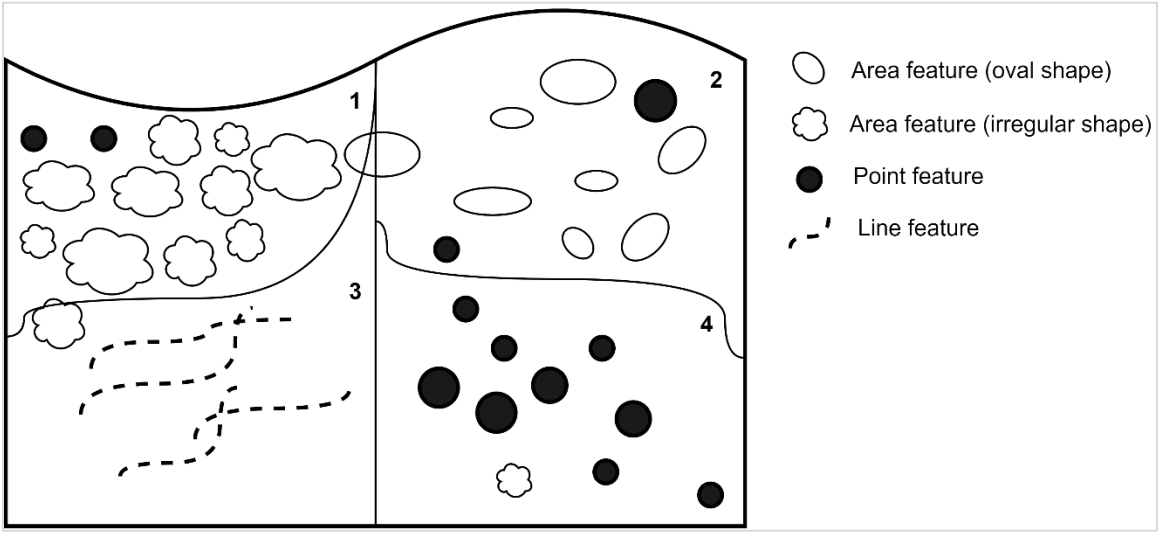


Figure 2.7 Landscape/Regional sub-extents delineation. Boundaries are used to demarcate areas that contain somewhat homogenous feature types.

2.5.4 Global Scale

Local landscapes and regions may be aggregated to give global representation of patterns. Global scale products (e.g., global landcover maps) are also decomposable into their constituent

landscapes or regions using some appropriate algorithms as shown in Nowosad and Stepinski (2018c). One of the earliest global vegetation maps was generated using the Holdridge classification system and general circulation models of the atmosphere via an aggregation of climatic zones (Monserud and Leemans 1992). Zones in this instance could be perceived as landscapes or regions with somewhat similar environmental conditions. A typical global scale analysis is the delineation of the global land area into vegetation, lakes, wetlands or water (Frey and Smith 2007). Similarly, Nowosad *et al*. (2019) combined change patterns in 9 km × 9 km mesoscale landscapes to construct global thematic maps depicting what, where, and the magnitude of change between 1992 and 2015. Given the availability of archived medium to course resolution satellite datasets, and in the wake of declining landcover types, it is becoming increasingly essential to derive global products consisting of information on specific landcover classes. For example, two 1 km × 1 km datasets from the Advanced Very High Resolution Radiometer (AVHRR) were combined to derive global maps showing tree cover, leaf longevity and leaf type (DeFries *et al*. 2000). Similarly, Landsat datasets were used to generate global maps depicting forest cover as well as change areas (Townshend *et al*. 2012). Recently, Li *et al.* (2020a) applied a deep-learning model to create global coral reef probability maps. They used global mosaic of Planet Dove satellite imagery with regional Millennium Coral Reef Mapping Project data for building the model. A wide variety of global landcover products highlight similar trends and reasonable agreements in terms of total area and spatial patterns at global scales. However, comparing individual/local classes in those datasets reveals limited agreement in spatial distribution (McCallum *et al*. 2006).

2.5.5 Scale in computer vision

The notion of scale effects on the extraction of patterns is well recognized in computer vision (Lindeberg, 1993). Lindeberg (1994) argues that a type of multiscale representation (i.e., scale-space theory) proposed by Witkin (1984), can be used for feature detection, feature classification and shape computation as these tasks can be expressed as combinations of Gaussian derivatives at multiple scales. Hay et al. (2002) provide a succinct synthesis of the scale-space theory and highlights its usefulness in exploring and quantifying landscape patterns. The authors suggest that combining remotely sensed data and blob-feature detection techniques with scale-space theory satisfy most of the requirements for multiscale landscape pattern analysis.

In CNN context, Sermanet and Lecun (2011) demonstrated a multiscale approach to traffic signal recognition using a CNN in which stage 1 contains high spatial resolution features while stage 2 yields coarse spatial resolution features; merging these multiresolution features improved the model's performance. Here, we identify three levels of scale within which computer vision methods, specifically, CNNs operate to capture spatial patterns in raster datasets. The first level of scale is a $n \times n$ filter (kernel) which convolves over a group or block of pixels (i.e., local scale) to capture patterns. In computer vision literature, a convolutional filter is equivalent to the concept of moving window commonly used in raster based geographical analysis. Figure 2.8a-c shows three variants of filters that may extract patterns at different scales. The second level of scale is a $m \times m$ *contextual window.* A contextual window can also be referred to as a tile. Figure 2.8d depicts a 10×10 input tile. Tiles are analogous to landscape or regional scale analysis in the geographical literature (i.e., using spatial or fishnet grids). Tiles may vary in size and can be used to capture different scales in the analysis. The third level is the global scale, and can be defined in terms of modeling and prediction phases. At the model building stage, global scale refers to

computing features by averaging over all available information in an input map (Zou *et al*. 2016, Guo *et al.* 2019). During the prediction phase, global scale denotes a set of contiguous tiles representing features (e.g., classifying landcover types in the AVHRR dataset) worldwide as in Li *et al.* (2020b). The definition of scale using filter and tile sizes are essential parameters for learning and capturing spatial patterns. Computer vision methods tend to focus principally on local and landscape/regional scales (i.e., filters and tiles), and global scale at modeling phase. Broader global scale structures are rarely used as whole model inputs; hence do not influence the extraction of pattern information.

The choice of a convolutional filter size for different data products requires a careful consideration (Basu *et al*. 2015). For instance, a 3 × 3 filter covers 9 pixels with an area (referred to as the field of view in computer vison) equivalent to 90 m × 90 m for Landsat data, while the same kernel covers 30 m × 30 m for Sentinel-2 data, and 3 m × 3 m for the National Agricultural Image Program dataset. It is important to note that process–pattern interactions represented in these field of views/local sub-extents may exhibit profound differences. Thus, the size of features or patterns unique to the underlying process should be considered. A large filter is useful for capturing large features or patterns while small filters may be effective for extracting patterns where features are relatively small (Peng *et al.* 2017). For example, Peng *et al.* (2017) provide evidence that a CNN with 15 × 15 filter outperforms a baseline CNN with 3 × 3 filter. It should be noted that a 1 × 1 filter does not capture contextual information but is used for channel-wise reduction or expansion in the number feature maps; this enables a model to learn spatial information across feature map channels (Chen *et al.* 2019). Dilated (astrous) filters use a dilation rate (i.e., pixel-neighbor spacing) to skip some defined set of pixels and can potentially expand a model's field-of-view (i.e., spatial extent of a filter) (Chen *et al.* 2017). This property

may be found useful where the underlying dataset, for example, consist of vegetation patches that are further apart. Similarly, an appropriate tile size (also termed the *contextual window*) selection should consider pixel spatial resolution as well as the size of features of interest in the scene. For high resolution data (e.g., 1 m × 1 m pixel size), a 28 × 28 tile dimension may be sufficient to capture contextual information on patterns of tree canopies and grass patches (Basu et al., 2015). However, with coarser spatial resolution data, tile size may be increased to include more data points while considering the size of features in the scene. To overcome challenges related to tile size, attempts have been made to train CNNs on multiscale contextual windows. Such multiscale CNNs have been proven effectively yield robust image representations across various scales for characterizing and classifying patterns (Zhao *et al.* 2018). Recently, Srivastava *et al*. (2020) showed that a multi-input Siamese CNN model improves urban landuse classification by effectively extracting scale invariant features.

In CNNs, the idea of capturing patterns at multiple scales can also appear somewhat implicit. As convolutional operations occur in different hierarchical CNN layers, both the resolution of patterns and feature map (sub-channels of input tiles) dimensions may change. Filters in successive layers operate on patterns at decreasing scales (i.e., reduced spatial resolution), and depending on a model architecture, input tiles spatial dimension may remain constant or reduce via down sampling operations. The consequence of this is that, patterns captured in lower CNN layers are at finer scales while those in higher layers are at coarser scales (Gong *et al.* 2014). Figure 2.8d-e illustrates convolution using a 3 × 3 filter over a 10 × 10 input tile. Down sampling 2.8d by a factor of 2 yields 2.8e (a 5 × 5 feature map). We note that whether down sampling occurs or not, the patterns in feature maps obtained in 2.8e are at a coarser spatial resolution.

Thus, the suitability and sensitivity of patterns extracted in multiple layers for pattern comparison will vary and require careful selection.

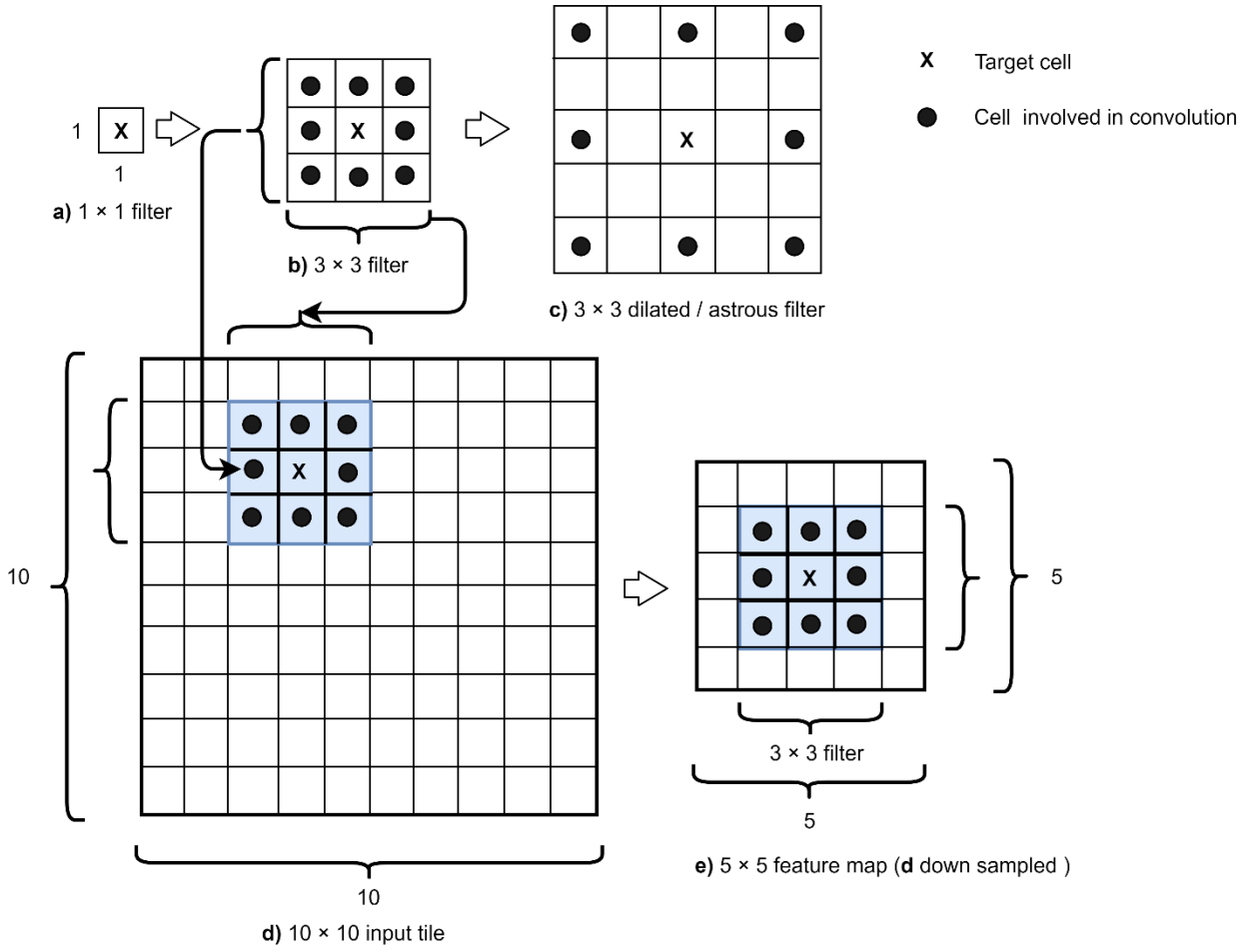


Figure 2.8 Scale levels in computer vision. (a) a 1 × 1 filter, (b) a 3 × 3 filter, (c) a 3 × 3 dilated filter, (d) a 10 × 10 tile (one channel), and (e) a 5 × 5 feature map after convolution and down sampling d. X denotes a target cell which is being convolved.

2.6 Characterizing patterns in computer vision

In recent years, computer vision algorithms are increasingly being employed in remote sensing and geographical analysis (Basaeed *et al.* 2016). In particular, there is a new focus on implementing computer vision models for extracting pattern information and mapping landcover patterns (Liu *et al.* 2018a).

2.6.1 Extracting process-pattern information using computer vision

Computer vision methods deploy simple to fairly complex techniques and functions to extract discriminative information from raster data. Additionally, computer vision methods can potentially exploit multimodal data as inputs such as combining optical remote sensing as well as synthetic aperture radar data (Liu *et al*. 2018c, Mohammadimanesh *et al*. 2019); thus, making the integration of different sensor data possible. Mean, variance and correlation computations are basic processing operations applied using moving windows. Composition, texture and structure which are fundamental components of image data and spatial patterns are well captured in such window operations (Volpi *et al*. 2013). Data transformation methods which involve the use of filters and activation functions (e.g., sigmoid function and rectified linear unit) may result in capturing discriminative patterns in remotely sensed imagery. For example, an activation function can introduce non-linearity in input data (Krizhevsky *et al.* 2012). This enables models to learn or extract complex spatial patterns or data structures.

Figure 2.9 depicts a conceptual framework on how computer vision methods and classical techniques process input data to discover and/or represent spatial patterns. Broken lines indicate potential cross feature utilization between computer vision and classical techniques. We use

double arrows to illustrate potential interdependence between spatial data (i.e., volume, mode, noise level) and pattern comparison methods used to extract and/or represent meaningful process-pattern relationships.

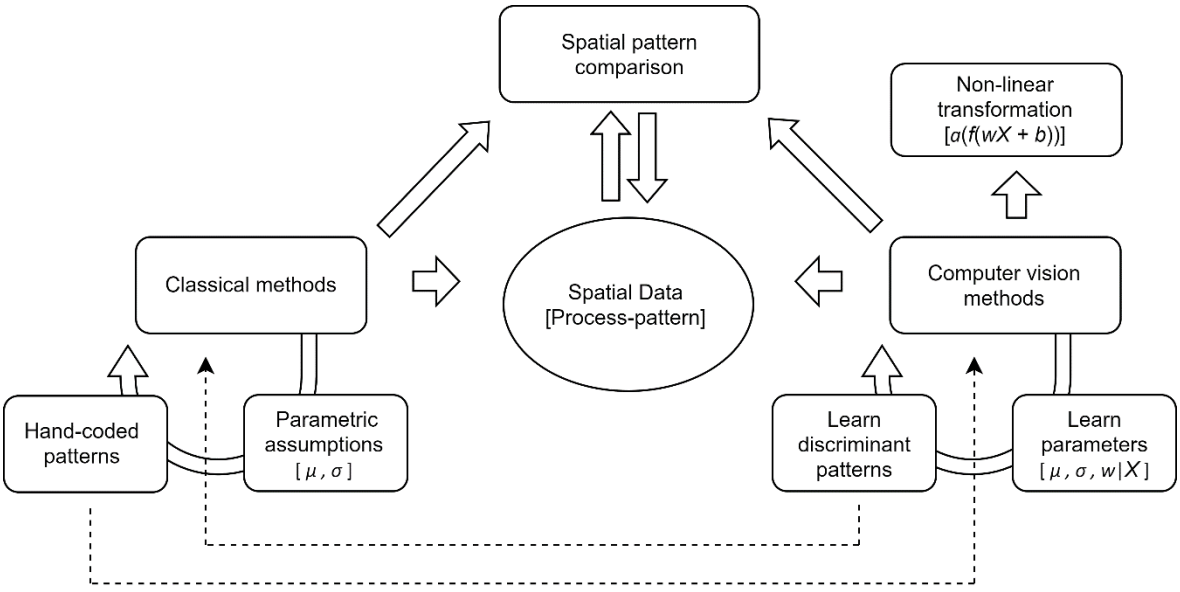


Figure 2.9 A summary of data processing to discover process-pattern information; classical versus computer vision methods. Pattern-process interactions symbolized in spatial data can be understood via spatial pattern comparison. X, w, and b denote data, weight and bias, respectively; μ and σ are mean and standard deviation of X; f denotes an activation function used for non-linear transformation.

Despite the successes reported, remotely sensed data analysis may still pose challenges for computer vision algorithms. The assumption of spatial stationarity over all image local regions will typically not hold for remote sensing imagery acquired under varying viewing angles, shadow intensity and topography. If not adequately alleviated during image pre-processing, these issues can degrade the performance of computer vision methods. Parameter selection is another major challenge as there is no standardized approach to selecting the best model parameters such

as filter/window size, number of features to learn, or model depth (Zhao and Du 2016). Moreover, inadequate training samples, training data errors, and limited ground-truth data are potential constraints to training and evaluating computer vision models (Elmes *et al*. 2020).

2.6.2 Comparing patterns using convolutional neural networks

Characterizing spatial patterns from big geospatial data sources, and detecting changes in these patterns, can be a daunting challenge for classical methods in which physical models are utilized to extract pattern information. Additionally, the complexity of spatial data and implicit spatial relationships constrains the capacity of conventional techniques to extract robust features to characterize spatial patterns (Shekhar *et al*. 2011).

CNN models employ a hierarchical set of spatial filters which are capable of extracting spatial information given a set of imagery (Chen *et al*. 2016, Basaeed *et al*. 2016, Zhang *et al.* 2017). The filters enable spatial weight sharing, and consequently constrain model parameters within reasonable size. Given an $n \times n$ filter and tile, the amount of pattern information to be learned can be manipulated. A large $n \times n$ filter and tile dimension will increase the number of pixels and therefore the amount of information available; spatial dependence information will however vary based on the degree of clustering or dispersion of the pattern of features in the underlying dataset.

Hand-coded filters (i.e., filters with predetermined weights) such as Gabor filters and Sobel filters have been employed in the past decades to detect and characterize patterns; such filters apply element-wise operation to their inputs to yield outputs showing a summary value over local regions. In CNN, activation functions transform inputs while filter weights are learned adaptively during training via backpropagation with loss functions such as hinge loss, mean

absolute error, and cross-entropy loss functions that minimize error between input and output variables. Pooling is implemented primarily to reduce input feature size with the aim to mitigate over overfitting as well as improving computational efficiency. Maximum pooling (max-pooling) takes the maximum value in a pooling kernel (local sub-extent), while average pooling computes the mean value. Max-pooling is frequently used as it is reported to efficiently extract discriminative features in most pattern recognition and object detection tasks (Krizhevsky *et al.* 2012). Pooling operations confer translation invariance in CNN. The translation invariance property enable CNN to learn to extract patterns at varying locations in a given scene. Thus, in geographical analysis and research where similar features or patterns are often located in different/localized regions of a given landscape, CNN holds potential to effectively extract spatial patterns of interest.

Figure 2.10 illustrates a hypothetical feature-based landscape comparison framework in which a pre-trained CNN is employed to extract features. Convolutional feature maps contain useful discriminative information and research has proven their suitability in change analysis (El Amin *et al.* 2016). In landscape pattern comparison, Albert *et al*. (2017) extracted features from fully connected CNN layers to characterize urban environments across several European cities. The authors showed that features from CNNs are robust for representing similarities and variabilities across urban scenes which generally exhibit somewhat complex texture and structure. Amirshahi *et al.* (2017) also propose image quality assessment metric based on CNN feature maps extracted from test and reference images at multiple CNN layers. The authors' image quality assessment problem is similar to pattern comparison in geographical research and emphasizes the superiority of CNNs feature maps for capturing spatial patterns that are useful for similarity assessment.

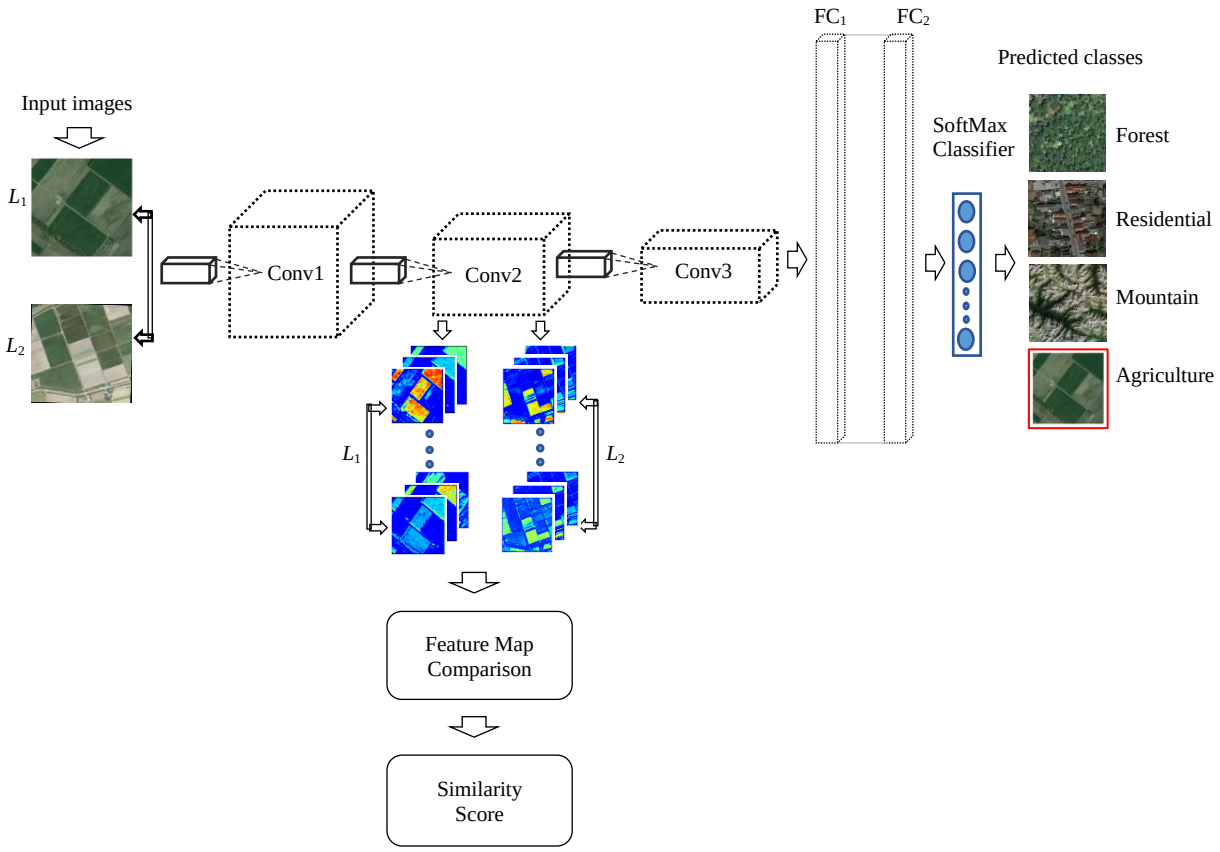


Figure 2.10 Convolutional feature-based similarity comparison. Feature maps are extracted from maps at two locations L1 and L2. Conv1, Conv2, and Conv3 denote convolutional layer 1, 2, and 3. FC1 and FC2 denote fully connected CNN layers. Each location image is propagated independently through a pre-trained CNN and a high dimensional feature map is derived. The resulting feature maps are then compared, and similarity score computed using some algorithm (e.g., the structural similarity index).

2.7 Outlook for computer vision methods in SPC analysis

ANNs offer potential to incorporate different sensor data as inputs. Given that these models do not rely on parametric statistical assumptions of data values, they may be suitable for analyzing

non-Gaussian, multi-modal, noisy and/or missing data, and categorical and continuous data (Rogan and Chen 2004). Fully convolutional CNNs are potential tools for GIScience and geographical research as spatial and topological information which are of fundamental importance in pattern comparison remain intact in model outputs. With the abundance of medium to high resolution imagery, feature-based pattern comparison research has promising potential for improving landscape similarity mapping and pattern change detection accuracy. Feature maps extracted from pre-trained CNNs layers contain hierarchical features at variable resolutions, yet retain discriminative local information that can potentially characterize changes in spatial patterns (El Amin *et al*. 2016). CNNs hold potential to extract process-pattern information, but this may require ad hoc parameterization (i.e., filter size, tile or grid size and input channels).

Understanding how spatial processes generate spatial patterns is an example of a *causality dilemma* whereby spatial patterns often interact with and influence relevant processes, leading to a progressive patterning phenomenon and vice versa (Turner 1989a, McIntire and Fajardo 2009). Behind the characterization of spatial patterns generated by processes is the various dimensions and choice of scale for analysis. We note that while scale recognized as a contextual element in geographical research and computer vision, there are differences between how scale and resolution impact computer vision techniques and traditional pixel-based or object-oriented analysis. For example, given that filters of varying dimensions can be utilized in a CNN model, features of different resolution in raster images may be effectively represented. Additionally, the hierarchical nature of feature extraction in CNN layers, results in capturing multiresolution features, thus minimizing scale effects.

2.8 Conclusion

Pattern comparison strides across many domains with different overarching objectives. For example, while in geographical research the aim may be to understand pattern generating processes, in medical imaging the objective may be focused on identifying unique patterns of certain tumors in order to relate them to the underlying causal elements. In either case, scale will directly impact the ultimate outcome of the analysis carried out. We therefore emphasize that in CNN models, scale (i.e., filters and tile sizes) needs to be mapped to the underlying pattern generating process of interest to minimize spurious information while extracting "significant" pattern signals.

Computer vision methods, specifically, spatially explicit CNN models have improved pattern recognition and object detection tasks (Xie *et al*. 2020). For instance, variations in image scenes resulting from illumination, shadows, and other artefacts not related to the underlying pattern generating process can be effectively handled in CNNs. However, investigation into CNN feature maps suitability and deployment for pattern comparison, from local to global scale is warranted. Additionally, given the complexity of remotely sensed data, it is essential future research investigates the robustness of pattern information extracted from CNNs using a combination of scales (e.g., varying filters and input tiles) on multimodal datasets (synthetic aperture radar versus Landsat) and multiresolution datasets (e.g., 30 m × 30 m Landsat versus 8 m × 8 m Sentinel-2).

Acknowledgements

The authors acknowledge funding provided by the Natural Sciences and Engineering Research Council, Canada and received by Dr. Robertson.

References

Albert, A., Kaur, J., & Gonzalez, M. (2017). Using convolutional networks and satellite imagery to identify patterns in urban environments at a large scale. *In Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 1357–1366.

Altaf, F., Islam, S. M. S., Akhtar, N., & Janjua, N. K. (2019). Going deep in medical image analysis: Concepts, methods, challenges, and future directions. *IEEE Access*, *7*, 99540–99572.

Amirshahi, S. A., Pedersen, M., & Yu, S. X. (2017). Image quality assessment by comparing CNN features between images. *IS and T International Symposium on Electronic Imaging Science and Technology*, 42–51.

Bajocco, S., Salvati, L., & Ricotta, C. (2011). Land degradation versus fire: A spiral process? *Progress in Physical Geography*, *35*(1), 3–18.

Basaeed, E., Bhaskar, H., Hill, P., Al-Mualla, M., & Bull, D. (2016). A supervised hierarchical segmentation of remote-sensing images using a committee of multi-scale convolutional neural networks. *International Journal of Remote Sensing*, *37*(7), 1671–1691.

Basu, S., Ganguly, S., Mukhopadhyay, S., DiBiano, R., Karki, M., & Nemani, R. (2015). Deepsat: a learning framework for satellite imagery. In *Proceedings of the 23rd SIGSPATIAL international conference on advances in geographic information systems*, 1-10.

Bernal, J., Kushibar, K., Asfaw, D. S., Valverde, S., Oliver, A., Marti, R., & Llado, X. (2019). Deep convolutional neural networks for brain image analysis on magnetic resonance imaging: a review. *Artificial Intelligence in Medicine*, *95*, 64–81.

Blaschke, T. (2010). Object based image analysis for remote sensing. *ISPRS Journal of*

Photogrammetry and Remote Sensing, *65*(1), 2–16.

Boots, B., & Csillag, F. (2006). Categorical maps , comparisons , and confidence. *Journal of Geographical Systems*, 109–118.

Boyd, D. S., & Danson, F. M. (2005). Satellite remote sensing of forest resources : three decades of research. *Progress in Physical Geography*, *1*, 1–26.

Bruzzone, L., & Prieto, D. F. (2000). Automatic Analysis of the Difference Image for Unsupervised Change Detection. *IEEE Transactions on Geoscience and Remote Sensing*, *38*(3), 1171–1182.

Camps-valls, G., Gómez-chova, L., Muñoz-marí, J., Rojo-álvarez, J. L., & Martínez-ramón, M. (2008). Kernel-Based Framework for Multitemporal and Multisource Remote Sensing Data Classification and Change Detection. *IEE Transactions on Geoscience and Remote Sensing*, *46*(6), 1822–1835.

Carleer, A. P., Debeir, O., & Wolff, E. (2005). Assessment of very high spatial resolution satellite image segmentations. *Photogrammetric Engineering and Remote Sensing*, *71*(11), 1285–1294.

Carranza, M. L., Hoyos, L., Frate, L., Acosta, A. T. R., & Cabido, M. (2015). Measuring forest fragmentation using multitemporal forest cover maps: Forest loss and spatial pattern analysis in the Gran Chaco, central Argentina. *Landscape and Urban Planning*, *143*, 238–247.

Chen, G., Hay, G. J., Carvalho, L. M. T., & Wulder, M. A. (2012). Object-based change detection. *International Journal of Remote Sensing*, *33*(14), 4434–4457.

Chen, L., C., Papandreou, G., Schroff, F., & Adam, H. (2017). Rethinking Atrous Convolution for Semantic Image Segmentation. http://arxiv.org/abs/1706.05587

Chen, Yilong, Wang, K., Liao, X., Qian, Y., Wang, Q., Yuan, Z., & Heng, P. A. (2019). Channel-Unet: A Spatial Channel-Wise Convolutional Neural Network for Liver and Tumors Segmentation. *Frontiers in Genetics*, *10*, 1–13.

Chen, Yushi, Jiang, H., Li, C., Jia, X., & Ghamisi, P. (2016). Deep Feature Extraction and Classification of Hyperspectral Images Based on Convolutional Neural Networks. *IEEE Transactions on Geoscience and Remote Sensing*, *54*(10), 6232–6251.

Comber, A., & Wulder, M. (2019). Considering spatiotemporal processes in big data analysis: Insights from remote sensing of land cover and land use. *Transactions in GIS*, 879–891.

Csillag, F., & Boots, B. (2005). Toward Comparing Maps as Spatial Processes. *In Developments in Spatial Data Handling*, 641–652.

Dabiri, Z., & Blaschke, T. (2019). Scale matters: a survey of the concepts of scale used in spatial disciplines. *European Journal of Remote Sensing*, *52*(1), 419–435.

Dark, S. J., & Bram, D. (2007). The modifiable areal unit problem (MAUP) in physical geography. *Progress in Physical Geography*, *31*(5), 471–479.

DeFries, R. S., Hansen, M. C., Townshend, J. R. G., Janetos, A. C., & Loveland, T. R. (2000). A new global 1-km dataset of percentage tree cover derived from remote sensing. *Global Change Biology*, *6*(2), 247–254.

Dungan, J. L., Perry, J. N., Dale, M. R. T., Legendre, P., Citron-Pousty, S., Fortin, M. J., Jakomulska, A., Miriti, M., & Rosenberg, M. S. (2002). A balanced view of scale in spatial statistical analysis. *Ecography*, *25*(5), 626–640.

Dungan, Jennifer L. (2006). Focusing on feature-based differences in map comparison. *Journal of Geographical Systems*, *8*(2), 131–143.

Duro, D. C., Coops, N. C., Wulder, M. A., & Han, T. (2007). Development of a large area

biodiversity monitoring system driven by remote sensing. *Progress in Physical Geography*, *31*(3), 235–260.

El Amin, A. M., Liu, Q., & Wang, Y. (2016). Convolutional neural network features based change detection in satellite images. *First International Workshop on Pattern Recognition. International Society for Optics and Photonics*, *10011*, 100110W.

Elmes, A., Alemohammad, H., Avery, R., Caylor, K., Eastman, J. R., Fishgold, L., Friedl, M. A., Jain, M., Kohli, D., Bayas, J. C. L., Lunga, D., McCarty, J. L., Pontius, R. G., Reinmann, A. B., Rogan, J., Song, L., Stoynova, H., Ye, S., Yi, Z. F., & Estes, L. (2020). Accounting for training data error in machine learning applied to earth observations. *Remote Sensing*, *12*(6), 1–39.

Falkowski, M. J., Wulder, M. A., White, J. C., & Gillis, M. D. (2009). Supporting large-area, sample-based forest inventories with very high spatial resolution satellite imagery. *Progress in Physical Geography*, *33*(3), 403–423.

Foody, G. M. (2004). Thematic Map Comparison : Evaluating the Statistical Significance of Differences in Classification Accuracy. *Photogrammetric Engineering & Remote Sensing*, *70*(5), 627–633.

Franklin Janet. (1995). Predictive Vegetation Mappimg: geographic modelling of biospatial patterns in relation to environmental gradients. *Progress in Physical Geography*, *19*(4), 474–499.

Frey, K. E., & Smith, L. C. (2007). How well do we know northern land cover? Comparison of four global vegetation and wetland products with a new ground-truth database for West Siberia. *Global Biogeochemical Cycles*, *21*(1).

Garry D., P. (2000). Scaling and ecological dynamics: Self-organizaton, hierachical structure,

and ecological resilience. *Climate Change*, *44*, 291–309.

Gillanders, S. N., Coops, N. C., Wulder, M. A., Gergel, S. E., & Nelson, T. (2008). Multitemporal remote sensing of landscape dynamics and pattern change: Describing natural and anthropogenic trends. *Progress in Physical Geography*, *32*(5), 503–528.

Gong, Y., Wang, L., Guo, R., & Lazebnik, S. (2014). Multi-Scale Orderless Pooling of Deep Convolutional Activation Features. *In European Conference on Computer Vision*, 392–407.

Gopal, S. (2016). Artificial Neural Networks in Geospatial Analysis. *International Encyclopedia of Geography: People, the Earth, Environment and Technology*, 1–7.

Guo, Y., Ji, J., Lu, X., Huo, H., Fang, T., & Li, D. (2019). Global-Local Attention Network for Aerial Scene Classification. *IEEE Access*, *7*, 67200–67212.

Gustafson, E. J. (1998). Quantifying Landscape Spatial Pattern : What Is the State of the Art ? *Ecosystems*, *1*(2), 143–156.

Hansen, M. C., & Loveland, T. R. (2012). A review of large area monitoring of land cover change using Landsat data. *Remote Sensing of Environment*, *122*, 66–74.

Hay, G. J., Dubé, P., Bouchard, A., & Marceau, D. J. (2002). A scale-space primer for exploring and quantifying complex landscapes. *Ecological Modelling*, *153*(1–2), 27–49.

Hay, Geoffrey J., Blaschke, T., Marceau, D. J., & Bouchard, A. (2003). A comparison of three image-object methods for the multiscale analysis of landscape structure. *ISPRS Journal of Photogrammetry and Remote Sensing*, *57*(5–6), 327–345.

Hay, Geoffrey J., Marceau, D. J., Dube, P., & Bourchad, A. (2001). A Multiscale Framework for Landscape Analysis : Object-specific analysis and upscaling. *Landscape Ecology*, *16*(514), 1–49.

He, H. S. (2008). Forest landscape models: Definitions, characterization, and classification.

Forest Ecology and Management, *254*(3), 484–498.

Hengl, T. (2006). Finding the right pixel size. *Computers and Geosciences*, *32*(9), 1283–1298. https://doi.org/10.1016/j.cageo.2005.11.008

Hussain, M., Chen, D., Cheng, A., Wei, H., & Stanley, D. (2013). Change detection from remotely sensed images: From pixel-based to object-based approaches. *ISPRS Journal of Photogrammetry and Remote Sensing*, *80*, 91–106.

Jain, A. K., & Mao, J. (1996). Artificial Neural Network: A Tutorial. *Communications*, *29*, 31–44.

Jasiewicz, J., Netzel, P., & Stepinski, T. F. (2014a). Landscape similarity, retrieval, and machine mapping of physiographic units. *Geomorphology*, *221*, 104–112.

Jelinski, D. E., & Wu, J. (1996). The modifiable areal unit problem and implications for landscape ecology. *Landscape Ecology*, *11*(3), 129–140.

Jones, E. L., Rendell, L., Pirotta, E., & Long, J. A. (2016). Novel application of a quantitative spatial comparison tool to species distribution data. *Ecological Indicators*, *70*, 67–76.

Joseph, L. N., & Possingham, H. P. (2008). Grid-based monitoring methods for detecting population declines: Sensitivity to spatial scale and consequences of scale correction. *Biological Conservation*, *141*(7), 1868–1875.

Keane, R. E., Cary, G. J., Davies, I. D., Flannigan, M. D., Gardner, R. H., Lavorel, S., Lenihan, J. M., Li, C., & Rupp, T. S. (2004). A classification of landscape fire succession models: Spatial simulations of fire and vegetation dynamics. *Ecological Modelling*, *179*(1–2), 3–27.

Klapka, P., Halás, M., Netrdová, P., & Nosek, V. (2016). The efficiency of areal units in spatial analysis: Assessing the performance of functional and administrative regions. *Moravian Geographical Reports*, *24*(2), 47–59.

Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). ImageNet Classification with Deep Convolutional Neural Networks. *Advances In Neural Information Processing Systems*, 1–9.

Kruger, N., Janssen, P., Kalkan, S., Lappe, M., Leonardis, A., Piater, J., Rodriguez-Sanchez, A. J., & Wiskott, L. (2013). Deep hierarchies in the primate visual cortex: What can we learn for computer vision? *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *35*(8), 1847–1871.

Kulha, N., Pasanen, L., Holmström, L., De Grandpré, L., Gauthier, S., Kuuluvainen, T., & Aakala, T. (2020). The structure of boreal old-growth forests changes at multiple spatial scales over decades. *Landscape Ecology*, *35*(4), 843–858.

Kulha, N., Pasanen, L., Holmström, L., De Grandpré, L., Kuuluvainen, T., & Aakala, T. (2019). At What Scales and Why Does Forest Structure Vary in Naturally Dynamic Boreal Forests? An Analysis of Forest Landscapes on Two Continents. *Ecosystems*, *22*(4), 709–724.

Law, S., Seresinhe, C. I., Shen, Y., & Gutierrez-Roig, M. (2020). Street-Frontage-Net: urban image classification using deep convolutional neural networks. *International Journal of Geographical Information Science*, *34*(4), 681–707.

Lecun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, *521*(7553), 436–444.

Levin, S. A. (1992). The Problem of Pattern and Scale in Ecology. *Ecology*, *73*(6), 1943–1967.

Li, F., & Zhang, L. (2007). Comparison of point pattern analysis methods for classifying the spatial distributions of spruce-fir stands in the north-east USA. *Forestry*, *80*(3), 337–349.

Li, J., Knapp, D. E., Fabina, N. S., Kennedy, E. V., Larsen, K., Lyons, M. B., Murray, N. J., Phinn, S. R., Roelfsema, C. M., & Asner, G. P. (2020b). A global coral reef probability map generated using convolutional neural networks. *Coral Reefs*, *39*(6), 1805–1815.

Li, W., & Hsu, C. Y. (2020). Automated terrain feature identification from remote sensing

imagery: a deep learning approach. *International Journal of Geographical Information Science*, *34*(4), 637–660.

Lindeberg, T. (1993). Lindeberg_Detecting salient blob-like image structures and their scales with a scale-space primal sketch-a method for focus-of-attention. *International Journal of Computer Vision*, *11*(3), 283–318.

Lindeberg, T. (1994). Scale-space theory: A basic tool for analyzing structures at different scales. *Journal of Applied Statistics*, *21*(1–2), 225–270.

Liu, T., Abd-elrahman, A., Morton, J., Wilhelm, V. L., Liu, T., Abd-elrahman, A., Morton, J., & Wilhelm, V. L. (2018a). Comparing fully convolutional networks , random forest , support vector machine , and patch-based deep convolutional neural networks for object- based wetland mapping using images from small unmanned aircraft system. *GIScience & Remote Sensing*, *55*(2), 243–264.

Liu, Z., Li, G., Mercier, G., He, Y., & Pan, Q. (2018b). Change Detection in Heterogenous Remote Sensing Images via Homogeneous Pixel Transformation. *IEEE Transactions on Image Processing*, *27*(4), 1822–1834.

Long, J., Nelson, T., & Wulder, M. (2010). Regionalization of Landscape Pattern Indices Using Multivariate Cluster Analysis. *Environmental Management*, *46*, 134–142.

Long, J., & Robertson, C. (2018). Comparing spatial patterns. *Geography Compass*, *12*(2), 1–18.

Mairota, P., Cafarelli, B., Boccaccio, L., Leronni, V., Labadessa, R., Kosmidou, V., & Nagendra, H. (2013). Using landscape structure to develop quantitative baselines for protected area monitoring. *Ecological Indicators*, *33*, 82–95.

Marceau, D. J., Gratton, D. J., Fournier, R. A., & Fortin, J. P. (1994). Remote sensing and the measurement of geographical entities in a forested environment. 2. The optimal spatial

resolution. *Remote Sensing of Environment*, *49*(93), 105–117.

Marceau, D. J., & Hay, G. J. (1999). Remote sensing contributions to the scale issue. *Canadian Journal of Remote Sensing*, *25*(4), 357–366.

McCallum, I., Obersteiner, M., Nilsson, S., & Shvidenko, A. (2006). A spatial comparison of four satellite derived 1 km global land cover datasets. *International Journal of Applied Earth Observation and Geoinformation*, *8*(4), 246–255.

McIntire, E. J. B., & Fajardo, A. (2009). Beyond description: The active and effective way to infer processes from spatial patterns. *Ecology*, *90*(1), 46–56.

Mohammadimanesh, F., Salehi, B., Mahdianpari, M., Gill, E., & Molinier, M. (2019). A new fully convolutional neural network for semantic segmentation of polarimetric SAR imagery in complex land cover ecosystem. *ISPRS Journal of Photogrammetry and Remote Sensing*, *151*(April), 223–236.

Monserud, R. A., & Leemans, R. (1992). Comparing global vegetation maps with the Kappa statistic. *Ecological Modelling*, *62*(4), 275–293.

National Science Foundation. (2018). https://doi.org/https://www.nsf.gov/pubs/2016/nsf16595/nsf16595.pdf%0A%0A

Nelson, T. A., & Robertson, C. (2012). Refining spatial neighbourhoods to capture terrain effects. *Ecological Prosesses*, *1*(1), 1–11.

Niesterowicz, J., Stepinski, T. F., & Jasiewicz, J. (2016). Unsupervised regionalization of the United States into landscape pattern types. *International Journal of Geographical Information Science*, *30*(7), 1450–1468.

Nogueira, K., Penatti, O. A. B., & dos Santos, J. A. (2017). Towards better exploiting convolutional neural networks for remote sensing scene classification. *Pattern Recognition*,

61, 539–556.

Nowosad, Jakub, & Stepinski, T. F. (2018a). Global inventory of landscape patterns and latent variables of landscape spatial configuration. *Ecological Indicators*, *89*, 159–167.

Nowosad, Jakub, & Stepinski, T. F. (2018b). Towards machine ecoregionalization of Earth's landmass using pattern segmentation method. *International Journal of Applied Earth Observation and Geoinformation*, *69*, 110–118.

Nowosad, J., & Stepinski, T. F. (2018c). Spatial association between regionalizations using the information-theoretical V-measure. *International Journal of Geographical Information Science*, *32*(12), 2386–2401.

Nowosad, Jakub, Stepinski, T. F., & Netzel, P. (2019). Global assessment and mapping of changes in mesoscale landscapes: 1992–2015. *International Journal of Applied Earth Observation and Geoinformation*, *78*, 332–340.

Olson, D. M., Dinerstein, E., Wikramanayake, E. D., Burgess, N. D., Powell, G. V. N., Underwood, E. C., D'amico, J. A., Itoua, I., Strand, H. E., Morrison, J. C., Loucks, C. J., Allnutt, T. F., Ricketts, T. H., Kura, Y., Lamoreux, J. F., Wettengel, W. W., Hedao, P., & Kassem, K. R. (2001). Terrestrial Ecoregions of the World: A New Map of Life on Earth. *BioScience*, *51*(11), 933.

Openshaw, S. (1984). Ecological fallacies and the analysis of areal census data. *Environment & Planning A*, *16*(1), 17–31.

Peng, C., Zhang, X., Yu, G., Luo, G., & Sun, J. (2017). Large kernel matters - Improve semantic segmentation by global convolutional network. *Proceedings - 30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017*, 1743–1751.

Potter, C., Tan, P. N., Steinbach, M., Klooster, S., Kumar, V., Myneni, R., & Genovese, V.

(2003). Major disturbance events in terrestrial ecosystems detected using global satellite data sets. *Global Change Biology*, *9*(7), 1005–1021.

Prieto, A., Prieto, B., Ortigosa, E. M., Ros, E., Pelayo, F., Ortega, J., & Rojas, I. (2016). Neural networks: An overview of early research, current frameworks and new challenges. *Neurocomputing*, *214*, 242–268.

Robertson, C., Long, J. A., Nathoo, F. S., Nelson, T. A., & Plouffe, C. C. F. (2014). Assessing Quality of Spatial Models Using the Structural Similarity Index and Posterior Predictive Checks. *Geographical Analysis*, *46*, 53–74.

Robertson, C., Nelson, T. A., Boots, B., & Wulder, M. A. (2007). *STAMP : spatial – temporal analysis of moving polygons*. 207–227.

Rogan, J., & Chen, D. (2004). Remote sensing technology for mapping and monitoring land-cover and land-use change. *Progress in Planning*, *61*(4), 301–325.

Sermanet, P., & Lecun, Y. (2011). Traffic sign recognition with multi-scale convolutional networks. *Proceedings of the International Joint Conference on Neural Networks*, *June*, 2809–2813.

Sharma, A., Liu, X., Yang, X., & Shi, D. (2017). A patch-based convolutional neural network for remote sensing image classification. *Neural Networks*, *95*, 19–28.

Shekhar, S., Evans, M. R., Kang, J. M., & Mohan, P. (2011). Identifying patterns in spatial information: A survey of methods. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, *1*(3), 193–214.

Shi, X., Ma, G., Chen, F., & Ma, Y. (2016). A Kernel-Based Similarity Measuring for Change Detection in Remote Sensing Images. *ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, *XLI-B7*, 999–1006.

Shin, H. C., Roth, H. R., Gao, M., Lu, L., Xu, Z., Nogues, I., Yao, J., Mollura, D., & Summers, R. M. (2016). Deep Convolutional Neural Networks for Computer-Aided Detection: CNN Architectures, Dataset Characteristics and Transfer Learning. *IEEE Transactions on Medical Imaging*, *35*(5), 1285–1298.

Springenberg, J. T., Dosovitskiy, A., Brox, T., & Riedmiller, M. (2014). *Striving for Simplicity: The All Convolutional Net*. 1–14.

Srivastava, S., Vargas Muñoz, J. E., Lobry, S., & Tuia, D. (2020). Fine-grained landuse characterization using ground-based pictures: a deep learning solution based on globally available data. *International Journal of Geographical Information Science*, *34*(6), 1117–1136.

Tewkesbury, A. P., Comber, A. J., Tate, N. J., Lamb, A., & Fisher, P. F. (2015). A critical synthesis of remotely sensed optical image change detection techniques. *Remote Sensing of Environment*, *160*, 1–14.

Tisseyre, B., Leroux, C., Pichon, L., Geraudie, V., & Sari, T. (2018). How to define the optimal grid size to map high resolution spatial data? *Precision Agriculture*, *19*(5), 957–971.

Townsend, P. A., Lookingbill, T. R., Kingdon, C. C., & Gardner, R. H. (2009). Spatial pattern analysis for monitoring protected areas. *Remote Sensing of Environment*, *113*(7), 1410–1420.

Townshend, J. R., Masek, J. G., Huang, C., Vermote, E. F., Gao, F., Channan, S., Sexton, J. O., Feng, M., Narasimhan, R., Kim, D., Song, K., Song, D., Song, X. P., Noojipady, P., Tan, B., Hansen, M. C., Li, M., & Wolfe, R. E. (2012). Global characterization and monitoring of forest cover using Landsat data: Opportunities and challenges. *International Journal of Digital Earth*, *5*(5), 373–397.

Turner, M. G., Dale, V. H., & Gardner, R. H. (1989b). Predicting across scales: Theory development and testing. *Landscape Ecology*, *3*(3–4), 245–252.

Urban, D. L., O'Neill, R. V, & Shugart, H. H. J. (1987). Landscape Ecology: A hierarchical perspective can help scientists understand spatial patterns. In *BioScience*, 37(2), 119–127.

Visser, H., & Nijs, T. De. (2006). The Map Comparison Kit. *Environmental Modelling & Software*, *21*, 346–358.

Volpi, M., Tuia, D., Bovolo, F., Kanevski, M., & Bruzzone, L. (2013). Supervised change detection in VHR images using contextual information and support vector machines. *International Journal of Applied Earth Observation and Geoinformation*, *20*(1), 77–85.

Warner, T. (2011). Kernel-Based Texture in Remote Sensing Image Classification. *Geography Compass*, *5*(10), 781–798. https://doi.org/10.1111/j.1749-8198.2011.00451.x

Wheatley, M. (2010). Domains of scale in forest-landscape metrics: Implications for species-habitat modeling. *Acta Oecologica*, *36*(2), 259–267.

Wickham, J., & Riitters, K. H. (2019). Influence of high-resolution data on the assessment of forest fragmentation. *Landscape Ecology*, *34*(9), 2169–2182.

Witkin, A. P. (1984). Scale-space filtering: A new approach to multi-scale description. *In ICASSP'84. IEEE International Conference on Acoustics, Speech, and Signal Processing*, 150–153.

Wu, H., & Li, Z. L. (2009). Scale issues in remote sensing: A review on analysis, processing and modeling. *Sensors*, *9*(3), 1768–1793.

Wu, J. (1999). Hierarchy and scaling: Extrapolating information along a scaling ladder. *Canadian Journal of Remote Sensing*, *25*(4), 367–380.

Wu, J. (2004). Effects of changing scale on landscape pattern analysis: Scaling relations.

Landscape Ecology, *19*(2), 125–138.

Wu, J., & Li, H. (2006). *Concepts of scale and scaling*. 3–15.

Wu, J., & Qi, Y. (2000). Dealing with scale in landscape analysis: An overview. *Geographic Information Sciences*, *6*(1), 1–5.

Wulder, M. A., Coops, N. C., Roy, D. P., White, J. C., & Hermosilla, T. (2018). Land cover 2.0. *International Journal of Remote Sensing*, *39*(12), 4254–4284.

Xie, Y., Cai, J., Bhojwani, R., Shekhar, S., & Knight, J. (2020). A locally-constrained YOLO framework for detecting small and densely-distributed building footprints. *International Journal of Geographical Information Science*, *34*(4), 777–801.

Ye, S., Pontius, R. G., & Rakshit, R. (2018). A review of accuracy assessment for object-based image analysis: From per-pixel to per-polygon approaches. *ISPRS Journal of Photogrammetry and Remote Sensing*, *141*, 137–147.

Zhang, H., Li, Y., Zhang, Y., & Shen, Q. (2017). Spectral-spatial classification of hyperspectral imagery using a dual-channel convolutional neural network. *Remote Sensing Letters*, *8*(5), 438–447.

Zhao, R., Pang, M., & Wang, J. (2018). Classifying airborne LiDAR point clouds via deep features learned by a multi-scale convolutional neural network. *International Journal of Geographical Information Science*, *32*(5), 960–979.

Zhao, W., & Du, S. (2016). Learning multiscale and deep representations for classifying remotely sensed imagery. *ISPRS Journal of Photogrammetry and Remote Sensing*, *113*, 155–165.

Zhou, Y., Narumalani, S., Waltman, W. J., Waltman, S. W., & Palecki, M. A. (2003). A GIS-based spatial pattern analysis model for eco-region mapping and characterization.

International Journal of Geographical Information Science, *17*(5), 445–462.

Zou, J., Li, W., Chen, C., & Du, Q. (2016). Scene classification using local and global features with collaborative representation fusion. *Information Sciences*, *348*, 209–226.

Chapter 3: Exploring the use of computer methods for spatial pattern comparison

3.1 Abstract

Detection of changes in spatial processes has long been of interest to quantitative geographers seeking to test models, validate theories, and anticipate change. Given the current 'data-rich' environment of today, it may be time to reconsider the methodological approaches used for quantifying change in spatial processes. New tools emerging from computer vision research may hold particular potential to make significant advances in quantifying changes in spatial processes. In this paper, two comparative indices from computer vision, the structural similarity (SSIM) index and the complex wavelet structural similarity (CWSSIM) index were examined for their utility in the comparison of real and simulated spatial datasets. Gaussian Markov random fields were simulated and compared with both metrics. A case study into comparison of snow water equivalent (SWE) spatial patterns over northern Canada was used to explore the properties of these indices on real-world data. CWSSIM was found to be less sensitive than SSIM to changing window dimension. The CWSSIM appears to have significant potential in characterizing change and/or similarity; distinguishing between map pairs that possess subtle structural differences. Further research is required to explore the utility of these approaches for empirical comparison cases of different forms of landscape change and in comparison to human judgments of spatial pattern differences.

3.2 Introduction

As environmental monitoring technologies increase in scope and scale, providing new and varied perspectives on biotic, abiotic, and social-ecological processes; there is renewed need to compare representations of spatial processes across time and space (Long and Robertson 2018). Motivations for comparison include: (1) detection of spatio-temporal changes, (2) comparison of model outputs, (3) calibration and/or validation of spatial models, (4) analysis of model uncertainty and sensitivity, and (5) assessment of map accuracy (Pontius, Huffaker, and Denman 2004 ; Visser and De Nijs 2006). In a pattern/process framework, detection of change in spatial process typically resolves to characterizing changes in temporally separable spatial patterns, and ideally, assessing the significance of this change relative to candidate spatial processes (Boots and Csillag 2006; Miller and Goodchild 2015). Increasingly, this is making use of time series of patterns/images obtained from growing archives of earth observation data obtained from satellite sensors (Greig, Robertson, and Lacerda 2018; Comber and Wulder 2019). In the geographic analysis context, we consider spatial pattern comparison to be a specific *geographic* instance of a broader class of spatial pattern comparison problems which are prevalent across a wide range of disciplines. Many disciplines, from astronomy (Groth 1986; Makowiecki and Alda 2008) to archeology (Papaodysseus et al. 2008; Zhou et al. 2017), to brain imaging (McIntosh et al. 1996; Luders, Cherbuin and Gaser 2016), rely on techniques to quantify and compare spatial patterns. Some of these have been informed by geographic spatial analysis, while others have not. The geographic context may provide particular local detail, in the form of local knowledge (Robertson and Feick 2017) or local values and understanding (Delmelle 2019) to help discern and interpret analyses of change and change attribution to underlying processes. Such locally situated change analysis may lead to more meaningful change analysis; moving away from

classical notions of randomness to more locally-relevant spatial processes (e.g., Walker and Schuurman 2015). However, what is common to all contexts for quantifying change in spatial processes is the use of spatial patterns to do so (McIntire and Fajardo 2009).

There are two principal settings for spatial pattern comparisons: a) comparing the same landscape over time in order to identify and characterize change, and b) comparing different landscapes at the same (or similar) time in order to quantify differences and similarities across spatial contexts. Much of the research into spatial pattern comparison in the GIScience and remote sensing literature has been focused on the first type of comparison problem. The problem of detecting significant changes (i.e., changes in underlying spatial processes) and disregarding trivial changes due to data collection methods or data errors is not straightforward and requires careful consideration of the processes and patterns of interest (Long and Robertson 2018). Simulation-based approaches to landscape comparison in Remmel and Csillag (2003) provide an example of the complexity associated with pattern quantification in binary landscapes. In a continuous-valued pattern scenario, change quantification is more complex. Despite many methods for change detection in remote sensing (Feng, Sui, and Tu 2016), detection of changes in pattern (often called spatial structure or configuration) are more limited; generally constrained to comparison of univariate global statistics or bivariate spatial pattern statistics (Long and Robertson 2018). The second setting of pattern comparison comprises similarity of specific data types or comparisons associated with specific types of geographic phenomena (Mairota et al. 2013). For example, we may want to identify similar sites to function as reference locations when evaluating disturbance from resource extraction. As global and regional archives of satellite images and routine mapping programs expand, the need for spatial pattern similarity analysis is likely to increase (Comber and Wulder 2019).

Boots and Csillag (2006) outline the dimensions of significant change detection in the context of categorical map comparison, stressing the importance of comparing *data-generating processes* from a spatial perspective. Yet the mechanics of how to realize such comparisons remain an ongoing research challenge, exacerbated in the big data era (Miller and Goodchild 2015). Much of the spatial analysis development for map comparison has concentrated on categorical maps (Lu et al. 2003; Coppin et al. 2004). The Map Comparison Kit software package implements a fuzzy-set comparison (Hagen 2002a; Hagen 2002b; Visser and De Nijs 2006), among other methods such as kappa and accuracy statistics. The Kappa statistic has been frequently deployed in tandem with other traditional metrics for measuring agreement or detecting changes on vegetation maps (Monserud and Leemans 1992; Hagen-Zanker, Straatman, and Uljee 2005). Many of these methods are criticized for their unreliability in quantifying change from no change. For instance, Mas (1999) highlights the sensitivity of classical change detection techniques to poor image enhancement methods. Moreover, there has been several renewed calls to refine the Kappa statistic (Pontius Jr and Millones 2011; van Vliet, Bregt and Hagen-Zanker 2011).

Classical methods of image comparison are point-based methods such as the mean squared error (MSE) or the Dice index (Dice 1945; Dosselmann and Xue 2005)). Other related metrics used in image analysis such as peak signal-to-noise ratio (PSNR) and the signal-to-noise (SNR) are related to the MSE. These metrics have intuitive meaning and are simple to compute. Further, it has been shown empirically and analytically that the MSE for example, is closely linked to the structural similarity (SSIM) index and hence may perform close to the metric (Dosselmann and Yang 2011). Regardless, these metrics quantify global discrepancies that often fail to capture perceived image quality (Wang and Bovik 2002). Additionally, such point-based similarity

metrics ignore spatial structure, and may extensively penalize images whose difference is just a pixel (Wang and Bovik 2006). Pontius (2000) argues that point-based metrics are liable to inaccuracies when there are errors in spatial co-registration. Similarly, Power, Simms, and White (2001) recognize that hierarchical fuzzy pattern matching is sometimes preferable to pixel-based map comparison as the latter can potentially penalize maps that differ owing to few pixel misalignment. Moreover, most point-based indices are sensitive to luminance and contrast (i.e., relative differences in signal level), and depending on the extent, might severely degrade the indices capability. These metrics are also sensitive to geometric and scale distortions resulting from shifts in image acquisition devices and sensor-solar orientations (Sampat et al. 2009). Consequently, such non-structural distortions may reflect and weaken the performance of the above indices in detecting real changes in landscape patterns, especially considering recent interest in drone-based image acquisition systems in environmental research and change analysis using multi-sensor fusion approaches.

Upon recognizing the intrinsic drawbacks of point-based metrics, neighborhood based indices such as the closest distance metric (CDM) and pixel correspondence metric (PCM) are increasingly employed in image comparison (Bowyer, Kranenburg, and Dougherty 2001; Prieto and Allen 2003). The use of spatial neighborhood-based indices can address some of these issues (Ping et al. 2004; Betts et al. 2006; Carl and Kuhn 2007). For instance, Wulder et al. (2007) utilized a local measure of spatial autocorrelation to assess model performance. Unlike point-wise indices, neighborhood-based indices rely on spatial correlation between neighboring pixels in map pairs as a way to incorporate local spatial structure. These indices are relatively resilient to small geometric distortions such as translation, rotation and scale shifts. But, an important setback to the above indices is that they fail to incorporate texture and structural information of

objects in the visual scene and are unable to model the properties of the human visual system (HVS) (Wang and Bovik 2006). In non-geographic contexts, HVS is often used as a benchmark for pattern detection methods – and without clearly defined candidate spatial processes, may be useful in evaluating some types of geographic spatial pattern comparisons.

In computer vision applications, it is widely recognized that texture and the structural component of patterns often reveal tangible information about underlying pattern-generating processes. Metrics capable of capturing structural information may be highly sensitive to processes producing geographic spatial patterns or assessing the goodness-of-fit in a spatial model describing landscape change (Robertson et al. 2014). For instance, in the analysis of rainfall patterns in Sri Lanka, Robertson et al. (2014) provide empirical evidence that the SSIM index is capable of revealing spatial structure that significantly improves the assessment of spatial interpolation models. Jones et al. (2016) study on space-use behavior of sperm whales in the Mediterranean Sea also emphasizes the SSIM capability to extract local-scale differences in space-use, providing information on spatial structure which cannot be uncovered by either pixel-by-pixel differencing or visual inspection. Additionally, Jasiewicz, Netzel, and Stepinski (2014) provide an example of quantitative measurement of landscape similarity focusing on patterns of landform elements. These studies demonstrate the potential of computer vision techniques in quantitatively-driven geographical analysis and research. We aim to build on these studies by comparing two popular image comparison methods derived from computer vision research.

The structural similarity indices, complex wavelet structural similarity (CWSSIM) index and the SSIM index, are two structurally sensitive computer vision methods for image comparison which may have utility in a spatial pattern comparison context. The indices account for potential sources of errors resulting from luminance and contrast as well as incorporating structural

information in landscape patterns. These methods have been benchmarked against their ability to mimic human judgment, and given that spatial pattern change in a geographic context is often assessed by humans in order to arrive at suitable management strategies, we foresee the metrics as potentially promising techniques for resolving complex problems in landscape change analysis and decision making processes.

In this paper, we investigate the comparison of spatial patterns, specifically in the case of continuous-valued spatial distributions on regular lattices. We extensively explore the CWSSIM and SSIM indices to examine their potential in spatial pattern comparison by discriminating maps generated as realizations of stochastic spatial processes. Further, to assess the overall generalizability of the metrics in real-world scenarios, a case study is presented where the indices are tested on snow water equivalent (SWE) maps. By comparing spatial and complex wavelet domain metrics, we hope to illuminate how these indices can be exploited in spatial pattern detection and characterization of change on continuous-valued landscape maps. With this approach, the study further sheds light on the potential sensitivity of CWSSIM index and its employment in spatial model assessment and/or validation.

3.3 Materials and methods

3.3.1 Neighborhood-based image comparison indices

Neighborhood-based indices for image data are widely developed in signal and image processing over the past decades. For example, the CDM, a geometry-based index proposed by Bowyer, Kranenburg, and Dougherty (2001), locates a possible match in a defined neighborhood for every image pixel, and ultimately returns an accumulation of matched and unmatched pixels.

The PCM is another closely related index proposed by Prieto and Allen (2003) which attempts to find pixel matches within a neighborhood of a given radius for corresponding pixels. The SSIM index proposed by Wang et al. (2004) also accounts for spatial correlations within a localized window. The SSIM index represents an extension of the universal image quality index in Wang and Bovik (2002). The SSIM index distinguishes itself from the aforementioned geometry-based indices by incorporating structural information of objects in the visual scene. Thus, aside from intensity and contrast, SSIM accounts for the attributes coding for structural information of objects in an image.

3.3.2 Theoretical concepts and configurations of the CWSSIM and SSIM indices

The SSIM index was developed to mimic the HVS and therefore predict human preferences in image quality assessment (Wang and Bovik 2002; Wang et al. 2004). The ideas motivating the SSIM formulation is that the HVS is highly adapted to extracting structural information from a visual scene. With the measurement of structural similarity and/or distortions, perceptual image quality could be effectively approximated. Although the CWSSIM configuration is much more complex, it is worth emphasizing that CWSSIM is an extended version of SSIM, representing an extension and computation of SSIM from the spatial domain into the complex wavelet domain, employing the steerable pyramid decomposition technique (Sampat et al. 2009; Li et al. 2010). A closely related variant of the CWSSIM index, the Image Quality Measure (IQM2) was proposed by combining structural similarity index and steerable pyramid wavelet transform (Dumic, Grgic, and Grgic 2014).The Dual tree wavelet transform (DTWT), proposed by Ioannidou and Karathanassi (2007), is another closely related method whose index is strictly computed via the dual tree approach in the complex wavelet transform domain.

3.3.3 SSIM index properties

In this section, we shed light on SSIM mathematical properties that confer its utility in landscape change detection compared to traditional methods. The SSIM index configuration as proposed by Wang et al. (2004), is described below. It is assumed that *x* and *y* respectively represent a local region defined by a moving window in the reference image and distorted image (i.e., image to be compared with the reference image) with identical dimensions. Using SSIM, the comparison task is decomposed into luminance, contrast and structure components. The luminance term $l(x,y)$, being slowing varying, is calculated as a function of the mean intensities (μ_x,μ_y) of image pairs. The contrast comparison $c(x,y)$ is derived from the local window standard deviations (σ_x,σ_y). The third component, structure, is estimated as the local covariance $s(\sigma_{xy})$ between *x* and *y*. This component therefore measures the degree of correlation between the map-pairs. The luminance component is given as:

$$l(x,y) = \frac{2\mu_x\mu_y + C_1}{\mu_x^2 + \mu_y^2 + C_1} \qquad \textit{Equation 3.1}$$

The contrast component is written as:

$$c(x,y) = \frac{2\sigma_x\sigma_y + C_2}{\sigma_x^2 + \sigma_y^2 + C_2} \qquad \textit{Equation 3.2}$$

Finally, the structure function appears as:

$$s(x,y) = \frac{\sigma_{xy} + C_3}{\sigma_x\sigma_y + C_3} \qquad \textit{Equation 3.3}$$

The constants C_1, C_2 and C_3 are intended to guard against the metric becoming unstable in somewhat homogeneous regions where the image local mean or standard deviation is close to

zero. The dynamic range L, of an 8-bit gray image is assumed as 255 and can be used to help determine reasonable values for the constants. The constants K_1 and K_2 are determined by Wang et al. (2004) to be far less than one; K_1 and $K_2 << 1$, $C_1 = (LK_1)^2$, $C_2 = (LK_2)^2$, and $C_3 = C_2/2$. The product of the above functions, after some basic algebra, yields the SSIM index as shown in equation 4.

$$SSIM(x,y) = \frac{(2\mu_x\mu_y+C_1)\cdot(2\sigma_{xy}+C_2)}{(\mu_x^2+\mu_y^2+C_1)\cdot(\sigma_x^2+\sigma_y^2+C_2)} \qquad \textit{Equation 3.4}$$

The SSIM index can be written in terms of measures of luminance, contrast and structure components as:

$$SSIM(x,y) = [l(x,y)]^{\alpha}[c(x,y)]^{\beta}[s(x,y)]^{\gamma} \qquad \textit{Equation 3.5}$$

In spatial pattern comparison context, the contribution of luminance, contrast and structure to spatial pattern variability in the form of landscape similarity can be controlled by users through adjusting or weighting equation 5 parameters, α, β, and γ, if there is prior knowledge on each component. Users can arbitrarily set non-zero weights for the parameters where ($\alpha > 0$, $\beta > 0$, and $\gamma > 0$). Table 3.1 shows the notations in SSIM and their meanings. In this study, the default parameter setting, $\alpha = \beta = \gamma = 1$ is adopted. This default parameter setting from Wang et al. (2004) weights the various components equally since there is no prior knowledge on the nature and extent of spatial dependence in the SWE maps. Setting all parameters to unity allows for a meaningful comparison of the metrics behavior in both simulation (i.e., known spatial structure) and real-world data (i.e., unknown spatial structure).

The maximum value of SSIM, *SSIM* (x,y) = 1 occurs if and only if x and y are identical. Note, also, that *SSIM* (x,y) is defined over a moving window (for example, 3 × 3) that is evaluated

over each local region of the image. As such, SSIM is a local metric aggregated into a global metric by computing the mean over local regions as common in computer vision applications.

In summary, SSIM operates by normalizing local patterns of pixel intensities using the underlying scene luminance and contrast terms within a defined local window. This feature can be leveraged to generate somewhat objective pattern comparison results since multitemporal maps may be contaminated with signals that are not representative of true landscape pattern change but originate from different acquisition times as well as the use of disparate data abstraction devices and methods. Such an attribute can also be exploited to significantly mitigate map preprocessing errors that impact pixel intensity values by altering image luminance and contrast.

3.3.4 CWSSIM index properties

The CWSSIM index is proposed based on the motivation that the phase pattern (i.e., the spatial configuration) of natural image signals contains more profound structural information than their magnitude component (Oppenheim and Lim 1981). The index therefore attempts to differentiate magnitude and phase shifts of a pair of signals in detecting and characterizing dissimilarity between image-reference pairs. The structural component of natural image signals is critical to detecting pattern changes in paired maps, and this, in a landscape context, could capture underlying processes generating a particular landscape pattern. The magnitude component is closely associated with luminance as in the case of SSIM, and is ultimately scaled out. Thus, potential differences originating from illumination (e.g., sun angle, shade etc.) can be reduced in landscape pattern similarity analysis. This is analogous to the use of band ratios instead of raw channel digital numbers in analysis of remotely sensed image data.

Given that the CWSSIM computation contains intensive mathematics, here we outline only the properties of the CWSSIM that relate to its potential utility as a spatial pattern comparison metric. Unlike the SSIM, for reference map and corresponding map whose similarity is to be evaluated, CWSSIM employs low-pass filtering mechanism to decompose the map-pairs into multiple channels (sub-bands). Successive filtering and down-sampling results in sub-bands with each level having a smaller dimension than the preceding channel (i.e., forming a wavelet pyramid). Ultimately, comparison is made between corresponding pairs of sub-bands or multiple frames. The comparison takes the form of measurement of the correlation in signal phase and magnitude between image-pairs, with emphasis on the phase component as it encodes relevant structural information. Extensive mathematical details on the metric can be found in (Sampat et al. 2009). Table 3.1 depicts the various notations in CWSSIM and their meanings. The CWSSIM index is given as:

$$CWSSIM(x,y) = \frac{2|\sum_{i=1}^{n} c_{x,i}c_{y,i}^{*}|+K}{\sum_{i=1}^{n}|c_{x,i}|^2+\sum_{i=1}^{n}|c_{y,i}|^2+K} \qquad \textit{Equation 3.6}$$

where c_x and c_y denote extracted or filtered (sub-channels or blurred-down-sampled versions) of the input image pairs (see Figure 3.1). That is, the CWSSIM operation initializes by simultaneously decomposing input maps x and y into sub-bands (sub-channels) c_x and c_y using the steerable pyramid decomposition technique as in (Sampat et al. 2009). Figure 3.1 illustrates input image decomposition process (i.e., for one input map, x) to generate three sub-bands (low pass band1, low pass band2, and low pass band3). These sub-bands (channels), c_x and c_y from input maps x and y are aligned and CWSSIM computed using a local moving window as in SSIM. The sub-bands correspond to samples of image pairs at the same spatial location. The local sub-band pixel values for an image-reference pair is denoted by: $c_x = \{c_{x,i} | i = 1,...,n\}$ and

$c_y = \{c_{y,i} | i = 1,...,n\}$. n represents the number of pixels defined by the local moving window. For a 3 × 3 window, $n = 9$. The window slides across each sub-band to compute the CWSSIM indices. Thus, the local neighborhood of each sub-band (wavelet) that contributes to the construction of the pyramid levels (hierarchical sequence of extracted sub-bands) is sampled. The K and c^* terms denote stabilizing constant and complex conjugate, respectively. K improves the robustness of CWSSIM estimates under low local signal to noise ratios.

Luminance and contrast changes are approximated in the wavelet domain as a pair-wise transformation of local pixel values, resulting in constant scaling of all sub-bands. This ensures that the metric is insensitive to luminance and contrast variations between structurally dissimilar landscape maps or patterns being compared, hence, the metric can potentially minimize differences related to these sources in spatial pattern comparison task. In short, the metric normalizes radiometrically dissimilar multisensor or multitemporal maps leaving only the structural differences, crucial information in detecting and quantifying landscape pattern change.

The translation term corresponds to a linear phase shift (horizontal and vertical shifts) in the Fourier domain. This shift is approximated assuming that translation is small relative to the size of the wavelet filter, and under slowly varying wavelet filter envelop. The scaling and rotation terms are approximated using the translation term, assuming that the image pairs are subjected to small scaling and rotation. The scale and translation invariance attribute of the metric does not only confer the advantage to output consistent similarity values at varying scale shifts in landscape pattern change detection, but can effectively reduce map-pair differences originating from geometric sources such as movement of image acquisition devices and sensor-solar orientations, co-registration discrepancies and related georeferencing mismatches. The illustration in Figure 3.1 shows a conceptual overview of the CWSSIM index operation.

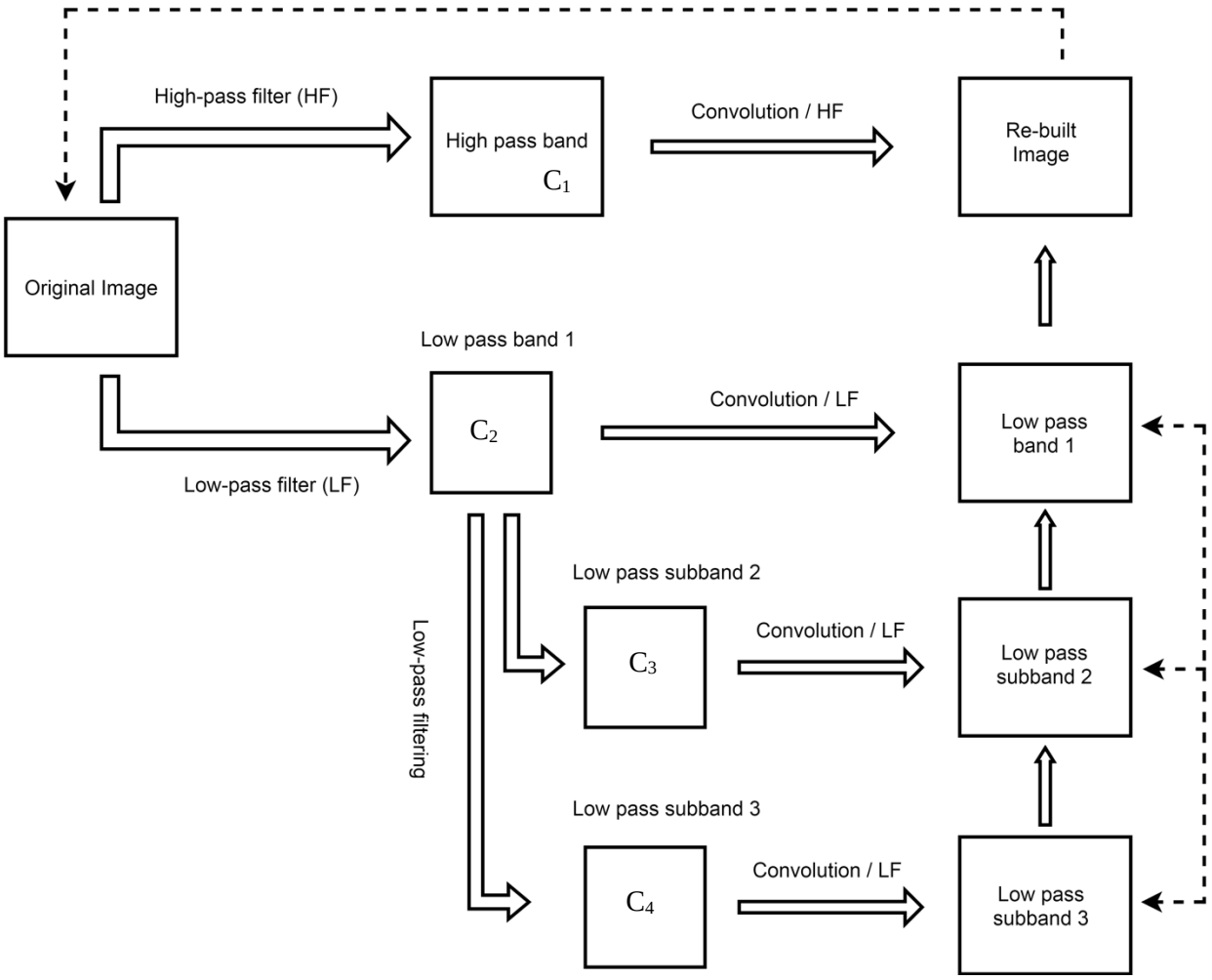


Figure 3.1 Schematics of CWSSIM index operation. This illustration depicts how the CWSSIM index operates on images. Each of the images being compared is simultaneously decomposed using the complex steerable pyramid decomposition method, yielding high pass and low pass bands; the filters are translation and rotation invariant. The CWSSIM index values are computed locally using a moving window across each of the wavelet sub-bands.

A graphical example of SSIM and CWSSIM is provided in Figure 3.2 which compares sub-images (a) and (c) to image (b) (all sampled from the larger image (d)). The cropped samples are selected such that the dominance of both land and water pixels create a discernible geometry to

visualize the metrics' judgement of similarity as the coastline shifts. Sub-image (a) can be considered an *across-landscape comparison* which represents an area up the coast from (b) the reference image, while sub-image (c) is shifted north of (b) due to a NAD27/NAD83 datum shift. Note that the bounding box for sub-image (b) and (c) are identical. In each case we might consider the landscapes similar; yet traditional point-based or even window-based comparison metrics would be limited in recognizing similarity in this case. Here, CWSSIM scores (a-b: CWSSIM = 0.514, c-b: CWSSIM = 0.571) are higher than SSIM scores for (a-b: SSIM = 0.411) and (c-b: SSIM = 0.460). This demonstrates the geometric invariance properties of the CWSSIM.

To summarize, the CWSSIM differentiates the measurement of magnitude and phase shifts, has greater sensitivity to phase (i.e., spatial pattern structural change) than magnitude change, and is insensitive to consistent relative phase shifts. The CWSSIM geometric invariance attribute could enable an effective mitigation of errors emanating from small shifts in data sampling equipment.

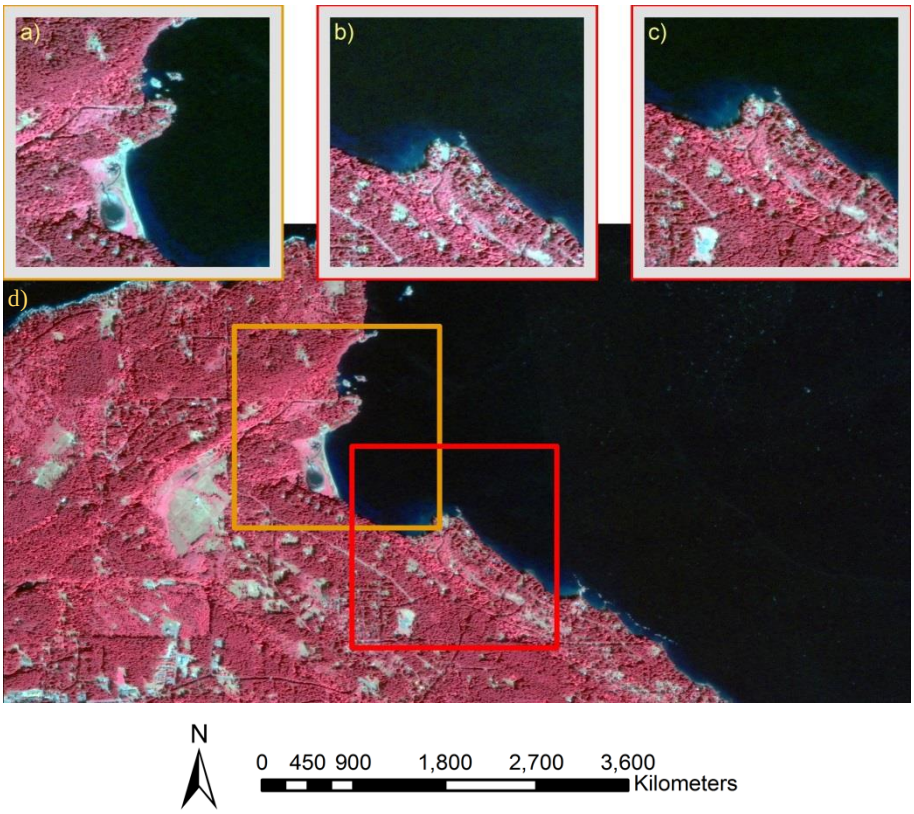


Figure 3.2 Sample landscape comparison of SSIM and CWSSIM. Multiple images for near-infrared false colour composite imagery of coastline on Gabriola Island, British Columbia, Canada. Sub-image (a) location corresponds to the orange bounding box, while sub-image (c) is located at the red bounding box; note that sub-image (b) bounding box is identical to sub-image (c), but data is different due to NAD27/NAD83 datum shift north of (b).

We explore these features of these metrics in the following analyses in order to better understand their potential in map or spatial pattern comparison, change detection, model assessment, and several other geographical applications as evidenced in Figure 3.2.

3.3.5 Simulating Spatial Patterns

Realizations of spatial stochastic processes were used to explore indices in a controlled setting. Moving average fields were simulated as their distinct patchy nature will allow for visual inspection of the extent of dissimilarities between paired images. Two separate sets of images were simulated on a 50 × 50 grid. Though spatial patterns as well as the outcomes of metrics computed on them are largely a function of spatial grain and extent (Gustafson 1998; Comber and Wulder 2019), the selected grid dimension considered here is useful for analyzing patterns characterized by small-scale moving average fields and ensures relatively simple patterns. Moreover, patterns in moving average fields can be easily detected by the HVS due to smooth nature of the spatial structure (Dungan et al. 2002). A key parameter which can be interpreted as the scale of spatial dependence in the underlying spatial process, the radius r, was set at $r = 3$ and 8, for each of the simulations. Note that r can be interpreted as the range (i.e., expressed in cells) of spatial dependence. Simulations were conducted in MATLAB programming software as shown in (Kroese and Botev 2015). For each simulation setting, 100 images were selected for matrix-wise comparison. Overall, 10,000 pair-wise comparisons were made. Throughout the succeeding sections of this paper, the two simulations will be referred to as mvr3 and mvr8, for simulations on radii $r = 3$ and 8, respectively. We use the prefix ‘mv’ to label series of simulated maps (e.g., mv314).

3.3.6 Parameterizing CWSSIM and SSIM

The CWSSIM and SSIM indices are expected to perform differently given their properties as described in the previous section. CWSSIM has advantages over SSIM owing to its translational

and scale invariant attributes; depending on parameter settings, it could incorporate 6 pyramid levels in the wavelet domain, a property which might appear like computing SSIM 6× across the same image-reference pairs. Another potentially influential factor on both indices' performance is the extent of complexity inherent in spatial patterns realized for simulations and SWE data structure, texture and edge intensity. SWE data and simulation realizations have limited complexity as both represent smoothly varying spatial structure.

The simulated and SWE maps were compared using CWSSIM and SSIM. In order to investigate their sensitivity to window size, three window parameterizations were adopted; 3 × 3, 6 × 6 and 11 × 11. In Wang and Bovik (2002), 8 × 8 window size is adapted for SSIM but was observed to yield unnecessary "blocking" artifacts, while in Wang et al. (2004) the use of 11 × 11 gaussian symmetric window overcomes this limitation. The variance and structure measures of SSIM were found to be sensitive to window sizes, 3 × 3, 5 × 5, and 7 × 7 (Jones et al. 2016). In CWSSIM, 7 × 7 window produced relatively high similarity values Wang et al. (2004). Jasiewicz, Netzel, and Stepinski (2014) employ an adaptive window parameter for retrieving similar features between landscape pairs. Our 3×3 window setting aligns well with mvr3 images, and the idea here is to intentionally inspect SSIM performance and sensitivity at known spatial relationships and window dimensions. Here, we intend to underline the metrics' sensitivity taking into consideration the extent of spatial dependence among features in the visual scene. This we hope, could represent an important starting point in fine-tuning window parameters to any task at hand.

For CWSSIM, default settings were used for other parameters; the stabilizing term, K was set to zero ($K = 0$); guardb (gd), the edge size to be discarded from the four image boundaries, was also set as gd = 0, the number of orientations (ori) = 16, and number of levels (level) = 6. The

utilization of these default parameter settings is motivated by evidence that as pyramid levels increase, and holding ori = 16, maximum CWSSIM index value is realized as the index becomes more resilient to translations and rotations at advanced levels of pyramid decomposition (Sampat et al. 2009). This will therefore allow for exploring the full capacity of the metric. For implementation, SSIM constants, K_1 and K_2 were adjusted to 0.01 and 0.03, respectively as suggested in Wang et al. (2004). The analysis of SWE maps and simulated data was executed by running SSIM and CWSSIM indices implementation codes using MATLAB.

3.3.7 Case Study: Mapping snow water equivalent change in northern Canada

In order to extend the application of the metrics from simulation scenarios to empirical data, we tested SSIM and CWSSIM on SWE maps. The window parameter settings implemented in simulation comparisons were employed in the SWE data. Sample of SWE maps were compared at 3 × 3, 6 × 6 and 11×11 window sizes. The SWE data are coarse resolution ($25km \times 25km$ pixels) covering parts of the Canadian Arctic with temporal resolution stretching from 1979 to 2017 at daily intervals. The maps used were clipped out of the global SWE data archive produced by the Globesnow-2 project, which combines passive microwave remote sensing with extensive field data to parameterize daily SWE data over the northern hemisphere (Takala et al. 2011). In many remote polar regions where climate change is greatly impacting ecosystems, remotely sensed SWE data is the only data available for understanding landscape scale patterns of snow and related hydrological processes and functions. As such, understanding changes in spatial structure of SWE over large regions is an important information. For the purpose of this practical demonstration, 1980 and 2013 datasets were chosen. These datasets were found to contain sufficient maps at desired temporal resolution covering the months in winter when snow

abundance and spatial distribution was sufficient for testing the metrics performance. Within each year, comparisons were made on a monthly basis to account for seasonal variability in snow and water dynamics. Ten images were selected in each particular month of the year and compared with the other year's imagery, yielding approximately 100 pair-wise matching in each month's dataset. To rank SWE maps based on perceived similarity to a given reference map on SSIM and CWSSIM indices, Wilcoxon rank tests was performed using R programming software.

3.3.8 Metric Comparison Analysis

Similarity values for different simulations and comparisons were conducted using rank test for statistical significance between ranks. Significance tests were carried out on similarity values using Wilcoxon signed rank test. Hypothesis tests were performed in all scenarios of CWSSIM and SSIM comparisons using variable window sizes. Comparisons were relative rather than dependent on absolute metric values. The approach also renders it feasible to compare agreement and/or discrepancy in SSIM and CWSSIM judgment as to which image-reference pairs are most similar.

3.4 Results

In this section, we report the results of our experiments. For each image pair comparison, we compute the MSE score to compare with the structurally sensitive metrics. It is worth noting that the MSE is a widely deployed classical metric which captures the difference in error signals between paired images but not the visually perceptual structural components.

3.4.1 Similarity index values distribution

Figure 3.3 depicts the distribution of similarity values for the simulation realizations under different spatial dependence parameters. In general, as the window size increases, the metrics generate lower similarity values. As well when window dimension is aligned with spatial scale, similarity is higher, but when it is mis-specified and/or larger, similarity values are lower and more dispersed. SSIM values are notably higher than that of CWSSIM. In Figure 3.3 (a), cross-simulation (i.e., cross spatial process) comparison is denoted by mvr3_mvr8_w6 (i.e., mvr3 versus mvr8 comparison using a window with size 6), and mvr3_w3 denotes within simulation (i.e., within spatial process) comparison (i.e., mvr3 versus mvr3 using a window with size 3). It can be seen from Figure 3.3 (a) that CWSSIM values are more dispersed around the median while SSIM values tend to exhibit less variability. A plot of the metrics variances, Figure 3.3 (b) further illuminates on the inherent variability in the indices estimates of similarity with changing window size. SSIM scores remain consistently higher than CWSSIM; these values point to the extent of penalization of image dissimilarities by the metrics. Interestingly, however, as can be seen from Figure 3.3 (a) there appears to be little overlap between SSIM and CWSSIM; this is probably driven by higher variance characteristic of CWSSIM values. While lower similarity scores are expected for cross-simulation comparison, SSIM generated higher values than did CWSSIM. This suggests that with SSIM, the cross-process comparison with a slightly mis-specified window may be estimated more similar than a within-process comparison with a poorly defined window-size as seen in the within-process (mvr3_w11) comparison (see Figure 3a). It is important to emphasize that the Wilcoxon signed rank test generated $p < 0.05$ (95% confidence interval) in all test scenarios; thus, the null hypothesis that SSIM and CWSSIM values are not significantly different is rejected throughout the analysis. This rendered the deployment of ranks

values in comparing agreement and discrepancy in the metrics judgment relevant as the index values themselves are not directly comparable.

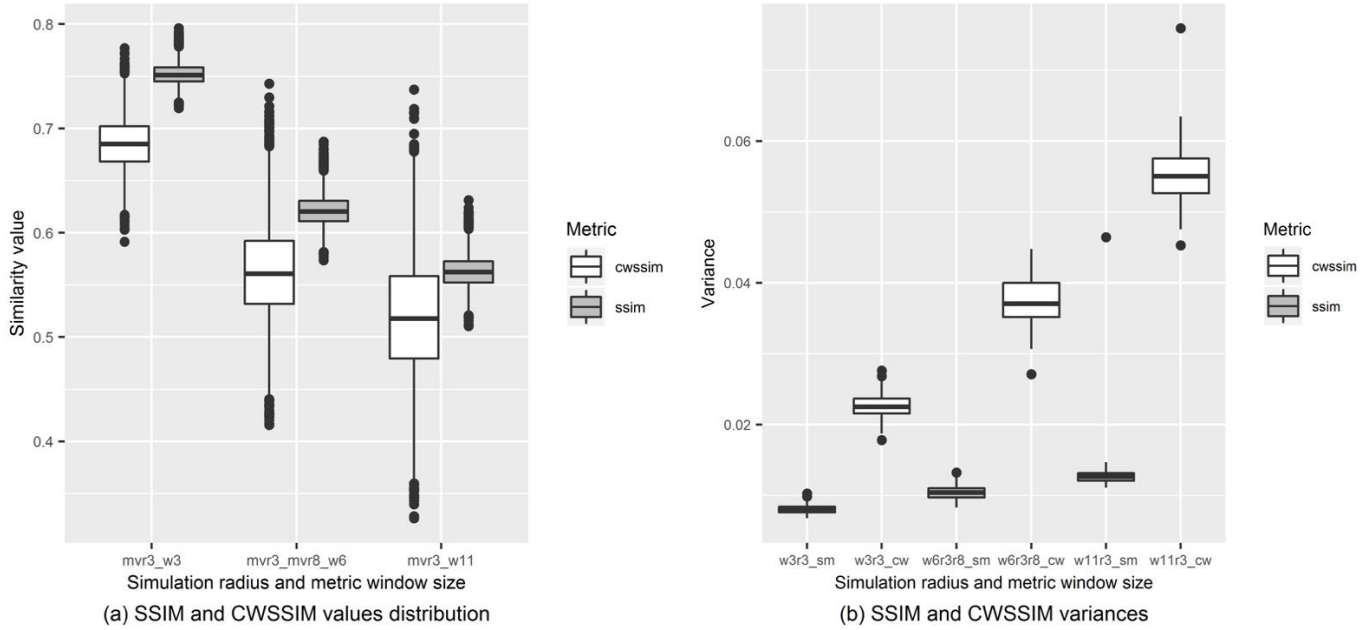


Figure 3.3 Similarity values and variance distributions. As observed from (a), SSIM values are generally higher than CWSSIM, and separable due to low variability. Conversely, CWSSIM scores exhibit lower values with higher dispersion than SSIM.

3.4.2 SWE similarity maps and similarity values distribution

The distributions of similarity values in the SWE datasets are summarized in Figure 3.4. SSIM values are generally higher and separable from CWSSIM, except for February and April where overlaps can be seen in the box plots. As observed in simulation scenario, CWSSIM has greater variability than SSIM, while SSIM values remain generally higher. The February SWE map

similarity in Figure 3.4b shows SSIM to yield higher similarity values when comparison was made using window 3 × 3. SSIM index values however appear to decrease as window size increases, while that of CWSSIM remained virtually constant. It can be seen that the median value for SSIM and CWSSIM are almost identical at window size 6 × 6. Thus, at larger window size, SSIM indicates dissimilarity compared to other months, while CWSSIM does not. Figure 3.5 depicts SWE images captured on different dates in February 1980 and 2013 – February 2nd 1980 (a: FEB800202), February 1st 2013 (b: FEB130201), February 10th 1980 (c: FEB800210), and February 10th 2013(d: FEB130210). We observe inverse patterns in snow distribution, with low SWE in 1980 in the southwest of the images, and higher SWE in 1980 in the northeast region of the images. We further notice an inverse trend in snow distribution in 2013 images, this time, lower SWE in 2013 in the northeast region, and higher snow in the southwest region.

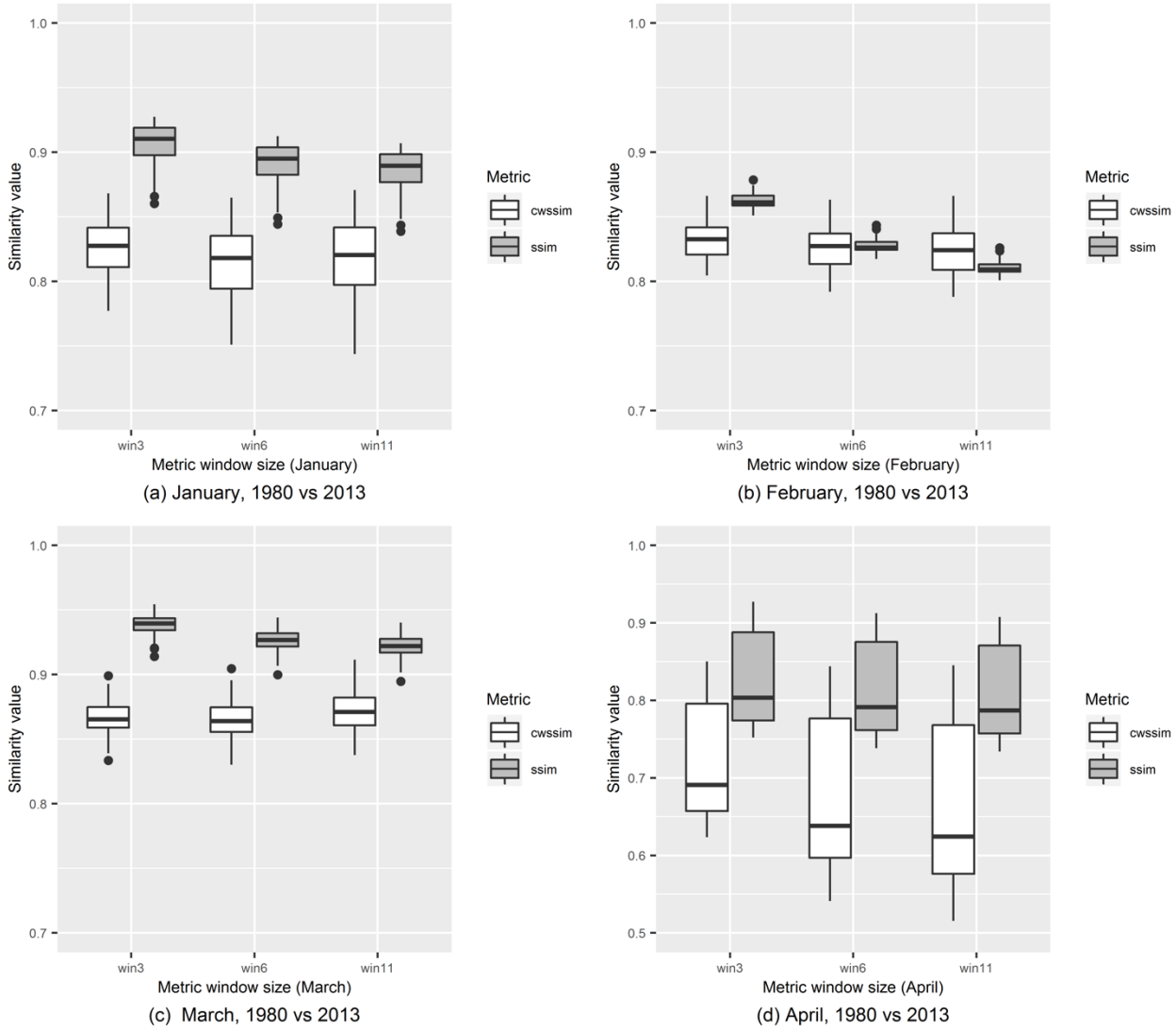


Figure 3.4 Box plots of SSIM and CWSSIM values distribution. Comparison is monthly or seasonal based (a) January 1980 versus January 2013; (b) February 1980 versus February 2013, (c) March 1980 versus March 2013, and (d) April 1980 versus April 2013). The difference between the sample means and variances is statistically significant ($p < 0.05$).

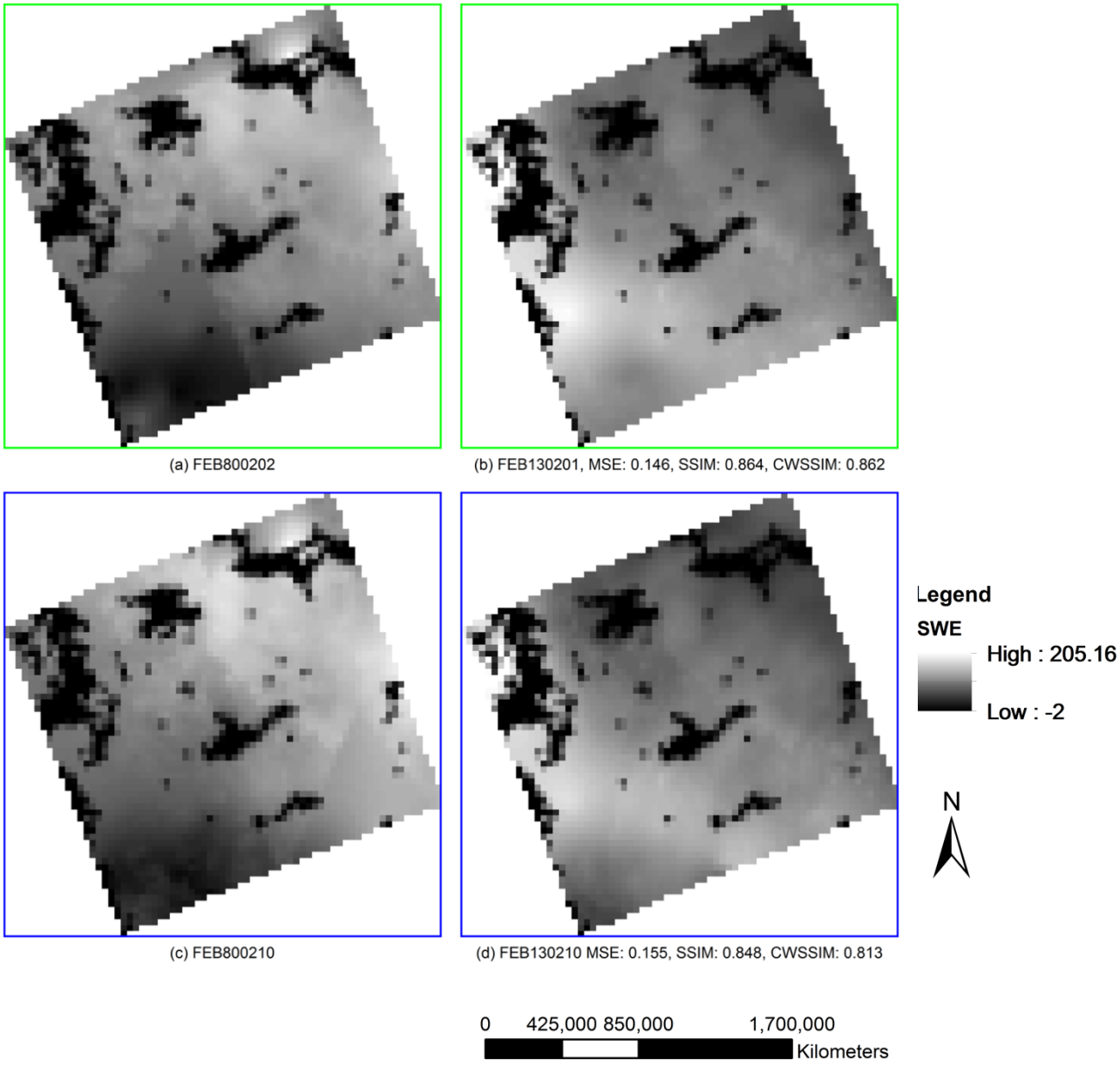
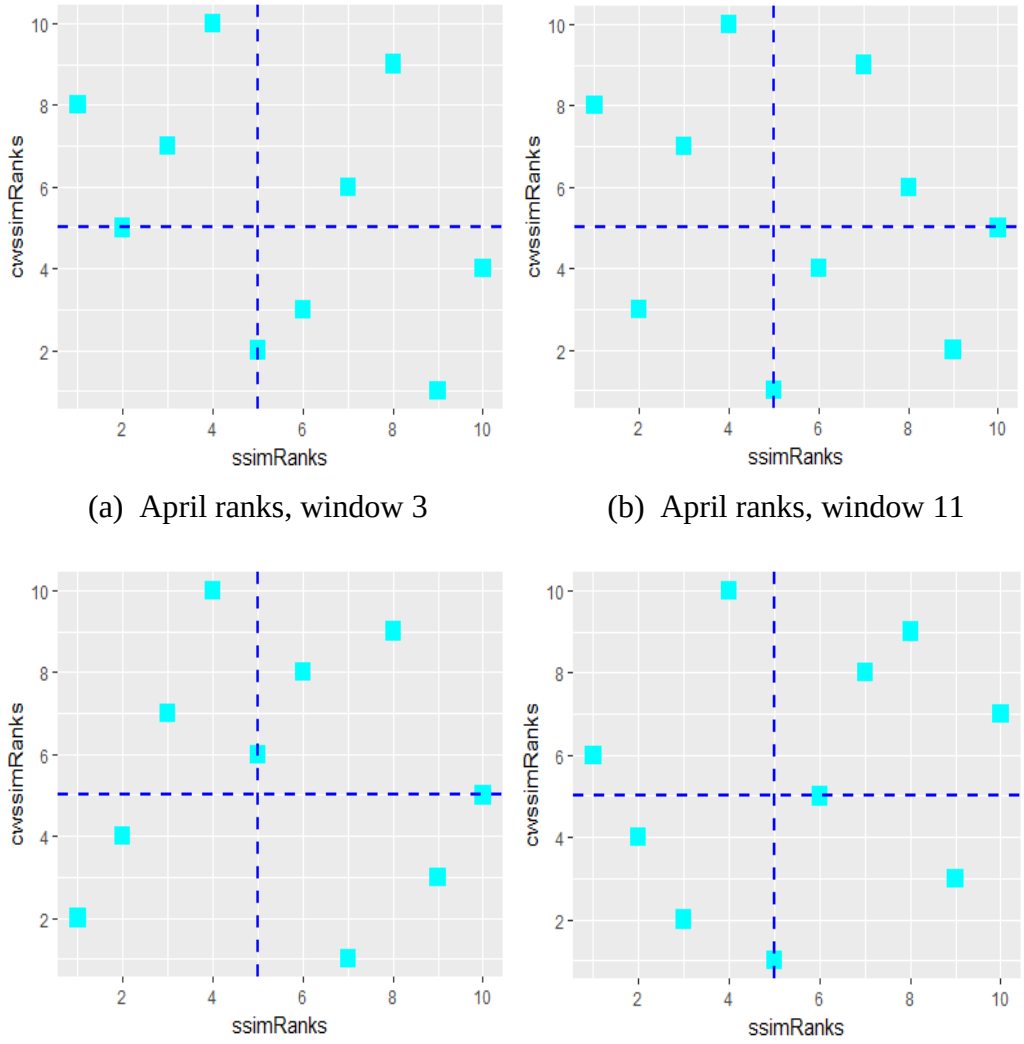


Figure 3.5 SWE distribution in February 1980 and 2013. As can be seen from the images, there is higher SWE in northeast compared to the southeast in 1980 for both (a) and (c). Compared with the 1980 images, the 2013 images (b) and (d), appear to depict an inverse trend – higher SWE distribution in the southeast. Similarity values are derived from comparing (a-b), and (c-d), where (a) and (c) are reference images.

In Figure 3.6, it can be noticed that increasing window size results in map ranks shift, especially, for SSIM relative to CWSSIM. This demonstrates the indices sensitivity to varying window dimensions. For CWSSIM, we observe some consistency in image ranks in the first and fourth quadrants of Figure 6(a) and 6(b). While it is not certain about CWSSIM behavior with differing window sizes, it appears clear that SSIM is relatively sensitive to changing window configurations. We also note that there are clear rank discrepancies between the metrics across all scenarios.



(a) April ranks, window 3

(b) April ranks, window 11

(c) February ranks, window 3

(d) February ranks, window 11

Figure 3.6 Plot of SSIM and CWSSIM ranks. (a) and (b) depict April SSIM and CWSSIM ranks, (c) and (d) are SSIM and CWSSIM ranks for February. Each square box on the graph represents an image (i.e., SWE map). We arbitrarily split ranks distributions into four quadrants to demonstrate agreement and discrepancy on SSIM and CWSSIM indices.

In Figures 3.7-3.8 red, green and blue frames respectively represent reference image, metrics agreement and discrepancy in ranks. Unlike the simulation case, in the SWE dataset, comparison and ranking employing window size 3 and 6 yielded repetitive maps and ranks, thus we focus on window dimensions 3 and 11 instead. Figure 3.7 shows samples of SWE maps from February 1980 (b) and February 2013 (a and c). SSIM and CWSSIM indices show discrepancy in assigning ranks to map FEB13101. In contrast, the indices ranking agree for FEB1304. As noticed previously, limited variability exists in the metrics values on the same index using window settings 3; cross index values are however different. MSE estimates are opposite and disagree with both metrics' similarity assignments.

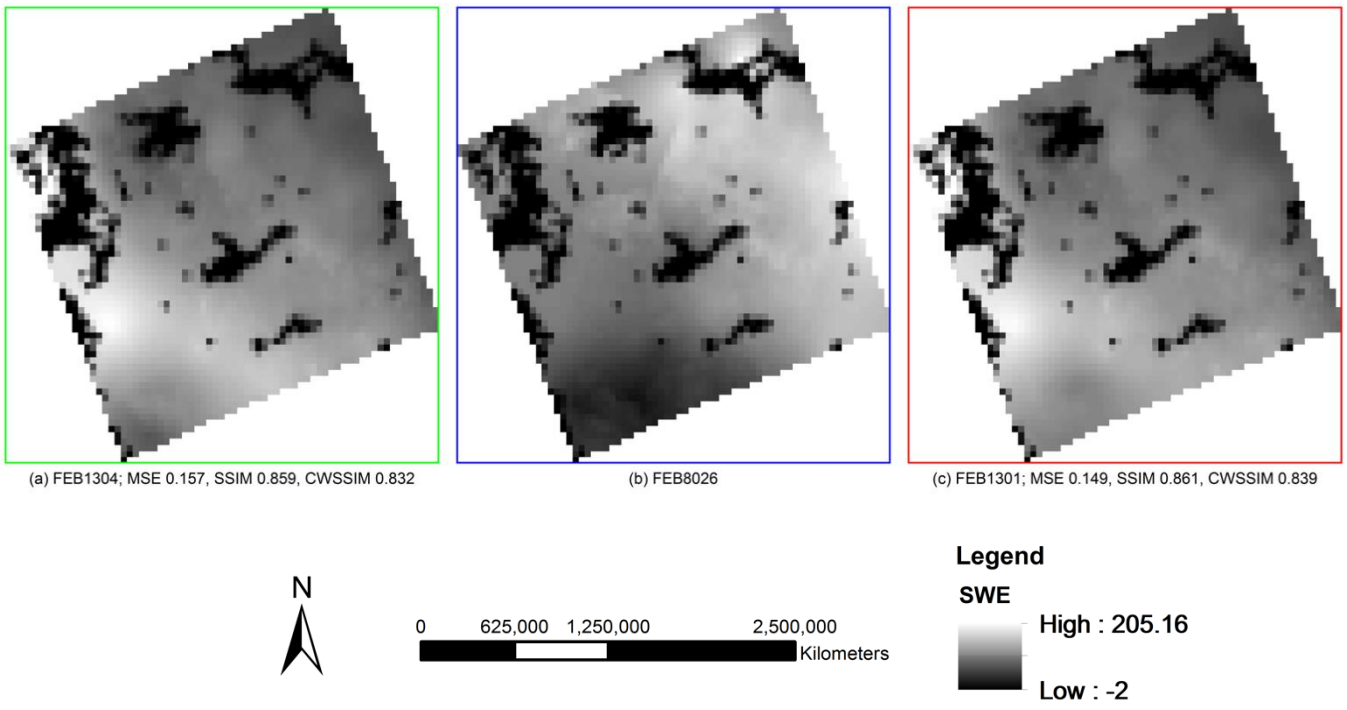


Figure 3.7 Sample of SWE maps in February 1980 and 2013. (b) is reference image, (a) and (c) are SWE maps taken on different dates. SSIM and CWSSIM show discrepancy in assigning ranks to map (c) FEB13101. Conversely, the indices ranking agrees for (a) FEB1304.

Similarity maps in April SWE dataset are shown in Figure 3.8 using window size 3. The ranks of SSIM and CWSSIM match for APR1301; also, MSE estimates align with both metrics. Again, the closeness of the metrics values on the same scale is emphasized.

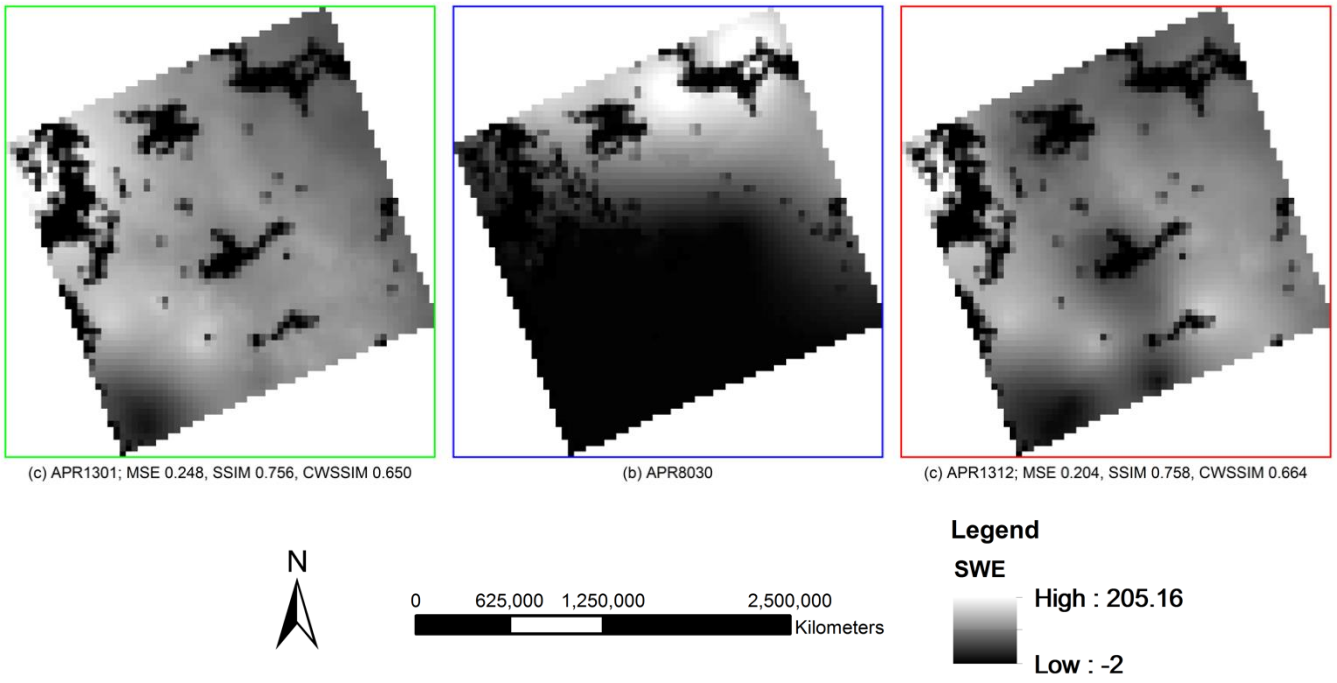


Figure 3.8 SWE maps in April 1980 and 2013. (b) is reference image, (a) and (c) are SWE maps on different dates. Window size 3 was maintained for similarity measurements. Ranks of SSIM and CWSSIM turn out to match for (a) APR1301.

Figures 3.9 and 3.10 depict surface snow cover dynamics from January to April in the year 2013. The reference image was capture on 2nd January, 2013. Window size 3 × 3 was used to compute SSIM values. Although SSIM index shows fluctuating snow levels through figures 3.9(b) to 3.9(h) probably due to daily snow variability, a consistent decrease in snow level is obvious from Figures 3.9(i) to 3.9(l). The snow melting process is further emphasized in figure

10 where both CWSSIM index and SSIM index exhibit apparent decline in surface snow cover in 2013, from first April through to twenty fifth April.

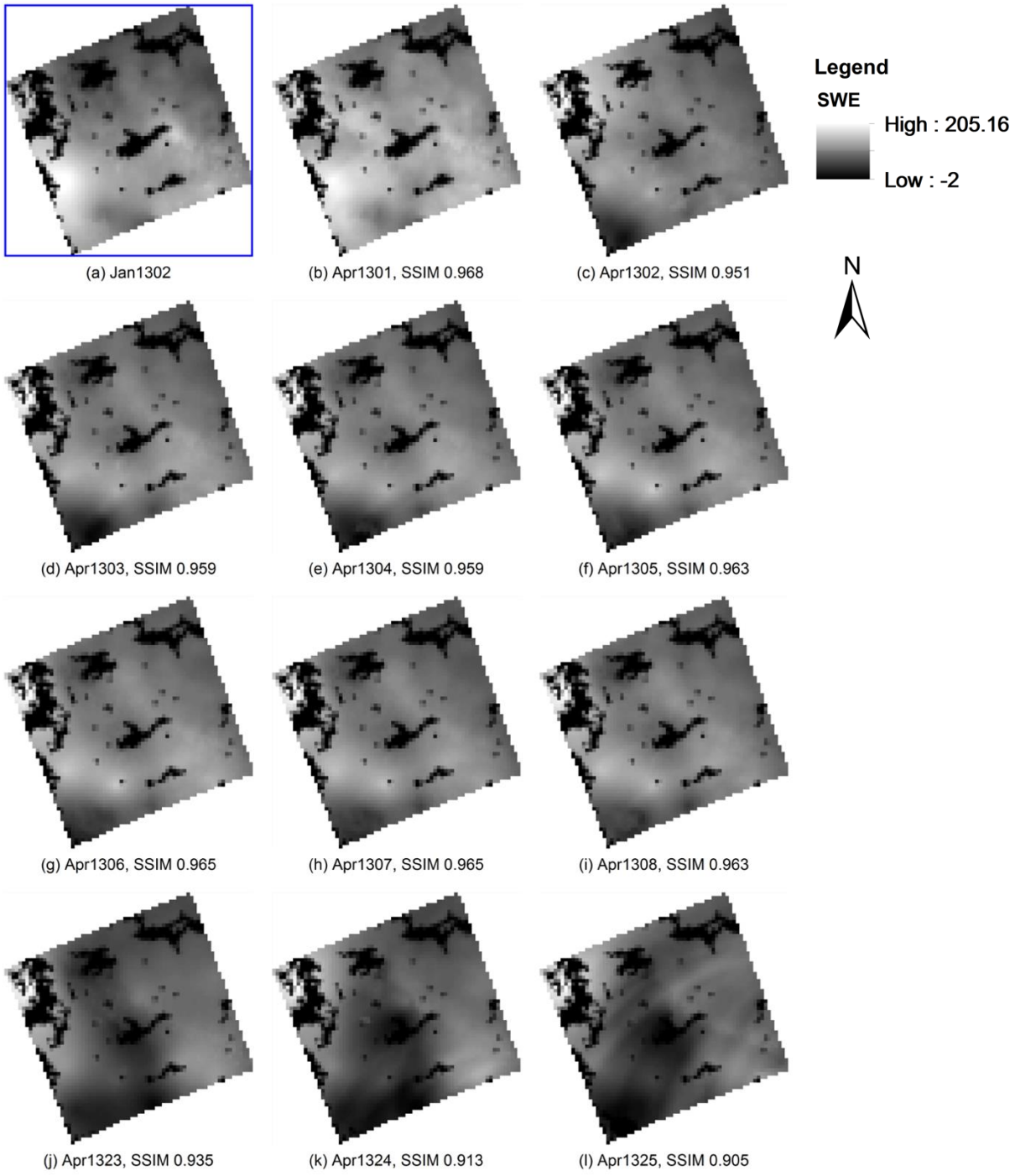


Figure 3.9 Comparison of January SWE map and April SWE maps. (a) January SWE map, and (b) - (l) are April SWE maps. SWE map Jan1302 represents the reference image. A gradual decline in similarity values can be noticed from the beginning towards the end of April.

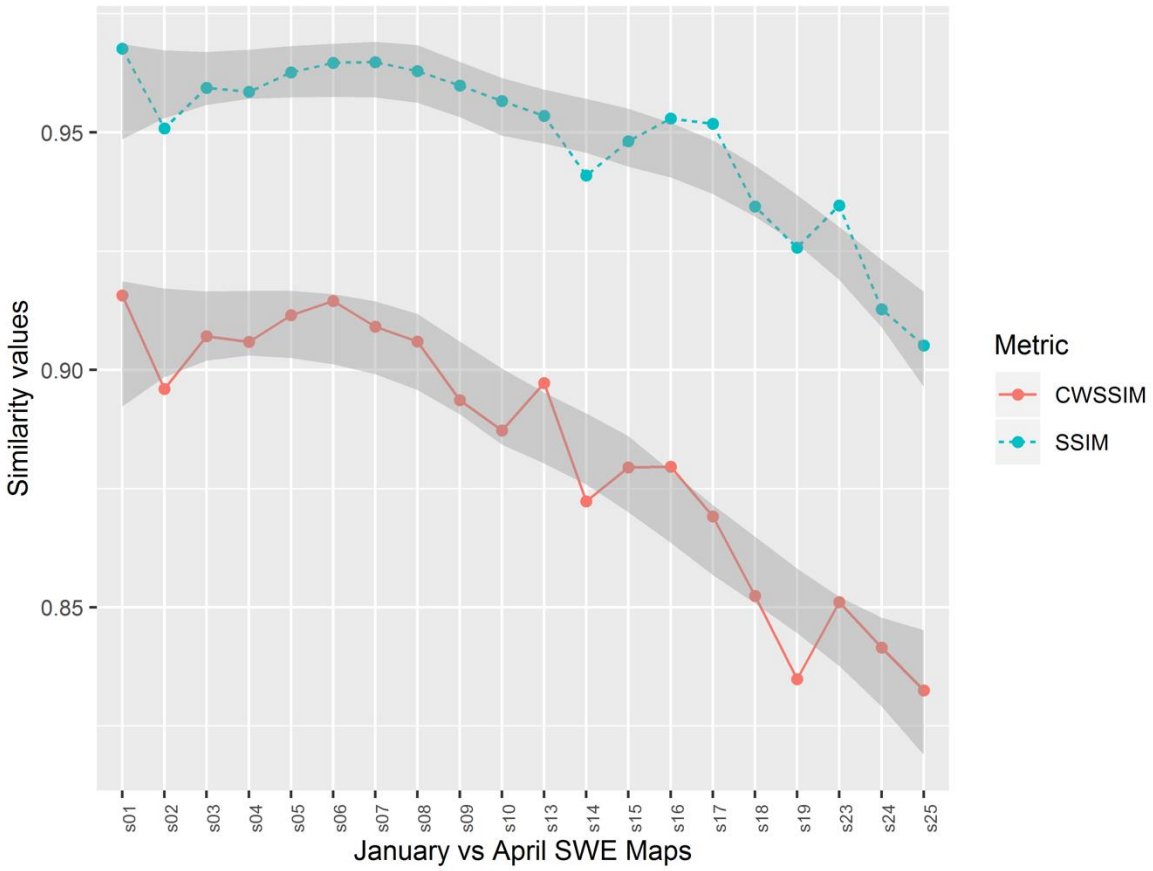


Figure 3.10 Trend in snow melting process in northern Canada, April 2013. On the horizontal axis, labels s01 to s25 represent samples of SWE maps in April, and the corresponding similarity scores on the vertical axis. Note that here, comparison is made using January SWE map as reference. Overall, the plot depicts a closely related trend in similarity values recorded by the metrics, though SSIM values are higher than CWSSIM.

3.5 Discussion

Using Wilcoxon signed rank test, similarity estimates on SSIM and CWSSIM indices were found to be statistically significant for within simulation (mvr3 versus mvr3) and between simulation (mvr3 versus mvr8). For each simulation, the test was significant when windows

3 × 3, 6 × 6, and 11 × 11 were used. It was only in one scenario that there was no statistical significance in mvr3 versus mvr3 using window size 11 × 11 (i.e., p-value = 0.793). Wilcoxon tests for most image-reference pairs yielded p-value < 0.05 for our simulation data. When ranks of image-reference pairs on both indices were compared, patterns of both agreements and discrepancies were observed. There were more frequent agreements in mvr3 versus mvr3 with window size set to 11 × 11 than in mvr3 versus mvr3, and mvr3 versus mvr8, using window 3 × 3 and 6 × 6, respectively.

For CWSSIM, the variance pattern shown in Figure 3.3b is expected; being lowest for mvr3 versus mvr3 compared to comparisons where process parameters were different. The differences in variance for SSIM were less pronounced. This exposes SSIM's lower power to detect subtle changes in spatial patterns or spatial pattern generating processes and perhaps emphasizes CWSSIM insensitivity to non-structured changes (Sampat et al. 2009), a feature of possible utility in spatial pattern comparison (e.g., see Figure 3.2). In the cross-process comparison of mvr3 versus mvr8 images, the metrics' similarity scores reduced substantially below mvr3 versus mvr3 comparison on window 3 × 3 but remained higher on that same data with window 11×11. Given that there is higher spatial dependence in mvr8, we anticipate lower similarity values for mvr3 versus mvr8 than in mvr3 versus mvr3 at the same window dimensions. However, as depicted in Figure 3.3a, this turns out to be true on the smallest window 3 × 3 and not on the largest window 11 × 11. Such an observation points to the effects of spatial smoothing and loss of local spatial information on the degree of spatial dependence existing in the map-pairs as the window size increases. Thus, we notice that within simulation (i.e., similar underlying geographical process), there occurred decreased similarity values, but with increased image-pair rank agreement on both indices at window size 11 × 11. This sheds light on the

overall importance of window dimensions on the sensitivity of the metrics and emphasizes the importance of window choice and scale in geographical pattern comparison. An inappropriate window size is likely to miss the real variability in the underlying pattern structure. These inferences further highlight the widely debated issue of scale in geographical research. Therefore, to properly quantify and characterize spatial patterns using these metrics, an appropriate window choice should be adopted vis-a-vis prior knowledge pertaining to scale and extent of spatial dependence in processes generating spatial patterns (Comber and Wulder 2019).

Jones et al. (2016) research on sperm whales space use behavior in the Mediterranean Sea demonstrates the influence of window size on the various components of SSIM index. The authors found that varying window size from 3 × 3 to 5 × 5 showed noticeable differences in variance (contrast) and covariance (spatial structure) components in the maps being compared. Similarly, in our experiments with varying window sizes, the metrics yielded different similarity values. In moving from window size 3×3 to 11×11, sharp changes in map-pair similarity values as well as rank shifts occurred compared to traversing from window dimension 3 × 3 to 6 × 6. To overcome window selection issues in geographical research, Wu (2004) suggests that the choice of local neighborhood size (i.e., window size) should include the spatial resolution of the data or maps being compared, ecological questions intended to be answered via map comparison, and possible scale-dependence in the pattern generating processes.

In remote sensing data acquisition, sun angle change, uneven topography, sensor platform shift and/or angle displacement are common problems requiring image registration and radiometric correction (Dai and Khorram 1998; Rogan and Chen 2004). The CWSSIM is inherently translation, rotation, and scale insensitive to marginal variations between map-pairs (Gao, Rehman, and Wang 2011; Sampat et al. 2009); such attributes render the metric robust to

the non-structured distortions mentioned above. We use real-world maps to emphasize the geometric insensitive attribute of CWSSIM in Figure 3.2. Higher variance in CWSSIM values observed in our analysis, for example, demonstrates its inherent translational, rotational, and scale invariance property; thus, the index can potentially capture real and detailed structural changes between pairs of landscape maps without compromising its performance. This implies CWSSIM has the tendency to characterize and quantify differences even in patterns generated by closely related but somewhat diverging geographical processes. Contrarily, the SSIM index yielded relatively small standard deviations, suggesting that the image-reference pairs are produced by similar spatial pattern-generating processes which are probably difficult to isolate. Given that similar process can create distinct spatial patterns, differences in spatial patterns resulting from the same spatial processes may not be well characterized or observed using SSIM. In geographical analysis, especially comparing continuous-valued maps, image mis-registration can lead to false accuracy reports; the rotation and translation invariance attribute of CWSSIM can be useful for mitigating such errors. Other studies such as (Gao, Rehman, and Wang 2011) and (Rehman et al. 2013) emphasize CWSSIM's potential to discriminate and classify different image sets.

The SWE maps presented as case studies provide an insight into the potential deployment of computer vision methods in quantifying real geographical data. Wilcoxon signed rank tests in SWE data produced p-value < 0.05 for most samples, implying significant difference in the metrics evaluation of map-pair similarity. Figures 3.4 and 3.6 depict SSIM and CWSSIM values distribution and map ranks pattern change, respectively in SWE data. Shifts in ranks appeared less in the real geographical data (i.e., SWE data) which could be attributed to the relatively smooth transitional nature of snow patterns. Rank discrepancy tends to occur frequently where

the metrics values were close with less variability. As can be seen in figures 3.7, SSIM values are higher at small window size, suggesting that the metric is sensitive to local patterns in SWE variability. This property is useful for characterizing small scale changes in SWE. CWSSIM's translation and rotational invariance properties are probably smoothing or averaging out local variability; thus, accounting for the unchanged SWE distribution which were captured by SSIM in Figure 3.5. This implies CWSSIM may be less sensitive to local changes in SWE, but can be utilized to characterize structural changes in pattern rather than local changes in SWE maps. These observations, in all, highlight the metrics complimentary functionality and the potential to deploy them together to characterize changes in spatial patterns from local to large scale. In terms of snow spatio-temporal change, the metrics portray low snow levels in February and April 1980 compared to the amounts in 2013. Figures 3.9 and 3.10 illustrate temporal changes in snow pattern throughout April 2013. Substantial snow melting occurred in the last week (i.e., 23rd to 25th April). Such subtle spatio-temporal changes in climatic processes can be important indicators in climate trend analysis but cannot be effectively revealed by traditional map/image comparison methods (Robertson et al. 2014; Jones et al. 2016).

The CWSSIM index discriminative power was demonstrated through its successful use in recognizing the similarity of palmprints (Zhang et al. 2007). Dumic, Grgic, and Grgic (2014) study also shed light on pattern recognition capability of a variant of CWSSIM. As mentioned previously, these applications can be considered analogous to landscape pattern or spatial model comparison, hence demonstrates the potential of these metrics in geographical research. For instance, the unique structural pattern recognition attributes of SSIM and CWSSIM shows that the metrics may have utility in map comparison or finding goodness-of-fit in spatial model

output and/or spatial explicit model validation (Pontius Jr and Schneider 2001; Pontius, Huffaker, and Denman 2004).

Recent trends in spatial pattern analysis have shown entropy-based metrics and deep-learning methods to be effective for extracting spatial structure at varying scales (Frazier 2019). For instance, Zhao and Zhang (2019) utilized a Wasserstein metric to estimate spatial configurational entropy of landscapes in digital elevation models and simulated datasets. Landscape patterns can also be characterized through information theory approaches that analyze spatial patterns by transforming them into two-dimensional spaces (Nowosad and Stepinski 2019). However, given that changes in entropy result from changes in processes, challenges remain in attempts to infer spatial patterns from processes (Vranken et al. 2015). The application of metrics from deep-learning methods such as neural networks offer potential for analysis and interpretation of patterns and processes in a data-driven framework to augment the limitations of hand-crafted pattern extraction metrics mentioned above (Buscombe and Ritchie 2018; Reichstein et al. 2019). More importantly, deep-learning approaches also take spatial context and texture into account while under-weighing illumination and luminance variations in underlying process-pattern change, and hence are able to effectively classify data with large and complex spatial structures (Buscombe and Ritchie 2018). These emerging tools could be compared with SSIM and CWSSIM to generate rank-order of images so as to establish a framework for selecting metrics for a given problem at hand (e.g., Dosselmann and Yang 2013).

We note that the use of CWSSIM in spatial pattern comparison might raise certain potentially challenging questions. For example, given that the metric is insensitive to non-structured distortions only at small scales, the question pertaining to the threshold distortions such as mis-registration and radiometric correction errors are considered extreme, and the scale at which

CWSSIM gains sensitivity and ceases to be geometric and scale invariant calls for further investigation. Also, as critiqued by Robertson et al. (2014) in the case of the SSIM index, presently there is no existing consensus on the indices scales in map comparison beyond relational mapping. Research could focus on development of a comprehensive scale for evaluating map similarity values and perhaps for commonly used sensors or data sources (e.g., Landsat analysis-ready data). For example, via computing similarity values for sequences of map-pairs, null distributions could be generated for certain geographical pattern types to enable hypotheses formulation and significance testing. Future research is also warranted to explore and exploit the geometric invariance property of the CWSSIM metric for implementation in irregular map comparison and landcover classification tasks.

The SWE maps employed in this study to represent real geographical data possess limited intrinsic structural complexity. Testing the indices on data sets with greater complexity in spatial structure, texture and edge patterns is suggested. Furthermore, future research might focus on assessing the metrics performance on maps yielded by independent pre-registration and radiometric correction methods or models. This will likely expose the degree of their invariance to scale, pixel misalignment, illumination and contrast changes. Finally, we suggest that human judgment of structural similarity is simultaneously assessed alongside CWSSIM index values to ascertain the extent to which the metric mimics human perception of landscape pattern similarity. Such analyses might be able to calibrate index values with decision-points in a natural resource management context.

3.6 Conclusion

To summarize, the paper concludes that computer vision metrics CWSSIM and SSIM are clearly of potential utility in geographical analysis, though with sensitivity to window size definition and need for further research to map index values to specific types of landscape change/similarity. The ranks yielded by both metrics disagree much more in the simulation scenario than in the SWE dataset, especially where similarity values were close. The metrics are both capable of detecting changes in real geographical patterns as seen in the SWE maps, but SSIM maps retain more spatially relevant information compared to CWSSIM maps. While SSIM values remained consistently higher than CWSSIM, there was higher variance in CWSSIM estimates. Though the behavior of CWSSIM with respect to changing window configuration is difficult to generalize, it demonstrated some slight improvements over SSIM when it comes to spatial pattern discrimination. This attribute may render the index robust for comparing output maps generated from spatial models. The metrics could enable the capture and comparison of subtle changes in spatial patterns. The CWSSIM, could be significantly sensitive to different model realizations, and could potentially distinguish between geographical or landscape patterns produced by closely associated but different spatial processes. However, we note that both metrics possess complimentary functionalities, suggesting the need to consider their tandem deployment, especially, where pattern comparison interest is to characterize medium to large scale changes in spatial patterns. We further emphasize that the most sophisticated pattern comparison metric may or may not extract features that depict the true underlying pattern generating process as seen in the SSIM and CWSSIM indices results in simulation and real-world datasets. It is possible to optimize different metrics to extract spatially discriminative patterns for specific landscape tasks. Therefore, for a broader spectrum deployment in landscape

pattern analysis and change detection, a suite of metrics ranging from simple (e.g., MSE) to the most sophisticated methods (e.g., deep learning methods) need to be considered. For example, ranking the scores of image-reference pairs under candidate metrics for a given landscape type may provide insight into which metric best captures the underlying changes in spatial patterns

Acknowledgements

The authors acknowledge funding provided by the Natural Sciences and Engineering Research Council, Canada to Dr. Robertson.

References

Betts, M. G., Diamond, A. W., Forbes, G. J., Villard, M. A., & Gunn, J. S. (2006). The importance of spatial autocorrelation, extent and resolution in predicting forest bird occurrence. *Ecological Modelling*, *191*(2), 197–224.

Boots, B., & Csillag, F. (2006). Categorical maps, comparisons, and confidence. *Journal of Geographical Systems*, *8*(2), 109–118.

Bowyer, K., Kranenburg, C., & Dougherty, S. (2001). Edge detector evaluation using empirical ROC curves. *Computer Vision and Image Understanding*, *84*(1), 77-103.

Buscombe, D., & Ritchie, A. C. (2018). Landscape classification with deep neural networks. *Geosciences (Switzerland)*, *8*(7), 1–23.

Carl, G., & Kühn, I. (2007). Analyzing spatial autocorrelation in species distributions using Gaussian and logit models. *Ecological Modelling*, *207*(2–4), 159–170.

Comber, A., & Wulder, M. (2019). Considering spatiotemporal processes in big data analysis: Insights from remote sensing of land cover and land use. *Transactions in GIS*, 879–891.

Coppin, P., Jonckheere, I., Nackaerts, K., Muys, B., & Lambin, E. (2004). Digital change detection methods in ecosystem monitoring: A review. *International Journal of Remote Sensing*, *25*(9), 1565–1596.

Delmelle, E. C. (2019). Toward a More Socially Impactful Geographical Analysis. *Geographical Analysis*, 1–9.

Dice, L. R. . (1945). Measures of the Amount of Ecologic Association Between Species. *Ecology*, *26*(3), 297–302.

Dosselmann, R., & Xue, D. Y. (2005). Existing and emerging image quality metrics. *Canadian Conference on Electrical and Computer Engineering*, *2005*, 1906–1913.

Dosselmann, R., & Yang, X. D. (2011). A comprehensive assessment of the structural similarity index. *Signal, Image and Video Processing*, *5*(1), 81–91.

Dosselmann, R., & Yang, X. D. (2013). A rank-order comparison of image quality metrics. *Canadian Conference on Electrical and Computer Engineering*, 1–4.

Dumic, E., Grgic, S., & Grgic, M. (2014). IQM2: new image quality measure based on steerable pyramid wavelet transform and structural similarity index. *Signal, Image and Video Processing*, *8*, 1159–1168.

Dungan, J. L., Perry, J. N., Dale, M. R. T., Legendre, P., Citron-Pousty, S., Fortin, M. J., Jakomulska, A., Miriti, M., & Rosenberg, M. S. (2002). A balanced view of scale in spatial statistical analysis. *Ecography*, *25*(5), 626–640.

Feng, W., Sui, H., & Tu, J. (2016). Object-Oriented Change Detection for Remote Sensing Images Based on Multi-Scale Fusion. *ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, *XLI-B7*, 483–491.

Frazier, A. E. (2019). Emerging trajectories for spatial pattern analysis in landscape ecology. *Landscape Ecology*, *34*(9), 2073–2082.

Gao, Y., Rehman, A., & Wang, Z. (2011). CW-SSIM Based Image Classification. *IEEE International Conference on Image Processing*, 1249–1252.

Groth, E. J. (1986). A pattern-matching algorithm for two-dimensional coordinate lists. *The Astronomical Journal*, *91*(5), 1244–1248.

Gustafson, E. J. (1998). Quantifying Landscape Spatial Pattern: What Is the State of the Art? *Ecosystems*, *1*, 143–156.

Hagen-Zanker, A., Straatman, B., & Uljee, I. (2005). Further developments of a fuzzy set map comparison approach. *International Journal of Geographical Information Science*, *19*(7),

769–785.

Hagen, A. (2002a). *Approaching human judgement in the automated comparison of categorical maps*. 1–6.

Hagen, A. (2002b). Multi-method assessment of map similarity. *5th AGILE Conference on Geographic Information Science*, 1–8.

Ioannidou, S., & Karathanassi, V. (2007). Investigation of the Dual-Tree Complex and Shift-Invariant Discrete Wavelet Transforms on Quickbird Image Fusion. *IEEE Geoscience and Remote Sensing Letters*, *4*(1), 166–170.

Jasiewicz, J., Netzel, P., & Stepinski, T. F. (2014). Landscape similarity , retrieval , and machine mapping of physiographic units. *Geomorphology*, *221*, 104–112.

Jones, E. L., Rendell, L., Pirotta, E., & Long, J. A. (2016). Novel application of a quantitative spatial comparison tool to species distribution data. *Ecological Indicators*, *70*, 67–76.

Kroese, D. P., & Botev, Z. I. (2015). Spatial Process Simulation. In *Stochastic geometry, spatial statistics and Random Fields. Lecture Notes in Mathematics, vol 2120. Springer, Cham* (pp. 369–404).

Li, Z. G., Xie, S. L., Yao, W., Rahardja, S. (2010). A Similarity Index in the Complex Wavelet Domain. *In Industrial Electronics and Applications (ICIEA) 5th IEEE Conference on 2010 Jun 15*, 1644–1649.

Long, J., & Robertson, C. (2018). Comparing spatial patterns. *Geography Compass*, *12*(2), 1–18.

Lu, D., Mausel, P., Brondizio, E., Moran, E. (2003). Change detection techniques. *International Journal of Remote Sensing*, *25*(12), 2365–2407.

Luders, E., Cherbuin, N., & Gaser, C. (2016). Estimating brain age using high-resolution pattern recognition: Younger brains in long-term meditation practitioners. *NeuroImage*, *134*, 508–

513.

Mairota, P., Cafarelli, B., Boccaccio, L., Leronni, V., Labadessa, R., Kosmidou, V., & Nagendra, H. (2013). Using landscape structure to develop quantitative baselines for protected area monitoring. *Ecological Indicators*, *33*, 82–95.

Makowiecki, W., & Alda, W. (2008). New sky pattern recognition algorithm. *In International Conference on Computational Science*, 749–758.

Mas, J. F. (1999). Monitoring land-cover changes: A comparison of change detection techniques. *International Journal of Remote Sensing*, *20*(1), 139–152.

McIntire, E. J. B., & Fajardo, A. (2009). Beyond description: The active and effective way to infer processes from spatial patterns. *Ecology*, *90*(1), 46–56.

McIntosh, A. R., Bookstein, F. L., Haxby, J. V., & Grady, C. L. (1996). Spatial pattern analysis of functional brain images using partial least squares. *NeuroImage*, *3*(3 I), 143–157.

Miller, H. J., & Goodchild, M. F. (2015). Data-driven geography. *GeoJournal*, *80*(4), 449–461.

Monserud, R. A., & Leemans, R. (1992). Comparing global vegetation maps with the Kappa statistic. *Ecological Modelling*, *62*(4), 275–293.

Nowosad, J., & Stepinski, T. F. (2019). Information theory as a consistent framework for quantification and classification of landscape patterns. *Landscape Ecology*, *34*(9), 2091–2101.

Oppenheim, A. V., & Lim, J. S. (1981). The Importance of Phase in Signals. *Proceedings of the IEEE*, *69*(5), 529–541.

Papaodysseus, C., Exarhos, M., Panagopoulos, M., Rousopoulos, P., Triantafillou, C., & Panagopoulos, T. (2008). Image and pattern analysis of 1650 B.C. wall paintings and reconstruction. *IEEE Transactions on Systems, Man, and Cybernetics Part A:Systems and*

Humans, *38*(4), 958–965.

Ping, J. L., Green, C. J., Zartman, R. E., & Bronson, K. F. (2004). Exploring spatial dependence of cotton yield using global and local autocorrelation statistics. *Field Crops Research*, *89*(2–3), 219–236.

Pontius Jr, R. G., Millones, M. (2011). Death to Kappa: birth of quantity disagreement and allocation disagreement for accuracy assessment. *International Journal of Remote Sensing*, *32*(15), 4407–4429.

Pontius Jr, R. G., & Schneider, L. C. (2001). Land-cover change model validation by an ROC method for the Ipswich watershed, Massachusetts, USA. *Agriculture, Ecosystems and Environment*, *85*(1), 191–203.

Pontius, R. G. (2000). Quantification error versus location error in comparison of categorical maps. In *Photogrammetric Engineering and Remote Sensing* , 66(8), 1011–1016.

Pontius, Robert Gilmore, Huffaker, D., & Denman, K. (2004). Useful techniques of validation for spatially explicit land-change models. *Ecological Modelling*, *179*(4), 445–461.

Power, C., Simms, A., & White, R. (2001). Hierarchical fuzzy pattern matching for the regional comparison of land use maps. *International Journal of Geographical Information Science*, *15*(1), 77–100.

Prieto, M. S., & Allen, A. R. (2003). A similarity metric for edge images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *25*(10), 1265–1273.

Rehman, A., Gao, Y., Wang, J., Wang, Z. (2013). Signal Processing : Image Communication structural similarity. *Signal Processing : Image Communication*, *28*(8), 984–992.

Reichstein, M., Camps-Valls, G., Stevens, B., Jung, M., Denzler, J., Carvalhais, N., & Prabhat. (2019). Deep learning and process understanding for data-driven Earth system science.

Nature, *566*(7743), 195–204.

Remmel, T. K., & Csillag, F. (2003). When are two landscape pattern indices significantly different? *Journal of Geographical Systems*, *5*(4), 331-351

Robertson, C., & Feick, R. (2017). Defining Local Experts : Geographical Expertise as a Basis for Geographic Information Quality. In *13th International Conference on Spatial Information Theory (COSIT 2017)*. Schloss Dagstuhl-Leibniz-Zentrum fuer Informatik.

Robertson, C., Long, J. A., Nathoo, F. S., Nelson, T. A., & Plouffe, C. C. F. (2014). Assessing Quality of Spatial Models Using the Structural Similarity Index and Posterior Predictive Checks. *Geographical Analysis*, *46*, 53–74.

Rogan, J., & Chen, D. (2004). Remote sensing technology for mapping and monitoring land-cover and land-use change. *Progress in Planning*, *61*(4), 301–325.

Sampat, M. P., Wang, Z., Gupta, S., Bovik, A. C., Markey, M. K., & Member, S. (2009). Complex Wavelet Structural Similarity : A New Image Similarity Index. *IEEE Transactions on Image Processing*, *18*(11), 2385–2401.

Takala, M., Luojus, K., Pulliainen, J., Derksen, C., Lemmetyinen, J., Karna, J. P., Koskinen, J., & Bojkov, B. (2011). Estimating northern hemisphere snow water equivalent for climate research through assimilation of space-borne radiometer data and ground-based measurements. *Remote Sensing of Environment*, *115*(12), 3517–3529.

van Vliet, J., Bregt, A. K., & Hagen-Zanker, A. (2011). Revisiting Kappa to account for change in the accuracy assessment of land-use change models. *Ecological Modelling*, *222*(8), 1367–1375.

Visser, H., & De Nijs, T. (2006). The map comparison kit. *Environmental Modelling and Software*, *21*(3), 346–358.

Vranken, I., Baudry, J., Aubinet, M., Visser, M., & Bogaert, J. (2015). A review on the use of entropy in landscape ecology: heterogeneity, unpredictability, scale dependence and their links with thermodynamics. *Landscape Ecology*, *30*(1), 51–65.

Walker, B. B., & Schuurman, N. (2015). The Pen or the Sword: A Situated Spatial Analysis of Graffiti and Violent Injury in Vancouver, British Columbia. *Professional Geographer*, *67*(4), 608–619.

Wang, Z., & Bovik, A. C. (2006). Modern image quality assessment. *Synthesis Lectures on Image, Video, and Multimedia Processing*, *2*(1), 1-156.

Wang, Z., Bovik, A. C., Sheikh, H. R., & Simmoncelli, E. P. (2004). Image quality assessment: from error visibility to structural similarity. *IEEE Transactions on Image Processing*, *13*(4), 600–612.

Wu, J. (2004). Effects of changing scale on landscape pattern analysis: Scaling relations. *Landscape Ecology*, *19*(2), 125–138.

Wulder, M. A., White, J. C., Coops, N. C., Nelson, T., & Boots, B. (2007). Using local spatial autocorrelation to compare outputs from a forest growth model. *Ecological Modelling*, *209*(2–4), 264–276.

Xiaolong Dai, & Khorram, S. (1998). Quantification of the impact of misregistration on the accuracy of remotely sensed change detection. *IEEE Transactions on Geoscience and Remote Sensing*, *36*(5), 1566–1577.

Zhang, L., Guo, Z., Wang, Z., & Zhang, D. (2007). Palmprint verification using complex wavelet transform. *IEEE Transactions on Signal Processing*, 2007–2010.

Zhao, Y., & Zhang, X. (2019). Calculating spatial configurational entropy of a landscape mosaic based on the Wasserstein metric. *Landscape Ecology*, *34*(8), 1849–1858.

Zhou, J., Yu, H., Smith, K., Wilder, C., Yu, H., & Wang, S. (2017). Identifying designs from incomplete, fragmented cultural heritage objects by curve-pattern matching. *Journal of Electronic Imaging*, *26*, 011022.

Zhou, W., & Alan C., B. (2002). A universal image quality index. *IEEE, Signal Processing Letters*, *9*(3), 81–84.

Appendix 3A

Table 3.1 Mathematical notations and meanings

Notation	Meaning
x	Reference image
y	Image to be compared with the reference image
c	Sub-band (sub-channel) of image
$c*$	Complex conjugate (transformed version) of c
c_x	Sub-band of reference image
c_y	Sub-band of image to be compared with the reference image
$c_{x,i}$	Pixel intensity for sub-band c_x
$c_{y,i}$	Pixel intensity for sub-band c_y
i	Pixel or cell value
$l(x,y)$	Luminance term
$c(x,y)$	Contrast term
$s(x,y)$	Structure term
n	number of pixels defined by window dimension
β	Weight parameter for contrast term
γ	Weight parameter for structure term

Table 3.2 Mathematical notations and meanings (continued)

Notation	Meaning
K	Small stabilizing positive constant for CWSSIM
K_1	Small constant to derive C_1
K_2	Small constant to derive C_2
C_1	Stabilizing constant derived from K_1
C_2	Stabilizing constant derived from K_2
C_3	Stabilizing constant derived from C_2
L	Dynamic range of pixel values
μ_x	image x mean intensity
μ_y	image y mean intensity
σ_x	Standard deviation of image x
σ_y	Standard deviation of image y
σ_{xy}	Covariance of map x and y
α	Weight parameter for luminance term
β	Weight parameter for contrast term
γ	Weight parameter for structure term

Chapter 4: Landscape Similarity Analysis Using Texture Encoded Deep-Learning Features on Unclassified Remote Sensing Imagery

4.1 Abstract

Convolutional neural networks (CNNs) are known for their ability to learn shape and texture descriptors useful for object detection, pattern recognition, and classification problems. Deeper layer filters of CNNs generally learn global image information vital for whole-scene or object discrimination. In landscape pattern comparison, however, dense localized information encoded in shallow layers can contain discriminative information for characterizing changes across image local regions but are often lost in the deeper and non-spatial fully connected layers. Such localized features hold potential for identifying, as well as characterizing, process–pattern change across space and time. In this paper, we propose a simple yet effective texture-based CNN (Tex-CNN) via a feature concatenation framework which results in capturing and learning texture descriptors. Using the Tex-CNN, gradient-based spatial attention maps (feature maps) which contain discriminative pattern information are extracted and subsequently employed for mapping landscape similarity. To enhance the discriminative capacity of the feature maps, we further perform spatial filtering, using PCA and select eigen maps with the top eigen value. We show that CNN feature maps provide descriptors capable of characterizing and quantifying landscape similarity. Using the feature maps' histogram of oriented gradient vectors and computing their Earth Movers Distances, our method effectively identified similar landscape types with over 60% of target-reference scene comparisons showing smaller Earth Movers Distance (EMD) (e.g., 0.01), while different landscape types tended to show large EMD (e.g.,

0.05) in the benchmark AID. We hope this study will inspire further research into the use of CNN layer feature maps in landscape similarity assessment, as well as in change detection.

4.2 Introduction

Earth system and environmental data have become abundant via a variety of sources ranging from model simulation data, citizen science, amateur drones, airborne sensors, commercial satellites, and easily accessible data such as Landsat (Dandois and Ellis, 2013; Miller and Goodchild, 2015). These data are available at unprecedented spatial and temporal resolutions and are widely used for understanding processes of environmental change across time and space. Given the rapidity of human-induced landscape disturbances, there is increasing interest in using environmental data resources to not only understand but also characterize and quantify landscape-scale disturbances, and to support decisions and policies aimed at remediating degraded landscapes (Townshend et al., 2012; Wulder et al., 2018).

Identifying the underlying processes that generate spatial patterns is critical to quantifying changes in patterns across space and time (Comber and Wulder, 2019). For instance, we ask questions like, where are degraded landscapes? What types of specific features are common or different between geographical locations? Such questions can be addressed through landscape pattern comparison. Traditional landscape similarity analysis tools, however, rely largely on change-detection analysis of classified landcover maps to predict or quantify process-driven changes. While these approaches have been successful, they are limited in uncovering the complex and non-linear nature of process–pattern relationships (Long and Robertson, 2018). Classification-based techniques also depend on the accuracy of the underlying map classification

method and incur challenges associated with legend harmonization and consistency/reproducible methods for data processing (Li et al., 2020a). Furthermore, processes and patterns are both interdependent and affect each other in many ways, thus complicating prediction efforts (Turner, 1989a). The growing historical archive of image data is increasingly being used to develop monitoring schemes and tools for understanding complex land-change processes (Comber and Wulder, 2019). Despite these advances, tools capable of extracting structural information from raw, unclassified land-image data in geographic context are limited.

Deep learning models are capable of learning to extract robust descriptors from image data. Such descriptors are useful representations of data structure, and hence hold potential for landscape research (Reichstein et al., 2019). For example, Tracewski et al. (2017) demonstrated the application of deep learning for characterizing different landcover types. Grinblat et al. (2016) also applied deep neural networks for plant species identification based on vein morphological patterns. The landscape similarity search algorithm proposed by Jasiewicz et al. (2014) illustrates the potential of computer vision methods to discover *similar* landscapes across space. Landscape similarity analysis is a fundamentally different problem than classification, seeking to quantitatively assess the similarity of whole landscapes, rather than label homogeneous elements within them. CNN models can be considered a recent class of spatially explicit models in geographic context which have demonstrated their effectiveness in many classification problems and show potential for application in a landscape similarity context (Janowicz et al., 2020).

Computer vision models, such as CNNs, contain filter banks which engage in spatial learning, to extract spatially discriminative features of increasing complexity through weight-sharing (Cimpoi et al., 2016). Lower CNN layer (e.g., first and second convolutional layer) feature maps

contain local information that captures fine-grain discriminative patterns useful for similarity mapping, while deep layer (i.e., layers close to network output) features lack geometric invariance, which weakens their robustness to map finely detailed landscape patterns across variable scenes (Gong et al., 2014). The layers of CNNs can preserve representative information about an input image with varying rotation and illumination characteristics (Mahendran and Vedaldi, 2014); consequently, pretrained CNNs can be employed to extract features for characterizing dynamic texture and dynamic scenes (Qi et al., 2016). Convolutional neural network filters exhibit consistent response to useful local regions of images; based on this property, Li et al. (2018) proposed a PatternNet that utilizes deconvolution (i.e., up-sampling) to discover discriminative and representative patterns in images. In a related study, Lettry et al. (2017) introduced a model capable of detecting repeated patterns in images. The authors provide evidence that consistent small patterns can be strongly expressed in the lower layers and hence are detected as major repetitions. Given the importance of texture in landscape aerial scenes, these properties may be particularly useful in recognizing different types of landscape scenes in aerial and satellite image data.

A variety of CNN architectures have been proposed to resolve image-classification problems in recent years (Kalantar et al., 2020; Flores et al., 2019). CNN layer depth, input size, and even training strategies adopted may influence the model performance and competitiveness with traditional machine learning techniques (Ghorbanzadeh et al., 2019). For instance, to learn multi-scale features which are robust to scale variation, and thus reduce misclassification rates, Liu et al. (2018d) proposed a method in which randomly cropped image patches are used for model development. Gong et al. (2018) also introduced a saliency-based feature extraction framework with anti-noise transfer network and found the approach to yield high classification accuracy on

benchmark datasets. CNNs with feature concatenation or fusion modules are simple but effective feature extraction frameworks that have been adopted to combine local and global image features for improving the performance of many scene classification and other pattern recognition tasks (Zhu et al., 2018; Zhuang et al., 2019; Petrovska et al., 2020). Ye et al. (2018b) presented a multi-stage model that extracts and fuses low-, middle-, and high-level features, and obtained 95% accuracy on the Aerial Image Dataset (AID). Kang et al. (2017) also developed a network that captures contextual information via the fusion of deep and shallow features to improve ship-detection accuracy. A framework with dilated convolution and skip connections was found to learn multiresolution discriminative features for scene classification (Fu et al., 2017a). Similarly, Gao et al. (2018) proposed a network in which feature maps generated from input images are passed on to a concatenating layer, forming a combined feature map with richer discriminative information. The authors concluded that their method significantly improved hyperspectral image classification. In a related study, Huang and Xu (2019) used weighted concatenation to combine features across all CNN layers, yielding overall accuracy of 95% in the AID. Similarly, Zeng et al. (2018) developed a two-branch CNN in which local and global features are independently extracted and concatenated. With extensive experiments, the authors demonstrated that feature concatenation resulted in over 90% accuracy for most scene classes in AID.

Despite the state-of-the-art performance of current CNN architectures, deep learning algorithms are generally perceived as "black-boxes" in both computer vision and across other domains; consequently, there have been intensifying calls to interrogate and reveal the inner workings of deep learning models in disciplines such as geography (Gahegan, 2020). Visualizing spatial attention maps (i.e., feature maps) is a fairly simple method of exploring how CNNs learn and make decisions on an input image. The approach may be gradient-based and involve

computing network output changes with respect to input (Simonyan et al., 2013), or utilize a deconvolution network that projects image features over a plane (Zeiler and Fergus, 2014). Zhou et al. (2016) also proposed converting the linear decision (regression) layer into a convolutional layer for generating class-based attention maps. To improve gradient-based feature map quality, guided backpropagation has also been introduced (Springenberg et al., 2014). As these approaches do not always produce class-specific feature maps (Zagoruyko and Komodakis, 2017), Selvaraju et al. (2017) proposed Grad-CAM, which integrates guided backpropagation and class activation maps, and thus yielding class-discriminative spatial attention maps. In related research, Omeiza et al. (2019) proposed Smooth Grad-CAM++ to improve the spatial resolution and localization of patterns in feature maps. Class-selective relevance mapping has also been proposed to derive feature maps that contain the most discriminative regions of interest in medical images (Hongyun Zhang et al., 2019).

In this study, we focused on landscape similarity assessment on unclassified imagery using gradient-based convolutional feature maps. Gradient with respect to an image is a sensitivity map measuring how changes at a pixel spatial location affect changes in CNN model predictions (Selvaraju et al., 2017). To derive gradient-based feature maps, a pre-trained CNN model is therefore required. To this end, we propose training and deploying a texture-encoded CNN model (Tex-CNN) to extract feature maps for landscape similarity comparison. A classical CNN was also trained to assess the effectiveness of the proposed Tex-CNN model. Using the trained Tex-CNN, we derived feature maps which were evaluated in terms of capturing discriminant properties of different landscape scenes. The similarity analysis framework adopted here could form the basis of landscape retrieval / search systems, as well as fuzzy landscape search where multiple landscape scene types co-occur.

The contribution of this study is, therefore, two-fold. (1) A gradient-based convolutional feature map approach to landscape similarity analysis was proposed. Using gradient-based features, the proposed landscape similarity assessment utilizes *significant spatial patterns* in a query and a candidate image for comparison. (2) A landscape similarity metric capable of detecting within- and between-landscape types was developed. The use of CNN feature maps to characterize landscape similarity is limited in geography, and more importantly, this study is the first exploration of gradient-based CNN features for comparing landscapes. The paper is arranged as follows: We first illuminate the importance of spatial feature maps in landscape comparison; next, the methodological pipeline is presented, followed by results, discussion, and conclusion.

4.3 Related Work

Prior to the emergence of state-of-the-art of CNNs capable of detecting and classifying objects and patterns, image texture processing was one of the earliest applications in which CNNs were employed to extract discriminative local features (Ustyuzhaninov et al., 2016; Gatys et al., 2017).

4.3.1 Representing Patterns in CNN Feature Maps

Convolutional *feature maps* can be thought of as spatial activation features encoding discriminative regions within a given input image (Girdhar and Ramanan, 2017). A feature map can also be viewed as detection scores resulting from the application of a filter over spatial locations in a 2D image; the activation value obtained at the *i*-th location quantifies the importance of the pixel at that location (Cao et al., 2016). Such locations may be linked, at least

conceptually, to “landscape features of interest” or those areas of the landscape that are discriminative of the landscape scene label. The potential of a convolutional-feature-based approach in urban landscape change detection was presented in El Amin et al. (2016). The authors demonstrated that CNN features can perform higher than “hand-crated features” and other state-of-the-art techniques. In related research, Albert et al. (2017) showed that features extracted from CNNs trained discriminatively on urban imagery effectively compare neighborhood similarity across European cities.

In landscape research where local-to-global changes or pattern similarity are sometimes of interest, CNN maps can be helpful. Feature maps represent local response regions of filters and thus encapsulate valuable pattern information (Zagoruyko and Komodakis, 2017). These local regions also encode information pertaining to the underlying pattern-generating process. Feature maps from convolutional layers represent local descriptors of particular image regions which can be aggregated into global descriptors for image retrieval (Yandex and Lempitsky, 2015). An image-retrieval framework is also closely related to the landscape-pattern comparison problem. For instance, CNN activations containing pronounced spatial information can be utilized for detecting repeated patterns (Lettry et al., 2017). The challenge to detect repetitive spatial patterns is similar to the landscape similarity analysis problem. It has been illustrated that convolutional layer activations are local region descriptors and outperform many state-of-the-art descriptors (Gatys et al., 2017; Girdhar and Ramanan, 2017); thus, if these feature maps are well-pooled, a compact representation of a given landscape can be derived. Additionally, Zagoruyko and Komodakis (2017) have shown that feature maps represent “knowledge learned” by a given network about the underlying pattern and can be transferred to other networks, to improve pattern detection. Furthermore, classical machine-learning algorithms for pattern detection or

classification, such as Random Forest, Support Vector Machine, and Maximum Likelihood being employed in landscape research, can be coupled with deep feature extraction models to boost performance (Wang et al., 2019). For example, it has been shown that feeding features from CNNs to other models improve results (Petrovska et al., 2020). We therefore postulate that CNN-feature-based frameworks hold the potential to enable the detection and quantification of spatial patterns in unclassified image data.

4.3.2 CNN-Feature-Based Image Retrieval

Image retrieval is an active research area in this era of "big data", where the objective is to find a set of images that are the most similar to a given query image. Content-based image retrieval (CBIR) is a widely applied technique for retrieving images in databases. In CBIR, low-level image descriptors (e.g., color, texture, and structure) are extracted to form an image representation; a suitable measure is then selected to estimate similarity between images. Several algorithms have been proposed for an improved CBIR. For example, Unar et al. (2019) combine both visual and textual features for image retrieval. Zhang et al. (2019) also developed an algorithm that segments an image into salient, non-salient, and shadowed regions, in order to extract spatially relevant information. Earth observation data now available in various archives could provide a wealth of information through effective search and retrieval techniques (Peng et al., 2019).

Recent research has shifted towards the use of features extracted from deep convolutional layers of CNNs for image matching and retrieval (Shi and Qian, 2019; Gu et al., 2019). The use of deep convolutional features for image retrieval is demonstrated in a study conducted by Yandex and Lempitsky (2015). Chen et al. (2019) propose region-of-interest deep convolutional

representation for image retrieval. Their approach first identifies regions of interest and proceeds to extract features from the fully connected layer. Shi and Qian (2019) also adapted the region-of-interest-based approach called strong-response-stack-contribution, by exploring spatial and channel contribution, to generate a more compact global representation vector for an object-based image retrieval challenge. Cao et al. (2016) applied adaptive matching by splitting feature maps and later spatially aggregating them into regions of interest for comparison. Liu et al. (2015) proposed extracting and pooling subarrays of feature maps as local descriptors for visual classification task and found that the method outperforms features from fully connected layers. The aforementioned applications demonstrate the capability of CNN features to represent discriminative patterns which are useful for quantifying change and similarity, and therefore hold potential for designing resource management and decision-making applications in geography.

4.4 Materials and Methods

4.4.1 Models' Architecture

In the context of landscape similarity mapping, global shape information present in fully connected layers is of less significance, as landscape patterns often lack unique or stable geometry across space. Given that lower layers capture local patterns (Cimpoi et al., 2016), we concatenated multi-layer features, to learn a discriminative representation of the data-generating process. In feature fusion, feature maps from three convolutional layers (i.e., conv1, conv2, and conv3) are concatenated followed by flattening into feature vectors to yield a dense layer (denoted FC1). One possible approach to improving CNN features' discriminative potential is to apply attention pooling strategies that takes the weighted sum of different feature maps instead of

concatenating features, but this technique exponentially increases model parameters as well. However, we adopted feature concatenation, as it has been proven to enable the extraction of multiscale features, potentially obviating the need for multiscale inputs during model development (Lim and Keles, 2020). Moreover, attention strategies are effective for object recognition tasks but may not tangibly improve landscape pattern discrimination.

Work similar to our approach is Andrearczyk and Whelan (2016) feature concatenation framework. Figure 4.1 illustrates the architecture of a classical CNN, while Figure 4.2 depicts our model architecture. For model design, we build on the VGG16 model architecture and filter constellation (Simonyan and Zisserman, 2015). Thus, 32, 64, and 128 Filters are used in the first, second, and third convolutional layers of the classical CNN and the Tex-CNN models. The proposed model architecture is intended to be simple with minimum parameters as possible for field deployment.

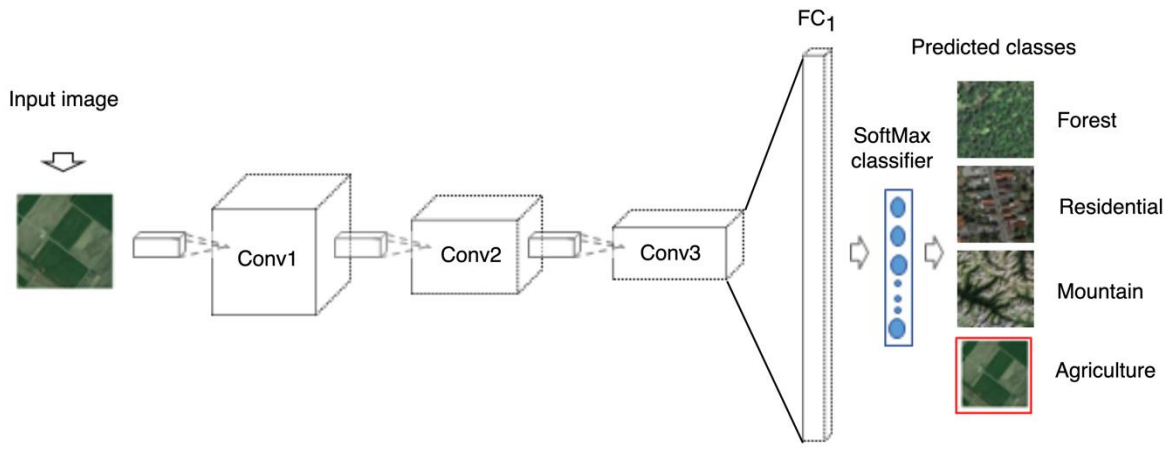


Figure 4.1 Architecture of a classical convolutional neural network (CNN). The CNN applies convolutional operations, as well as max-pooling, to process input tiles, but no feature concatenation is implemented.

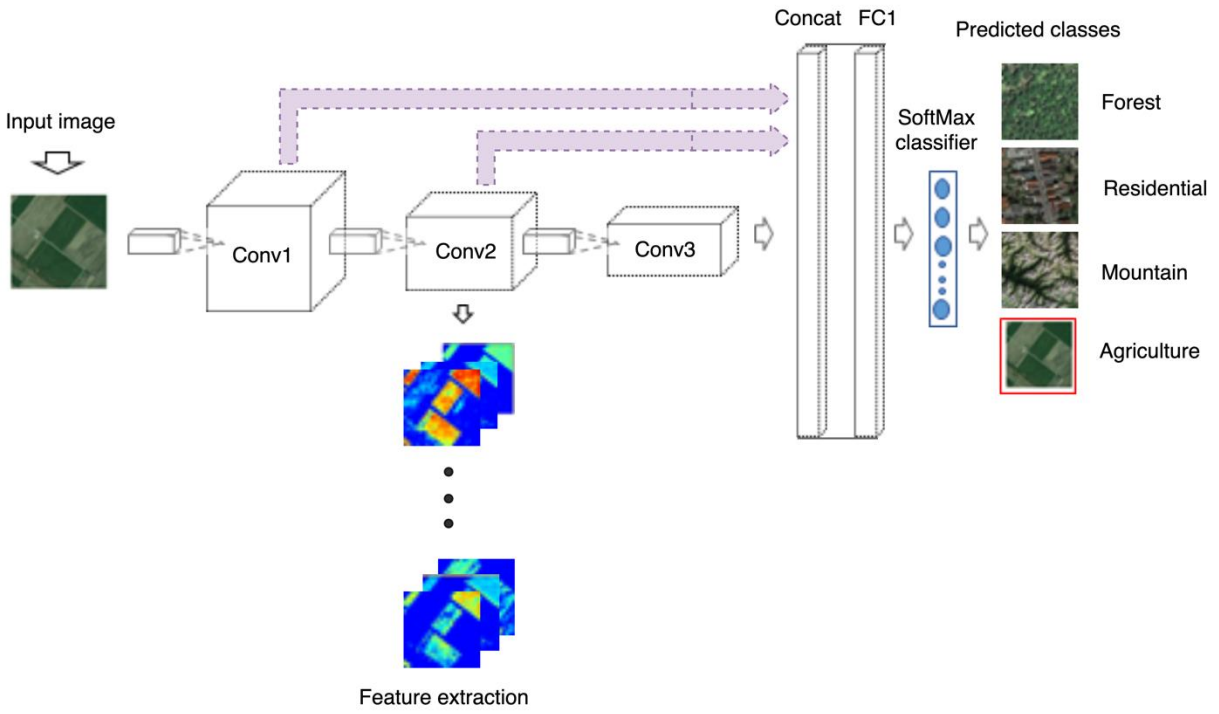


Figure 4.2 Architecture of our proposed texture-based (Tex-) CNN model. Con1, Conv2, and Conv3 denote convolutional layers, while FC1 denotes fully connected layer. Concat layer represents concatenation of Conv1, Conv2, and Conv3 feature maps.

4.4.2 Model Parameterization and Training

We opted for training our models from scratch, as this approach gives flexibility over model architecture. Although there is potential for data limitation, as well as over-fitting, in this framework (Nogueira et al., 2017), the approach facilitates feature maps comparison, as it ensures that features are the direct result of filters learned on data presented to models, compared to using pretrained networks in which filters learned from an entirely different domain than the task at hand. Given that the input image size is large enough (i.e., 225 × 225), we selected 7 × 7 convolutional kernels and used a fixed filter size throughout the convolutional layers. The space

between successive convolutional filter operations (stride) is set to 1 pixel. Filter receptive field size changes with layer depth and could result in profound differences in feature spatial resolution between successive layers. In the pooling layers, 2 × 2 max pooling with stride 2 is applied. The receptive field size at the third convolutional layer, therefore, becomes 46. We utilized 75% of the sample data for training and 25% for validation. To mitigate potential overfitting, 25% drop-out is used in convolutional layers, while 50% is applied to the FC1 layer (Hinton et al., 2012). The rectified linear unit (ReLU) is used as the activation function. Multiclass cross-entropy loss function is employed, and the models are trained for 30 iterations with Adam as the optimizer. Adam adaptively computes and updates gradients and is invariant to diagonal scaling of gradients (Kingma and Ba, 2015). The Keras-Tensorflow backend was used for building and supporting computations required to train the CNN models on a GPU with a NVIDIA-supported graphics card. Table 4.1 summarizes the models' architecture and parameters.

Table 4.1 A summary of models' architecture and parameters.

Layer Name	**Convolution**	**Max-Pooling**	**Activation**	**Drop-Out**
Conv-1	7 × 7 × 32	2 × 2	ReLU	25%
Conv-2	7 × 7 × 64	2 × 2	ReLU	25%
Conv-3	7 × 7 × 128	2 × 2	ReLU	25%
FC1	No	No	SoftMax	50%

4.4.3 Application Context: Landscape Comparison

Unclassified imagery, which is now ubiquitous due to the availability of sensors of varying types, offers the potential for landscape similarity queries. While land-cover classification in which pixels are labeled (classified) or objects are segmented and characterized is a predominant use of aerial and satellite imagery (Liu et al., 2015), in this modeling framework, we focus on characterizing whole scenes or landscapes. An implementation of this would be helpful for automating image retrieval and potentially provide a basis for mixed scenes and/or novel land-scene categories and/or descriptors. A conceptual representation for comparing unclassified images (aka landscapes/scenes) is depicted in Figure 4.3, using three landscapes/scenes denoted as X, Y, and Z, but the representation is expandable to multiple landscape types. Given an image, the feature map will be extracted for comparison, using EMD. EMD(X, X'), EMD(Y, Y'), and EMD(Z, Z') compute within-landscape similarity, while EMD(X, Y), EMD(Y, Z), and EMD(Z, X) estimate between-landscape similarity.

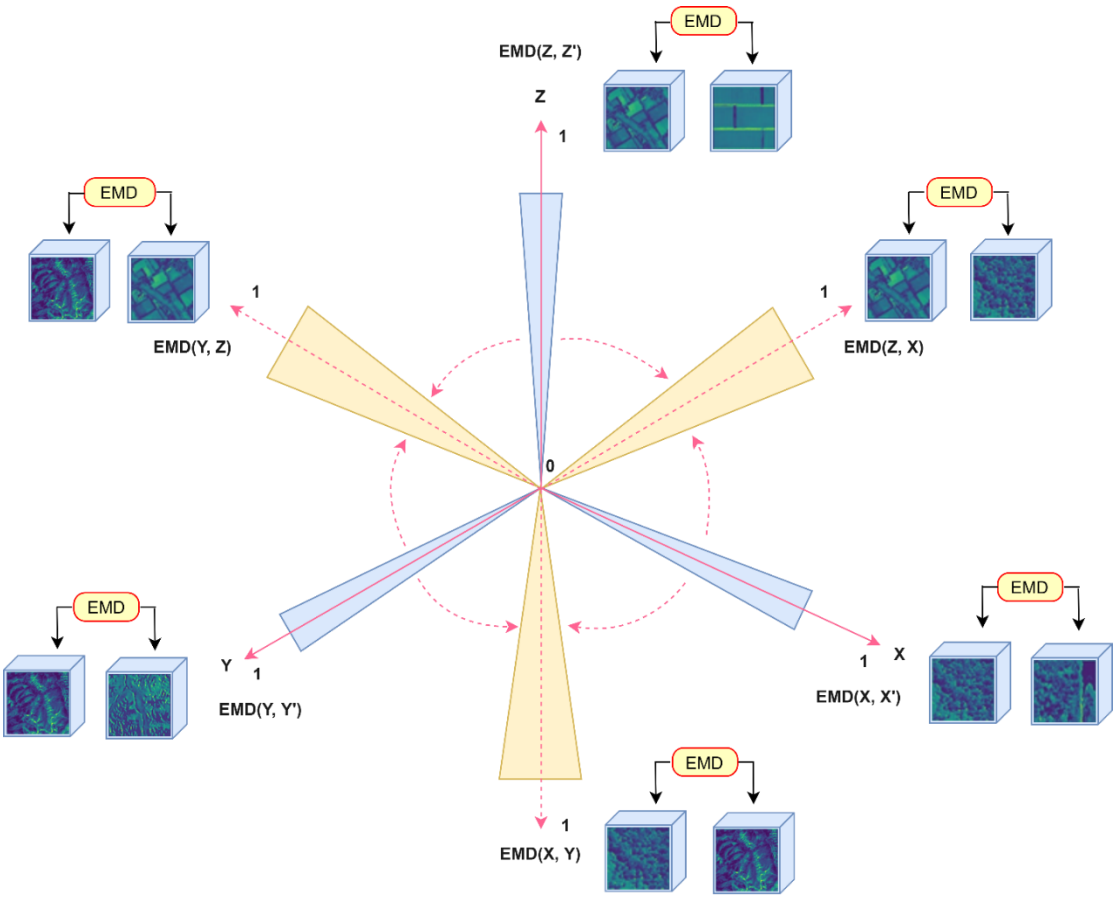


Figure 4.3 A conceptual framework for unclassified images/scenes comparison. Earth Movers Distance (EMD)(X, X'), EMD(Y, Y'), and EMD(Z, Z') denote within-landscape comparison, while EMD(X, Y), EMD(Y, Z), and EMD(Z, X) represent between-landscape comparison.

Benchmark datasets have long been used in computer vision for model development, due to the scarcity of labeled data, and the laborious processes required for generating such datasets, yet they remain relatively rare in geospatial research. The aerial imagery dataset is composed of high-resolution benchmark data recommended for training scene classification models (Xia et al., 2016). The AID contains multi-resolution images; the pixel spatial resolution varies from about half a meter to eight meters, providing a suitable dataset for training classical CNN and

Tex-CNN models. Although, there are 30 aerial scene types in the AID; this study considered 3 scene types – agriculture, forest and mountain were selected. Agricultural scene types are generally characterized by fields with conspicuous boundaries. Such fields may be bounded by regular or irregular boundaries and may contain green vegetation or bare soil. Forests scenes are landscapes most covered with trees, sometimes with intermittent patches. Mountains are terrain type scenes with variable morphology and may or may not contain vegetation. It is important to note that in reality these scene categories are not mutually exclusive: forests occur on mountains and agriculture can as well. We employ the notion of a dominant scene category as the basis for our modelling, and explore how scenes with lower similarity scores may represent mixed-scene landscapes. A common protocol in computer vision is to split a given dataset into training, validation, and test samples. This may sometimes result in high-accuracy reports resulting from overfitting. Owing to this caveat, and the need to find models capable of generalizing over a range of datasets for field application, we propose carrying out further validation by using a dataset from an entirely different sensor. As such, we employed Sentinel data to evaluate the generalizability of the models.

4.4.4 Datasets and data augmentation

The AID consists of diverse landscape types; to test the robustness and potential generalizability of the proposed CNN feature-based landscape similarity comparison, we selected three landscape types in which some landscape instances can be difficult to clearly distinguish. For example, the occurrence of vegetation on certain mountains may challenge the capability of a metric to accurately distinguish mountainous landscapes from agriculture and forest landscapes. However, considering only three landscapes reduces the sample size.

Convolutional neural networks are “data hungry” models; thus, training such models from scratch by using fewer samples and classes is likely to pose data limitation issues and overfitting. Overfitting occurs when the CNN performance is better on the training dataset but low on the test dataset (i.e., the model has high variance) (Hinton et al., 2012). We attempt to circumvent this challenge via the application of data augmentation. Data augmentation strategies have been shown to improve models' performance by increasing the size of feature space and patterns available for models to learn. To that end, we employed the Keras image data generator API to augment our training dataset. The original spatial extent of images in AID is 600 × 600 pixels; to generate 255 × 255 tiles, images belonging to the three scene types were selected and randomly cropped to yield 1000 images per scene. Given that the AID is multiresolution – image pixel sizes vary from about half a meter to eight meters, scale representation challenges are inherently reduced such that scale transformations may not make substantial difference following data augmentation. Bearing this in mind, horizontal flips and rotations (i.e., 45 – 180 degrees) were applied to the 1000 images in each scene category, resulting in 3000 images per aerial scene type, thus yielding 9000 training samples for three landscape types: agriculture, mountain, and forest. An additional 150 images in each landscape type were augmented, doubling the number of images per landscape type, and resulting in 900 test samples. In the Sentinel dataset, 600 images were used for independent validation of the potential generalizability of the method on medium-resolution satellite imagery. A separate set of 510 images from Sentinel dataset were combined with the 9000 images from AID to retrain the models. Table 4.2 describes the datasets and number images utilized in this study. We note that no data augmentation was applied to Sentinel dataset. Figure 4.4 illustrates samples of AID landscapes used in our experiments.

Table 4.2 Data types and specific application.

Data Source	Attribute	How Data is Utilized	No. of Images
AID	Aerial imagery, pixel spatial resolution vary between 0.5 m and 8 m	Training and testing models, and building similarity distributions	9000 images used training (75%) and validation (25%). 900 images used for testing (e.g., deriving confusion matrix)
Sentinel data	Open-source satellite data; 10 m pixel resolution	Visualizing feature maps, testing model generalization to medium resolution imagery	600 images used for testing and computing confusion matrix.
		Retraining models on Sentinel dataset	510 images combined with 9000 AID mages

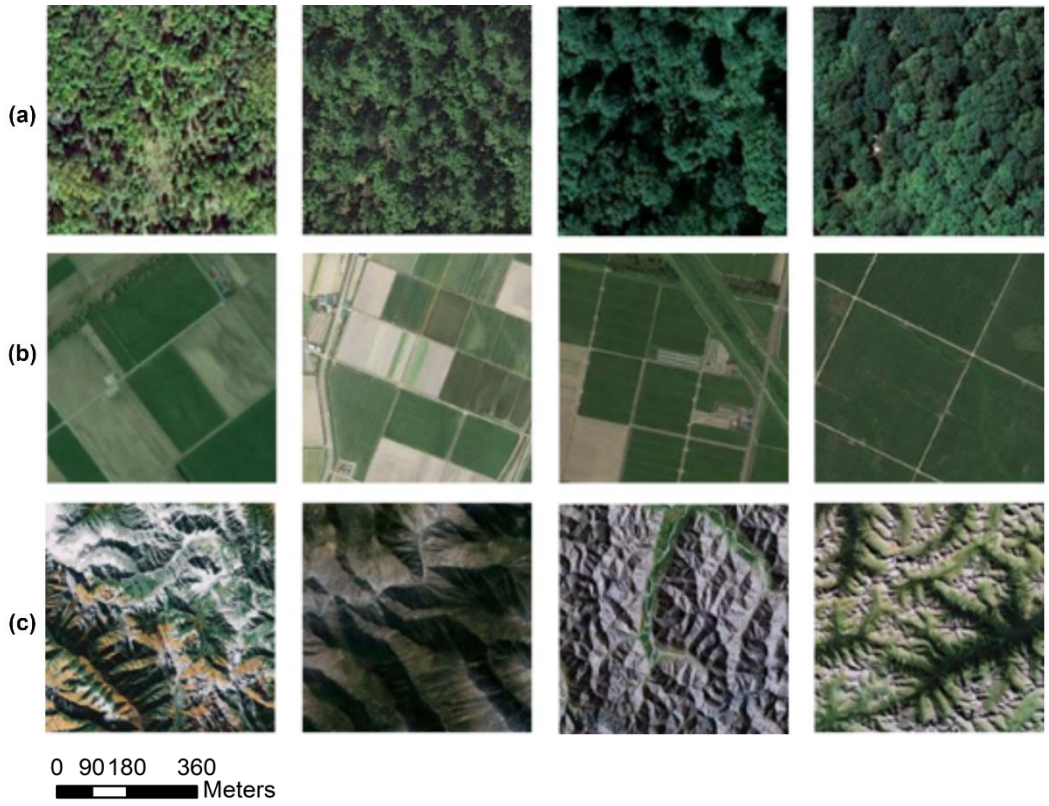


Figure 4.4 Selected landscape categories from the AID dataset. Row (a) forest landscape, row (b) agriculture landscape, and row (c) mountain landscape.

4.4.5 Activation/Feature Maps Derivation

Given a trained CNN model, gradient-based activation maps can be computed to allow for visualization of localized regions in an image that contribute significantly to a given output pattern. Using our trained classification model, activation maps were derived via backpropagation of filter responses to input pixel intensities (Selvaraju et al., 2017). ReLU was employed to constrain the backpropagation process to propagate only positive pixel values that activate filters; these pixel positions contain the highest weight and are therefore said to encode "significant patterns" or represent the signatures of the underlying pattern-generating process.

The gradient-based class activation map proposed by Selvaraju et al. (2017) is derived as follows: Let Y^c denote the score for a particular landscape scene. The gradient, with respect to Y^c (i.e., backpropagation gradient), is formulated as $\frac{\partial Y^c}{\partial A_{ij}^k}$. A^k denotes a set of CNN activation maps, and (i, j) are locations of pixels in the feature maps. Equations (1) and (2) summarize feature maps derivation.

$$\underbrace{\alpha_k^c}_{Feature\ map\ weight} = \overbrace{\frac{1}{Z}\sum_i \sum_j \underbrace{\frac{\partial Y^c}{\partial A_{ij}^k}}}^{Global\ average\ pooling} \qquad \textit{Equation 4.1}$$

$$L_{feature\ map}^c = ReLU\left(\sum_k \alpha_k^c\, A^k\right) \qquad \textit{Equation 4.2}$$

The weight term α_k^c captures the "significance" of feature map k for a target landscape type/scene. ReLU is applied to the weighted sum of feature maps, yielding heat-maps whose local regions highlight the most discriminant patterns in images. The resultant CNN activation maps pinpoint locations where the model focuses its attention on, since such locations contain significant spatial patterns. Therefore, activation maps can be referred to as "saliency maps" or "spatial attention maps".

4.4.6 Extracting HoG Vector from Feature Maps

For each landscape type, 50 landscapes were randomly selected from the AID for feature map extraction. The 50 landscape types were drawn from the original 150 samples, and hence were not augmented. The 50 landscape types chosen here are for the purpose of demonstration only, but our method can be extended to over 50 landscapes. Aerial scenes were collected from different countries to create the AID (Xia et al., 2016). Therefore, each randomly selected scene type can be treated as a potential landscape from different spatial location. As shown in Table 4.2, feature map similarity distributions were derived using AID, while Sentinel data was used to visualize gradient-based feature maps. We note that, since the number of filters in the second convolutional layer from which the feature maps were computed is 64, each image correspondingly yields 64 feature maps. Spatial filtering was performed using PCA to reduce the number of feature maps per image, yielding a more compact image descriptor (Yandex and Lempitsky, 2015). Such a step is inevitable when CNN feature maps are being compared; due to discriminative learning, not all filters respond to input images or pixels, and, as such, certain feature maps may contain no features/patterns where a filter is not activated by an input image (Yang et al., 2015; Xie et al., 2017; Luo et al., 2018). Using PCA, a feature map (i.e., eigen map) that has the highest eigen value was selected. Next, the HoG vector is extracted from each landscape type feature map. HoG has been shown to extract effective image descriptors for pattern recognition tasks. For example, human face recognition across standard datasets was found to improve with the use of HoG descriptors (Déniz et al., 2011). In related research, different plant species were effectively recognized from leaf patterns, using HoG descriptors (Quoc Bao et al., 2020). Setting the spatial parameters (i.e., cell size and cells per block) for extracting HoG features, however, requires a careful approach. In our implementation, the HoG

vector was extracted by considering *cell size 24 × 24* dimension and *cell-per-block* = 2 × 2 of each feature map. We deployed the EMD, a multivariate histogram distance measure to compare the resultant HoG vector representing reference and test feature maps. Rubner et al. (2000) illustrated the effectiveness of EMD in image retrieval context.

4.4.7 Formulating the Feature Map Comparison Metric

In the literature, there are a variety of pattern similarity comparison metrics, yet it is challenging to find robust and generic metrics to rely on when it comes to landscape similarity comparison. In this section, we illustrate how our convolutional feature map comparison metric was derived. Figure 4.5 is a depiction of our proposed convolutional feature-based landscape similarity comparison.

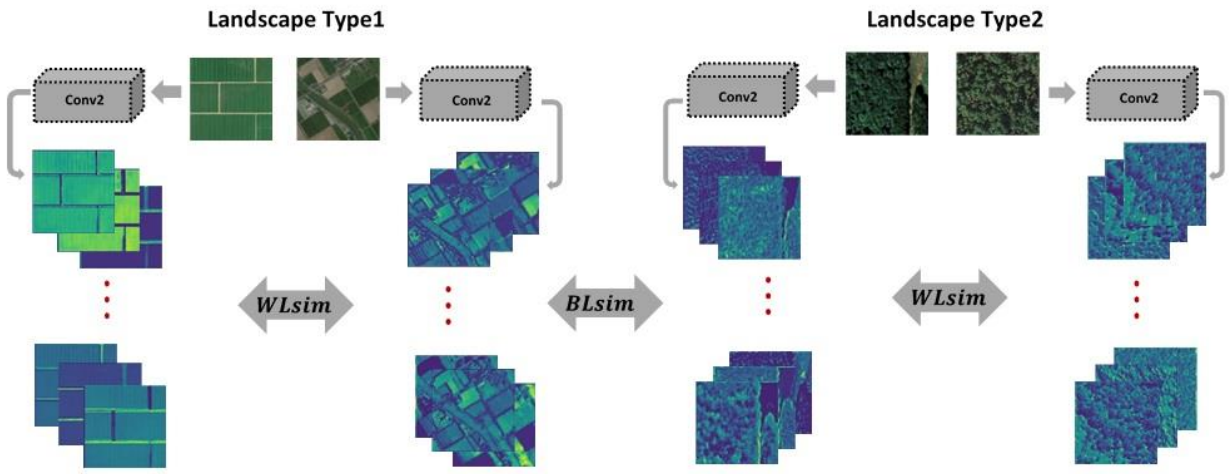


Figure 4.5 A framework for CNN-feature-based landscape similarity comparison. Notice that within-landscape comparison (WLsim) compares features in similar landscapes type 1 (agriculture landscapes) and landscape type 2 (forest landscapes), while between (an across) landscape comparison (BLsim) cross-compares feature maps in landscape type 1 versus landscape type 2.

Equations (3) and (4) illustrate our formulation and computation of within- and between-landscape similarities.

$$WLsim = EMD\left(HoG\left(L_{L1\ type,locX}\right), HoG\left(L_{L1,locY}\right)\right) \quad \textit{Equation 4.3}$$

$$BLsim = EMD\left(HoG\left(L_{L1\ type}\right), HoG\left(L_{L2\ type}\right)\right) \quad \textit{Equation 4.4}$$

where $L_{L1\ type}$ and $L_{L2\ type}$ represent different landscape categories from different spatial locations. $WLsim$ and $BLsim$ denote within- and between-landscape type comparison, respectively. For $WLsim$, we compare similar landscapes; example $L_{L1\ type}$, but from different locations (e.g., $locX\ vs\ locY$). For example, to compare agriculture landscapes, $locX$ will

represent a reference landscape, while $locY_{(1,2,3\ldots,n)}$ denotes agriculture landscapes (e.g., 225 × 225 grids) from other locations of interest. Landscapes whose spatial extents are large could be tiled into spatial grids of equivalent dimension as the model input size for comparison. $BLsim$ involves a comparison of two disparate landscape types (e.g., forest versus agriculture). $HoG(\cdot)$ computes HoG feature vector, given an input feature map, while $EMD(\cdot)$ estimates HoG feature vector similarity based on the EMD between landscapes. To test the proposed metric, 50 feature maps from each landscape type exhibiting the highest eigen value were randomly split into two subsets, thus yielding 25 images per subset, which are named G1 and G2 (e.g., agricG1 and agricG2 each contains 25 images belonging to agriculture landscapes). The random shuffle function in scikit-learn library effectively shuffles training data to ensure that images with similar features/patterns are thoroughly mixed. The random shuffle function was therefore applied to shuffle feature maps before splitting them into a subset of 25; this resulted in substantial intra-landscape variability in each subset. Using the proposed metric, a compact distribution based on EMD was computed for within- and between-landscape, by comparing each scene type; for example, in agricG1, a selected scene was compared with all other scenes in agricG2. This permutation schema was repeated for all the 25 scenes in agricG1.

4.5 Experimental Results

4.5.1 Landscape Type Prediction Models

Figure 4.6a–f depicts classification accuracies for landscape types on AID and Sentinel data. The confusion matrices are computed by deploying the models on the test images from AID and Sentinel datasets (i.e., 900 images for AID and 600 images for Sentinel-2). In Figure 4.6a,b, the

Tex-CNN and the classical CNN classification accuracy reports are similar except for mountainous scenes where Tex-CNN has higher classification accuracy. In Figure 4.6c,d, the first row of the confusion matrix shows that over 90% of the agriculture landscapes are misclassified as forest in Sentinel dataset. About 70% of the mountain landscapes are correctly classified by the Tex-CNN, while the classical CNN achieves only 25% classification accuracy. Figure 4.6e,f shows classification accuracies after fine-tuning the models. Fine-tuning was conducted using a combination of AID and Sentinel data to re-train the models. It can be observed that misclassification rates for agriculture landscapes have been substantially reduced after fine-tuning.

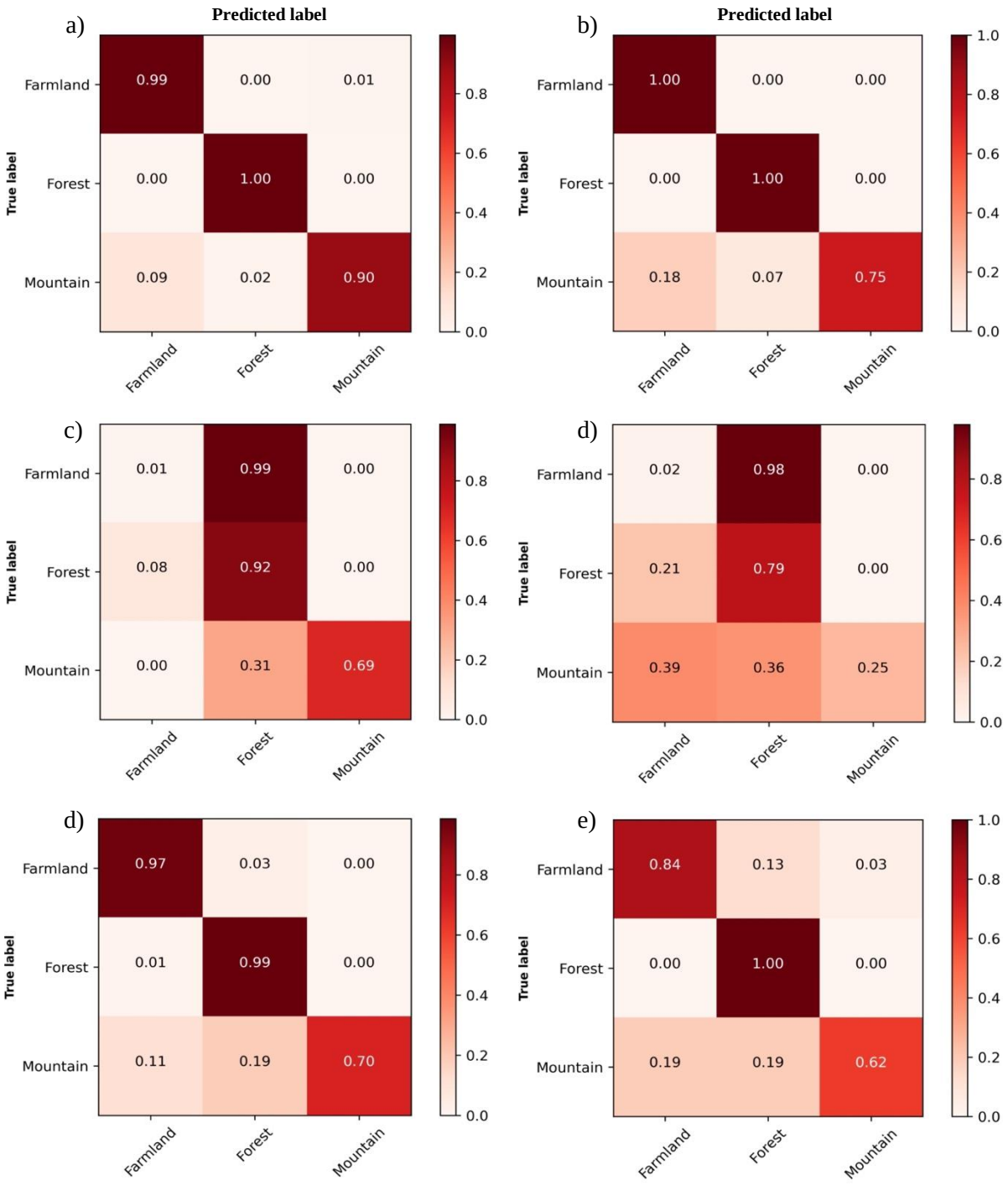


Figure 4.6 Confusion matrix for landscape-type classification accuracy. (a,b) Tex-CNN and classical CNN accuracy on AID. (c,d) Classification accuracy for Tex-CNN and classical CNN on Sentinel dataset. (e,f) Accuracy for Tex-CNN and classical CNN, respectively, after retraining on a combination of AID and Sentinel test data.

Table 4.3 compares the accuracy reports for reference state-of-the-art techniques and Tex-CNN on the AID. It can be seen that the Tex-CNN is competitive with other methods in terms of per-landscape type classification accuracy.

Table 4.3 Overall accuracy (OA) and selected per-scene class accuracy for reference and our proposed Tex-CNN on the AID.

Methods	**Agriculture (%)**	**Mountain (%)**	**Forest (%)**	**OA (%)**
TEX-Net-LF (Anwer et al., 2018)	95.5	99.9	95.75	92.96
Fine-Tuned SVM (Yu and Liu, 2018)	97.0	99.0	98.0	95.36
PMS (Petrovska et al., 2020)	98.0	99.0	99.0	95.56
CTFCNN (Huang and Xu, 2019)	99.0	100	99.0	94.91
GCFs + LOFs (Zeng et al., 2018)	94.0	99.0	99.0	96.85
MF2Net (Xu et al., 2020)	97.0	91.0	94.0	95.93
Classical CNN	100	75.0	100	91.67
Tex-CNN	99.0	90.0	100	96.33

4.5.2 Exploring CNN Layer Features Suitability for Landscape Comparison

Given that CNN layers process inputs hierarchically, feature maps spatial resolution become coarser with layer depth: Earlier layers contain finer resolution features, while deeper layer representation gives coarser features. We conducted visual assessment of feature map quality, as well as the potential utilization of the second- and third-layer feature maps. Layer-one features were not included in this analysis, as gradient-based features cannot be computed by using input image data as the penultimate layer. Figure 4.7 depicts feature maps with the highest eigen

values extracted from Tex-CNN. The feature maps are the result of applying PCA to layer two- and three-feature tensors. Notice how the spatial resolution changes across the layers. While layer-two eigen maps are fine-grained, with distinct patterns (e.g., agriculture landscape boundaries, tree clusters), this pattern is not clearly interpretable in layer-three eigen maps. In Figure 4.7, row (a), layer-two shows high-resolution features with conspicuous agricultural field boundaries. Contrarily, layer-three map depicts low-resolution features; the boundaries of individual parcels are blurred out. In Figure 4.7, row (b), layer-two shows fine-grained clusters of trees; layer-three, on the other hand, depicts coarse scale patterns which are not immediately recognizable as forest.

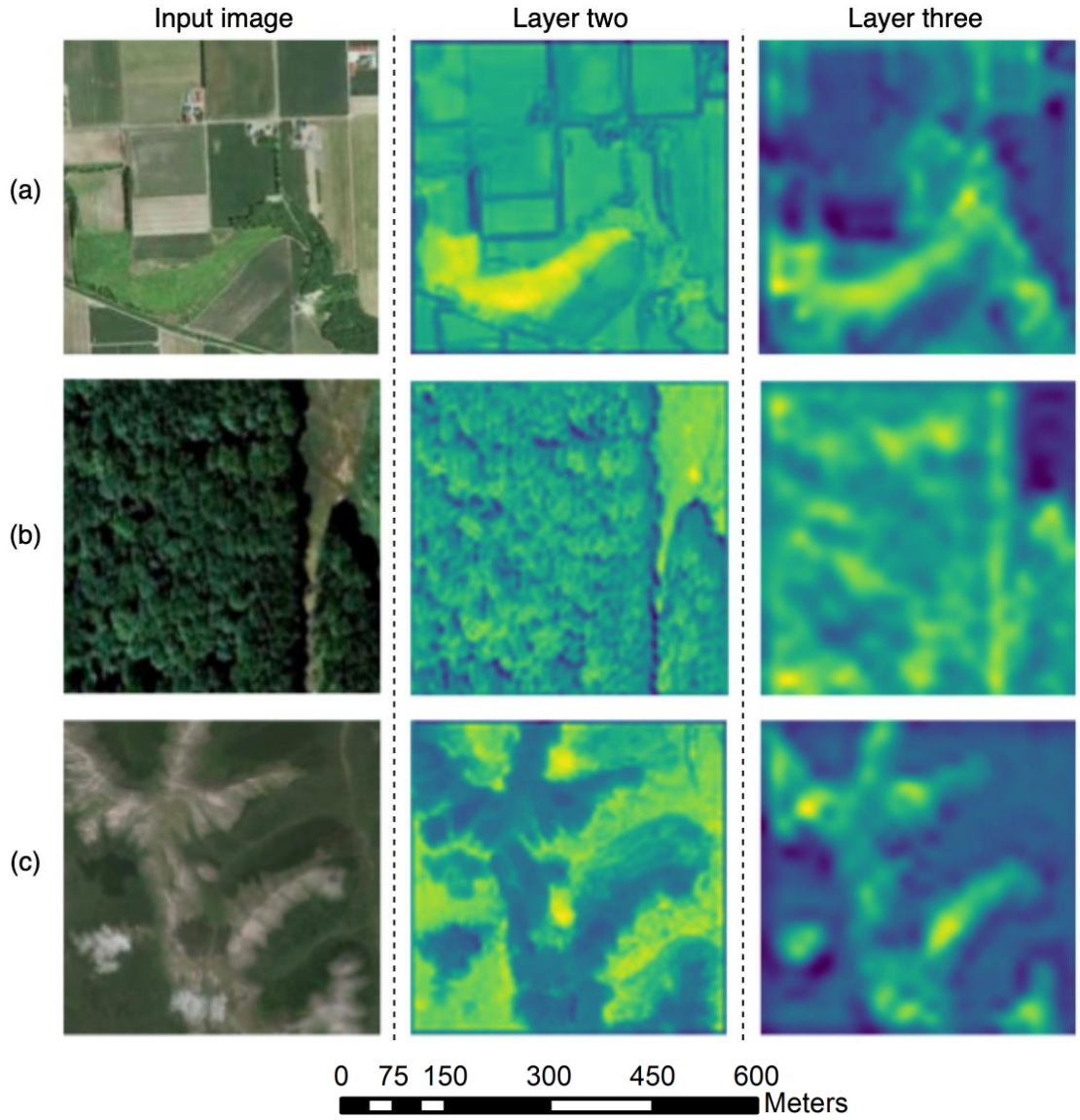


Figure 4.7 Original images and visualization of CNN feature maps reflecting their spatial resolution. Row (a) depicts agriculture landscapes, row (b) shows Forest landscapes, and row (c) represents Mountain landscapes. Column one shows input images. Columns two and three are the corresponding feature maps extracted from our Tex-CNN layers two and three, respectively. Note that the CNN features are eigen maps with the highest eigen values obtained after applying PCA to feature tensors in layers two and three.

4.5.3 Mountainous Terrains

We hypothesized that feature maps from within-landscape types would have lower EMD values, while those originating from disparate classes would have higher EMD values. We first conducted a Kolmogorov-Smirnov test to ascertain the validity of this hypothesis. As expected, it turns out that between-class feature distributions were statistically significantly different ($p < 0.05$). A sample of mountain landscapes from the AID and Sentinel datasets is depicted in Figure 4.8. Feature map regions that are highlighted in warmer colors represent the most significant discriminative patterns learned by the three filters; notice that most of these areas are predominantly less vegetated. Regions with cooler (blue) colors are found to be less significant, according to the model's weighting decision. Notice also that the filters sometimes perceive similar regions differently in terms of significant patterns – pixels that are found to be significant by one filter may be seen to have less weight by another filter, due to the discriminative learning behavior of CNNs.

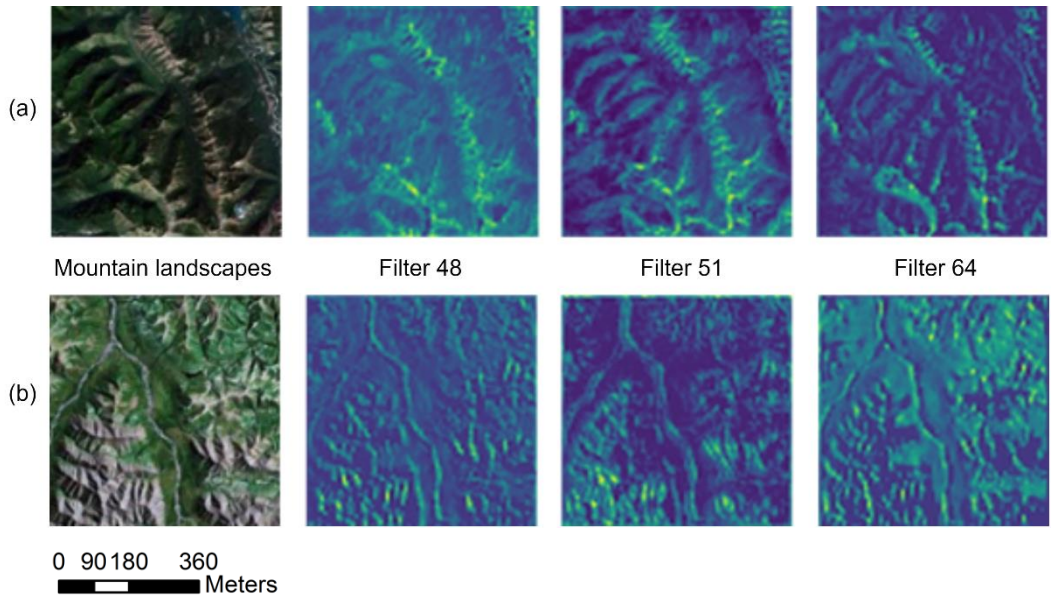


Figure 4.8 Mountain sample landscapes. Row (a) shows a sample mountain from Sentinel dataset. Row (b) shows a sample mountain from AID dataset. Feature maps are from Filters 48, 51, and 64.

Figure 4.9 shows the results for comparing mountainous landscapes and agriculture landscape types. It can be seen from Figure 4.9a,d that feature maps from similar landscapes display smaller distances, and hence their distribution falls to the left, characterized by smaller EMD. Over 60% of features in Wclass_mount of Figure 4.9a,b show EMD score of 0.01, while more than 50% of between class comparison yields EMD values higher than 0.05. Moreover, it can be observed that aside from shape differences, there is little overlap in the distributions of within class (Wclass_mount, Wclass_agricG1, and Wclass_agricG2).

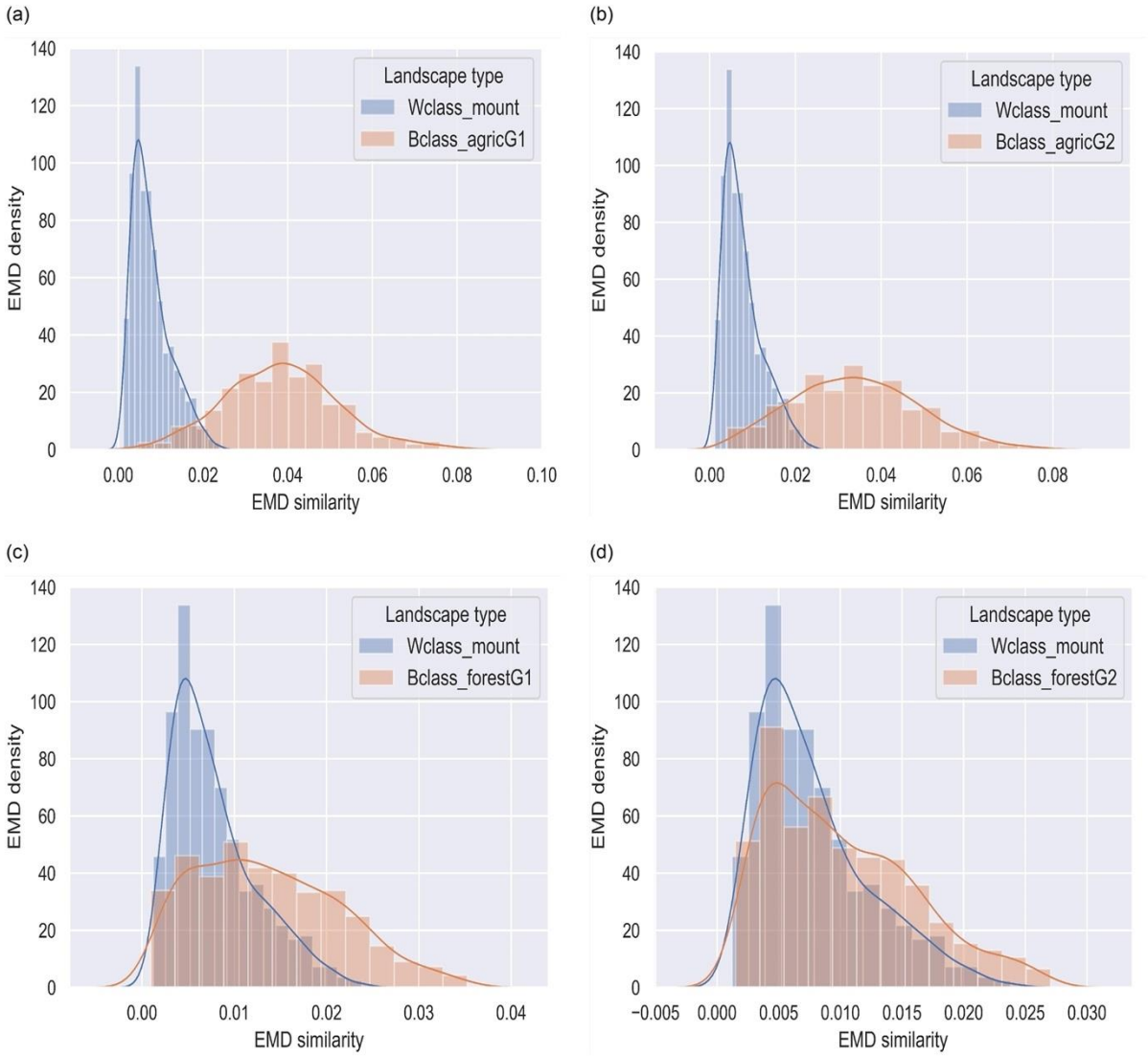


Figure 4.9 Landscape similarity comparison. EMD similarity distribution for mountain, forest, and agriculture patterns is depicted in (a–d). Mountain feature map comparison is within-class (i.e., mountain versus mountain). Between-landscape type similarity distribution is derived through mountain versus agriculture landscape (a,b), and mountain versus forest comparisons (c,d).

HoG can also be extracted directly from the original data (i.e., raw images) for comparison. We demonstrate this by computing EMD over the same set of original images used for extracting

CNN feature maps. Figure 4.10a,d presents within-class (i.e., Wclass_mount), between-class (i.e., (mountain versus agriculture landscape) and (mountain versus forest) EMD distributions. As can be seen in the derived CNN features, the mountain versus forest comparison poses challenges for real image comparison as well.

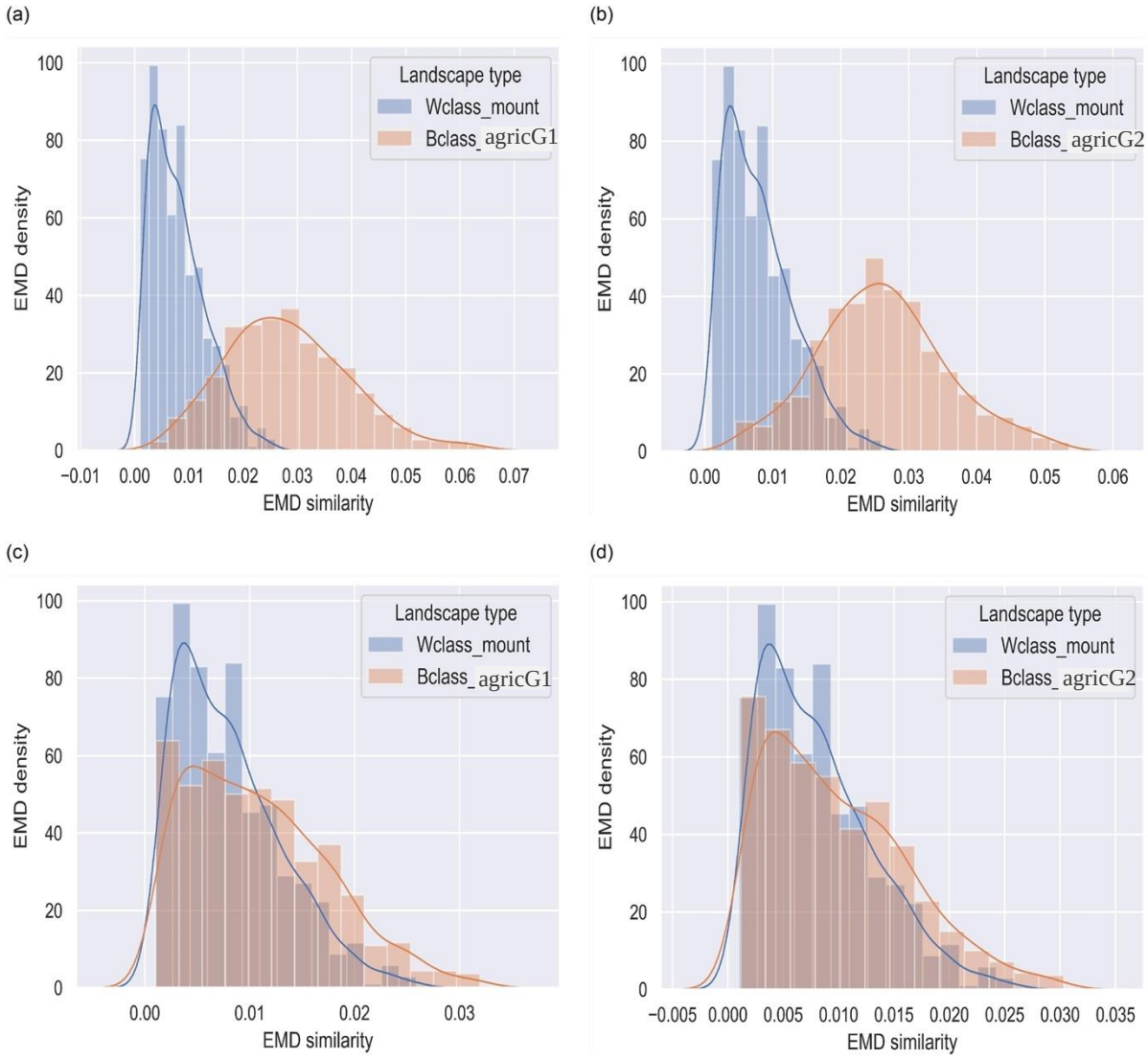


Figure 4.10 Original image histogram of oriented gradients (HoG) comparison. Image EMD values distribution for mountain, forest, and agriculture landscape patterns are depicted in (a–d). (a,b) Show within-class (mountain versus mountain) and between-class (mountain versus agriculture); meanwhile, (c,d) depict within-class (mountain versus mountain) and between-class (mountain versus forest).

4.5.4 Agriculture Landscapes

Figure 4.11 presents agriculture landscape samples and their corresponding feature maps. Convolutional filters are randomly selected to illustrate patterns learned on agriculture landscape types. It can be observed that the filters specialize in detecting different features. For example, Filter 43 recognizes agriculture landscape boundaries to be significant patterns, while Filter 8 weights blocks of vegetated areas higher. As shown in Figure 4.11a,b, the filters appear to assign significance to similar features in both AID and Sentinel datasets.

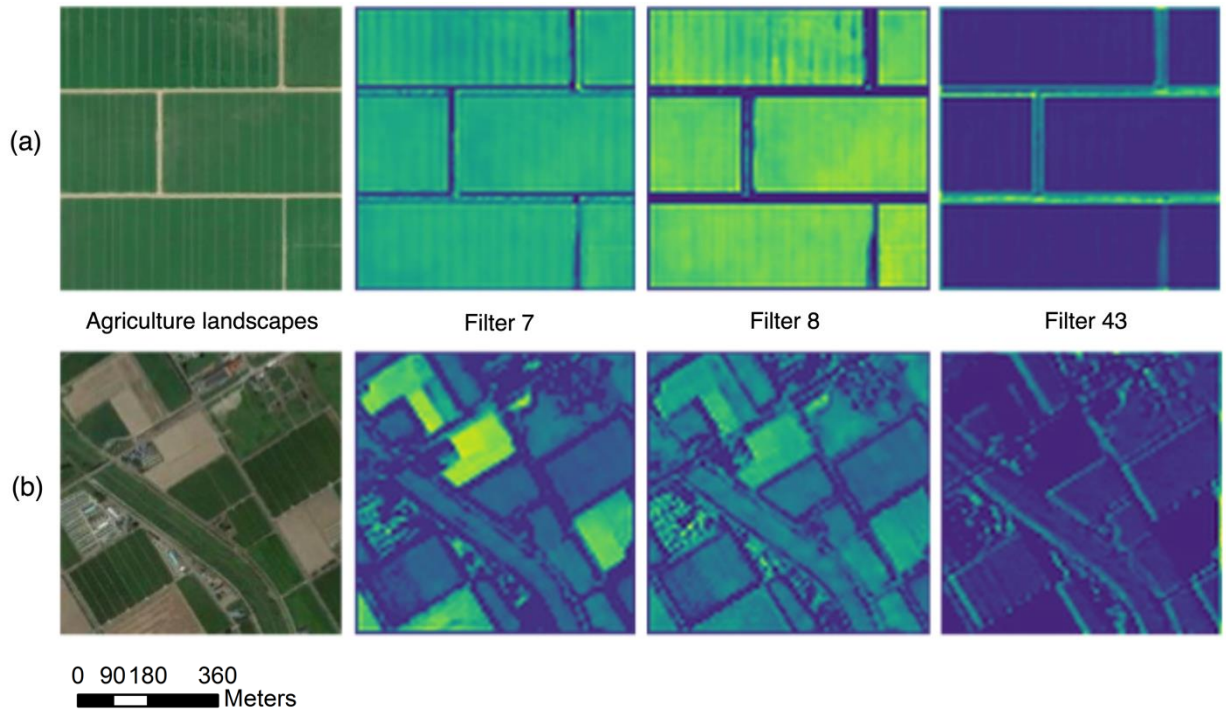


Figure 4.11 Agriculture landscapes and feature maps. Row (a) Sentinel dataset and row (b) AID dataset samples. Feature maps are extracted from Filters 7, 8, and 43. It can be seen that certain filters (e.g., Filter 43) specialize in detecting agriculture landscape boundaries, while Filters 7 and 8 detect regions with vegetation.

Figure 4.12a–d depicts within-landscape feature maps' similarity (Wclass_agric) and between-class similarity (Bclass_mountG1 and Bclass_mountG2, for mountains; Bclass_forestG1 and Bclass_forestG2, for forests). The Wclass_agric distribution shows most feature maps with EMD values close to zero, and over 65% of the feature maps show EMD values of 0.01. Conversely, Bclass_forestG1 and Bclass_mountG1 distributions tend to fall towards higher distances, with over 50% of feature maps having EMD value of 0.05.

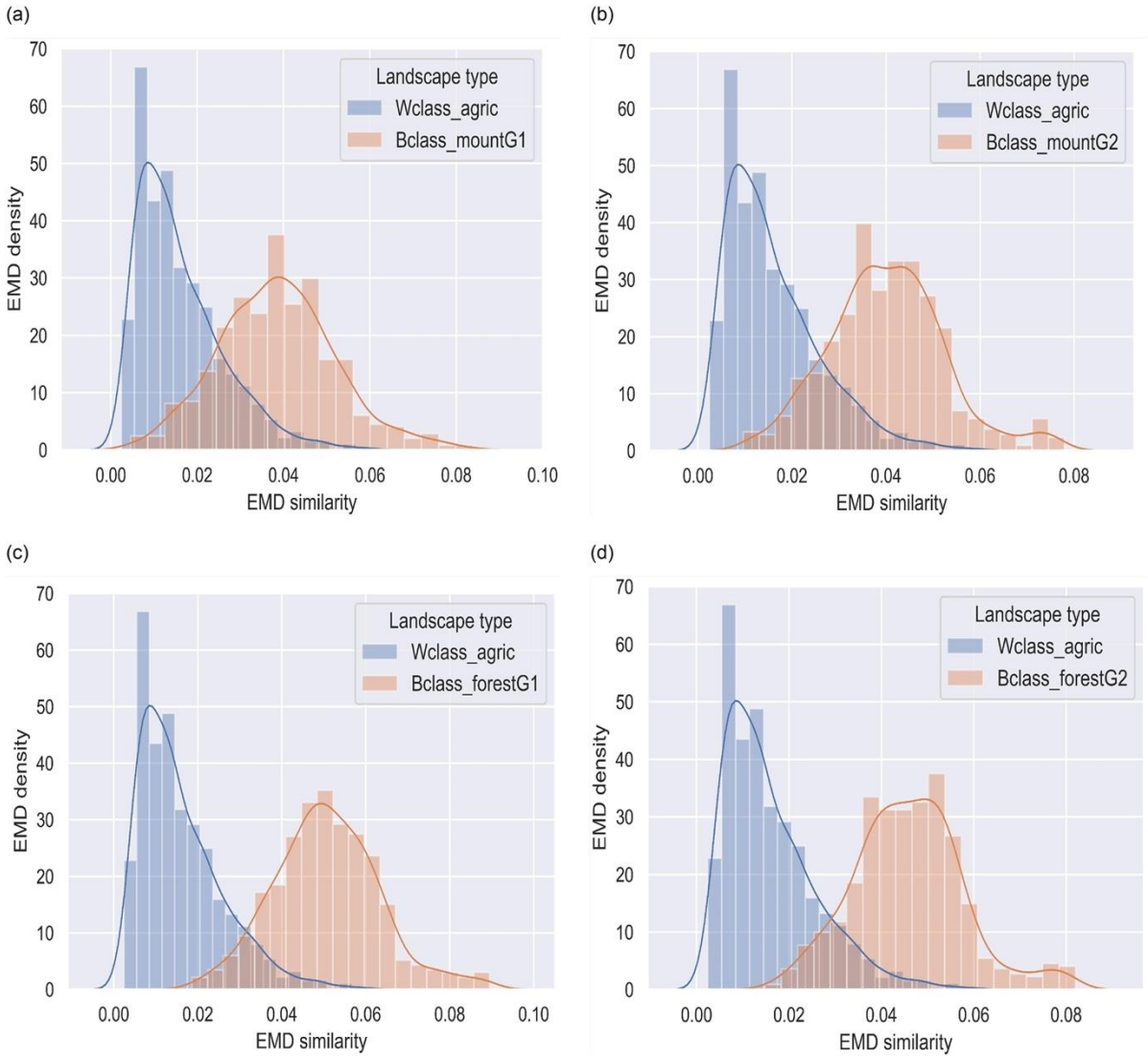


Figure 4.12 Landscape similarity comparison. Wclass_agric denotes within-agriculture landscape similarity. (a,b) Bclass_mountG1 and Bclass_mountG2 are distributions resulting from comparing agriculture landscape with mountains. (c,d) Bclass_forestG1 and Bclass_forestG2 are distributions generated by comparing agriculture landscapes with forests.

4.5.5 Forested Landscapes

Forest landscapes from the AID dataset and their feature maps are depicted in Figure 4.13a,b. Filters 11, 15, and 53 depict features at varying grain sizes, yet they represent discriminative features from an identical forest landscape.

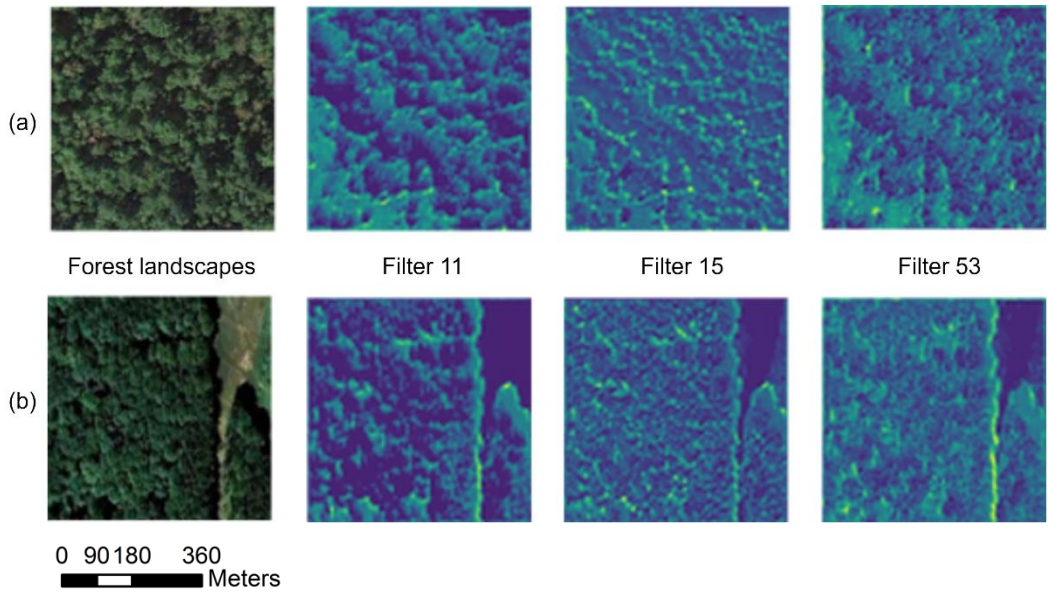


Figure 4.13 Forest landscapes from AID dataset. Row (a) denotes a sample image and its feature maps. Row (b) is sample of forest landscape from the different location. Notice that Filters 11, 15, and 53 depict features with varying grain sizes, yet they originate from an identical forest landscape.

Figure 4.14a,b illustrates the similarity distributions for within forest landscape (Wclass_forest) and forest versus agriculture landscape (Bclass_agricG1 and Bclass_agricG2). The two landscape types show distinct EMD similarity distribution with very little overlap. Moreover, high variance is noticeable in the between-landscape comparison. Feature maps in within-landscape comparison depict lower EMD scores, with over 60% of features showing

EMD values of 0.00–0.01, while over 70% feature maps in between-landscape comparison show 0.05 EMD similarity scores. Figure 4.14c,d compares forest landscapes with mountains. Within-class distribution (i.e., Wclass_forest) shows lower variance and relatively shorter EMD scores. However, though the distributions depict different shapes, there tend to be substantial overlap in within-class and between-class (Bclass_mountG1 and Bclass_mountG2) distributions.

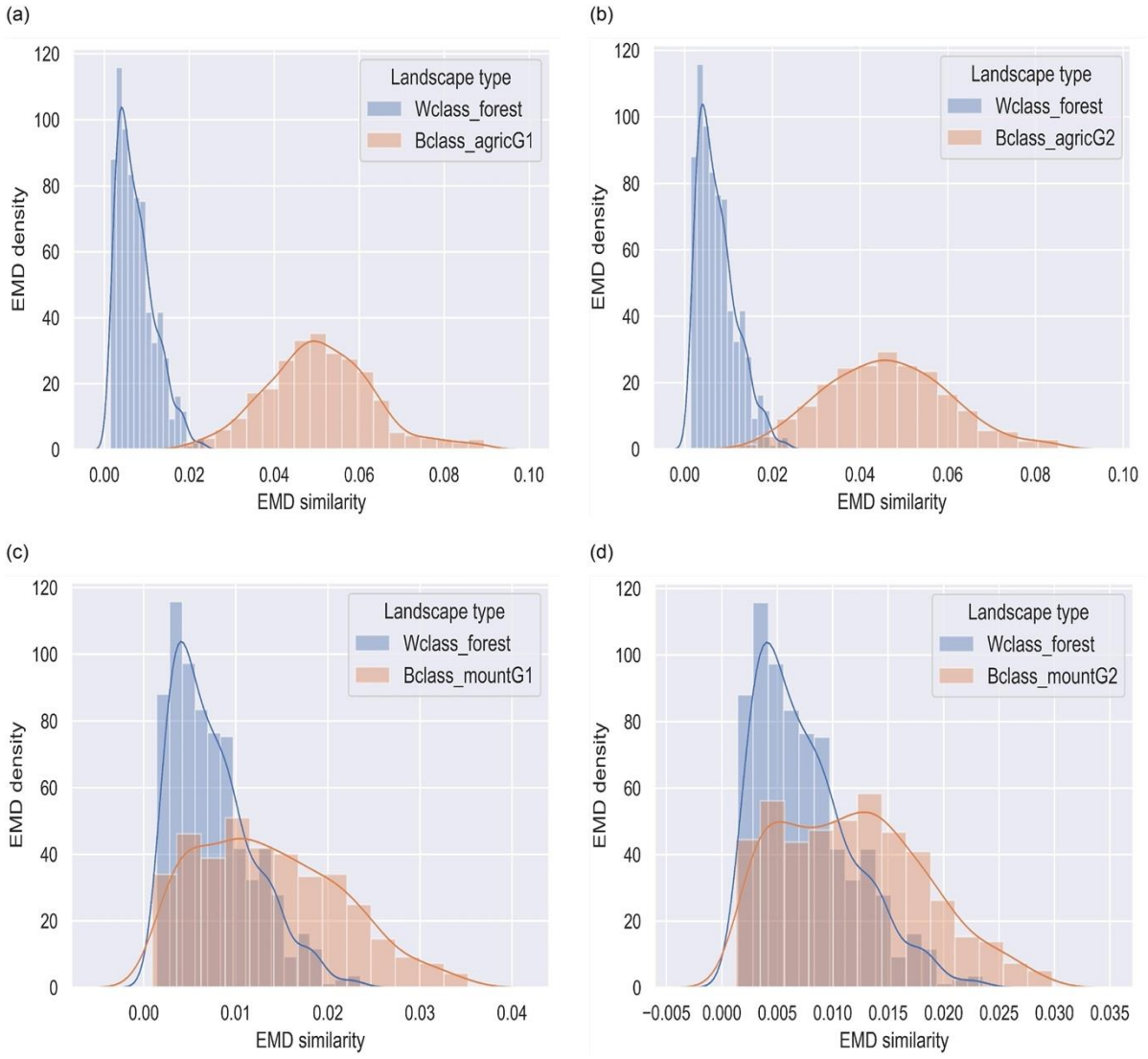


Figure 4.14 Landscape similarity comparison. (a,b) Distributions from within-landscape (Wclass_forest) and forest versus agriculture landscape types comparison (Bclass_agricG1 and Bclass_agricG2). (c,d) Distributions for forest versus mountain types comparison (Bclass_mountG1 and Bclass_mountG2).

4.6 Discussion

A comparison of the Tex-CNN accuracy reports on the test data, as shown in the confusion matrix (Figure 4.6), emphasizes the promising potential of textural information in capturing discriminative patterns. Although the performance of both models is virtually similar for agriculture and forested landscapes, we noticed a dramatic difference in the models' classification accuracies for mountainous terrain types. As seen in Figure 4.13, the feature maps display multi-resolution patterns in the forest landscape types. The feature concatenation method introduced may have encouraged the CNN to learn both fine and coarse grain spatial patterns (Basu et al., 2018). To illustrate the effectiveness of texture-encoded models, Table 4.3 presents the Tex-CNN classification results and that of the state-of-the-art models. Given that few landscape types were considered in this study, overall accuracy will not be a metric for an objective comparison; however, per-landscape accuracy shows that the Tex-CNN is competitive with other methods. The incorporation of texture features enhances model performance, especially for complex patterns (Andrearczyk and Whelan, 2016; Basu et al., 2018). It should be emphasized that the model is simple (i.e., small in size) and computationally efficient compared to other models (e.g., Huang and Xu, 2019; Anwer et al., 2018).

In Figure 4.6c,d, it can be observed that classifying landscapes in Sentinel data is challenging for both models, as they did not perform up to expectation in the first row of the confusion matrix. Over 90% of the agriculture landscapes tend to be classified as forest (i.e., false positive); contrarily, 92% and 79% of forest landscapes are correctly classified by the Tex-CNN and the classical CNN, respectively. This is partly explained by the relatively low spatial resolution of Sentinel data, as well as the data not being part of the training sample. Visual exploration of feature maps in Sentinel data shows most agriculture landscape boundaries disappearing

completely in higher layers of the CNN, thus making agriculture landscape samples appear as if they contain only vegetation patterns. The absence of boundary-like patterns likely triggers filter responses, leading to the misclassification of agriculture landscapes as forest. A profound reduction in misclassification rates, especially for agriculture landscapes, was achieved by adding Sentinel data (Figure 4.6e,f). Thus, presenting models with multimodal data at training time is likely to improve discriminative learning, while reducing misclassification errors.

The use of feature maps in pattern recognition is borne from the notion that the human visual system extracts the most relevant structural information from visual scenes in order to make decisions or characterize them semantically (Murabito et al., 2018). There is a great deal of analogy between landscape similarity comparison and assessment of feature maps similarity common in computer vision research (Reichstein et al., 2019). CNN feature maps are continuous-valued data which can avoid classification problems that arise in landscape research owing to landcover type discretization and artificial boundaries generation (Coops and Wulder, 2019). We adopted a novel approach to compare landscapes via the extraction of feature maps from specific landscape types. This framework leads to the availability sufficient feature templates describing a particular landscape and thus enabling robust similarity mapping. The PCA method resulted in objective selection of feature maps that best represent a given landscape. Feature map dimensionality reduction through PCA has been proven to not degrade but further improve the discriminative potential of convolutional features (Yandex and Lempitsky, 2015). Figure 4.7 shows samples of original images and their corresponding eigen maps. For landscape similarity comparison, layer-two feature maps were utilized. As can be seen in the figure, layer-two yields compact and high-resolution feature representations compared to layer-three. This

suggests that layer-two features may be suitable for similarity assessment, hence our adoption of the layer's feature maps.

In Figure 4.9a,b, mountainous landscapes show distinct differences with agriculture landscape types. The EMD values for the within class comparison (Wclass_mount) falls largely on the left, pointing to shorter distances and hence higher similarity. More than 60% of the feature maps show EMD values of 0.01. Over 50%, the feature maps between class comparison EMD values are as high as 0.05. This suggests that there exist significant discriminative features between these two distinct landscape types. Song et al. (2018) provide evidence that, by using feature map distances, it is possible to select the most discriminative patterns to represent mountainous terrains. The feature maps within the similar landscape also tend to depict higher EDM densities, which is an indicator of feature maps clustering (Zhou et al., 2016), and high-density (frequency) values imply that a large proportion of feature maps are similar. Figure 4.10a–d compares HoG features extracted directly from the original images. The EDM distributions are somewhat similar to the CNN feature maps, but it can be observed that the CNN features appear to be slightly sensitive; for example, fewer images in the between-class comparison fall in EDM of 0 – 0.01. Moreover, compared to the original image HoG features, it can be seen that EMD values' distribution tends to be peakier for within-class and a little flatter for between-class in the CNN feature comparison. This suggests that our Tex-CNN features may possess more image descriptors compared to raw image pixels.

When comparing mountains versus forested landscapes, EMD distributions appear to overlap. This challenge is not unexpected, given the diverse morphology of mountains in some images, especially given that some mountains contain forest. Furthermore, recalling that the model's performance at predicting mountainous terrains is low, it follows that feature maps derived for

certain input images that record poor scores may be of lower quality for landscape comparison. This suggests that, if a model is optimized to predict a particular landscape type with high accuracy, its corresponding feature maps will be of better discriminative quality and hence can be suitable for mapping landscape similarity (Zagoruyko and Komodakis, 2017).

Agriculture landscapes turn out to be the most easily discriminated patterns using the CNN model's feature maps (see Figure 4.12 a,b). As expected, the within-landscape type comparison shows smaller EMD values for agriculture landscape feature maps, with between-landscape distributions falling towards the right. Additionally, there is very little overlap in the distributions of within- and between-feature map distributions. Higher EMD values suggest lower similarity scores for landscapes being compared. Moreover, within-class feature maps exhibit somewhat low variance in EMD values. Over 65% of the Wclass_agric shows 0.01 EMD. This shows higher similarity compared to agriculture landscape versus forest comparison, where EMD values as large as 0.05 are recorded. The unique vertical and horizontal boundary features may be among the discriminative patterns the model learns in agriculture landscapes. Lower layers of CNN are superior in learning edges, blobs, curves, and fine-grained textural patterns (Grinblat et al., 2016). This observation emphasizes the high prediction accuracy recorded for the agriculture landscape type, as shown in the confusion matrix (Figure 4.6). Murabito et al. (2018) study found that saliency maps, a variant of gradient-based attention maps (i.e., feature maps), improve pattern detection.

Figure 4.14a–d depicts within-forest landscape and between landscapes, which consist of forest versus agriculture landscape (e.g., Bclass_agricG1), and forest versus mountain (e.g., Bclass_mountG1). Figure 4.14a,b emphasizes the existence of distinct discriminative features between forest and agriculture landscapes, as these two distributions show very little overlap.

More importantly, within-forest landscape (Wclass_forest) distribution shows lower EMD values, suggesting higher similarity scores. More than 60% of the feature maps have EMD values of 0.00–0.01, while over 70% of the between-landscape comparison shows 0.05 EMD similarity scores. However, the Wclass_forest versus Bclass_mount distributions show overlaps (Figure 4.14c,d), though the shape of the distributions suggest that the two landscapes belong to distinctively different class types. The Kolmogorov–Smirnov test further confirmed that the distributions are statistically significantly different ($p - value < 0.05$).

The remote-sensing and spatial-analysis literature has many metrics for comparing spatial patterns, yet this domain is largely fractured, and sometimes lacks generic toolsets for comparing continuous valued (i.e., unclassified) image data (Long and Robertson, 2018). Amirshahi et al. (2017) proposed extracting HoG and applying histogram intersection kernel to compare feature maps. Liu et al. (2020) also introduced a similarity distribution learning framework, using a CNN ensemble to incorporate feature uncertainty similarity at training time. The extracted features from the trained model are then employed in image retrieval and scene classification. Given that CNN feature maps are inherently discriminative and can potentially handle similarity uncertainties, we propose a metric to compare CNN feature maps' similarity via the computation of feature EMD. Our approach applies gradient-based computation to extract discriminative spatial patterns given an input image. The extracted feature maps contain local descriptors which are essential for pattern recognition. Utilizing EMD resolves the problem of histograms' bin size on similarity scores.

Our proposed metric effectively distinguishes agriculture landscape types from non-agriculture landscapes. Mountainous terrains and forested landscapes are discriminated, as their distributions are significantly different. A highly sensitive spatial pattern domain metric may be

able to overcome the overlaps seen in forested and mountainous landscapes distributions. We tested structural similarity and the complex-wavelet structural similarity metrics which capture spatial information but did not realize impressive results. We point out that our findings demonstrate the challenging nature of the AID dataset and its potential suitability for training models; despite containing fewer samples per scene categories, the images can be described as multi-scale (i.e., mountain patterns vary within the same landscape type). Such data can present challenges to CNNs without explicit multi-resolution encoding (Liu et al., 2018b). To surmount such a limitation, Li et al. (2018) suggested utilizing the last convolutional layer feature maps, since these enable the discovery of locally consistent spatial patterns. However, we chose not to apply these features, since they lack full geometric invariance, as well as fine-grain textural details (Gong et al., 2014). Figure 4.8 further emphasizes our claim, as it illustrates the low spatial resolution of layer-three feature maps. The last layer (i.e., FC1) encodes structure and global information (e.g., shape). As pointed out earlier, unlike object recognition, landscape patterns lack definite shapes; hence, features from this layer may not improve mountain versus forest discrimination substantially. Furthermore, given that the FC1 features are 1D vectors, the approach to computing the HoG adopted cannot be applied. The bag-of-words approach widely used in CBIR (Ahmed et al., 2019) could improve mountain versus forest distinction, but this approach was not considered in this work, as it is out of scope. The low classification accuracy of the models on Sentinel data (see Figure 4.6c,d) emphasizes the potential effects of spatial resolution on models' performance. Interestingly, however, the Tex-CNN outperforms the classical CNN, as it shows high classification accuracy for mountains. The inclusion of texture information may have improved the model's performance across scales.

4.7 Conclusion

The landscape-similarity mapping problem can be formulated as a challenge to detect repeated patterns, in other words, similar patterns across different locations, as shown in a study conducted by Lettry et al. (2017). The problem of comparing landscapes can also be considered in the context of image-retrieval tasks, as demonstrated by Yandex and Lempitsky (2015), using convolutional feature maps. Landscape similarity or change-detection problems may further be cast as image-quality assessment challenges, as demonstrated in Amirshahi et al. (2017). In this study, we showed that CNN-based features (aka spatial attention maps) contain discriminative descriptors of image quality and, hence, computing similarity over feature maps can be an effective and generic way to compare landscapes. Our approach provides evidence that a generic pattern-comparison metric can be developed from highly discriminative feature maps capable of mapping diverse landscape types.

The challenge encountered in the mixing of forest and mountain similarity distributions points to the potential occurrence of false positives when attempting to make search queries between forests and mountains. The models' performance being consistently low for mountains in AID and Sentinel data further emphasizes that scaling of features represented in feature maps might work for agriculture landscapes and forests but not for mountains. As mentioned previously, the morphology of the mountain class is highly variable; moreover, the presence of forest on mountains further complicates discrimination between the landscapes. In this context, a priori knowledge may help decrease false positives at the time of query.

One potential limitation of the proposal stems from the fact that mixed landscape samples were not considered in model development; widening the sample size to include scenes that contain a mixture of two or more landcover types could improve the metric's performance,

especially in discriminating mountains and forests. Such a fuzzy definition of landscape classes may be more useful for landscape-similarity and/or scene-retrieval applications in the future, as they more closely align with the complexity of landscapes in the real world. Additionally, the nested framework (i.e., PCA and HoG, and EMD) computations may increase the complexity of the proposed metric. Given that what constitutes the best approach to feature map selection approach remains an open question (Liu et al., 2015), an innovative and objective framework to select feature maps to enhance similarity detection, as shown by Rui et al. (2017) by utilizing feature map separability index, needs future consideration. Also, further research needs to consider expanding the number of landscape types to test the robustness and generalizability of the proposed metric. Independent validation datasets from different sensors, such as Sentinel, can be challenging for models trained on high-resolution aerial imagery; thus, it is essential that future research considers combining samples of multi-modal datasets for model development. The utilization of gradient-based CNN feature maps for landscape-change detection also warrants future research.

Author contributions: K.M. contributed to implementing modeling concepts, as well as writing the paper. C.R. contributed to developing modeling concepts and writing and reviewing the paper.

Acknowledgements

The authors would like to acknowledge funding provided by the Natural Sciences and Engineering Research Council, Canada to Dr. Robertson.

References

Ahmed, K. T., Ummesafi, S., & Iqbal, A. (2019). Content based image retrieval using image features information fusion. *Information Fusion*, *51*, 76–99.

Albert, A., Kaur, J., & Gonzalez, M. (2017). Using convolutional networks and satellite imagery to identify patterns in urban environments at a large scale. *In Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 1357–1366.

Amirshahi, S. A., Pedersen, M., & Yu, S. X. (2017). Image quality assessment by comparing CNN features between images. *IS and T International Symposium on Electronic Imaging Science and Technology*, 42–51.

Andrearczyk, V., & Whelan, P. F. (2016). Using filter banks in Convolutional Neural Networks for texture classification. *Pattern Recognition Letters*, *84*, 63–69.

Anwer, R. M., Khan, F. S., van de Weijer, J., Molinier, M., & Laaksonen, J. (2018). Binary patterns encoded convolutional neural networks for texture recognition and remote sensing scene classification. *ISPRS Journal of Photogrammetry and Remote Sensing*, *138*, 74–85.

Basu, S., Mukhopadhyay, S., Karki, M., DiBiano, R., Ganguly, S., Nemani, R., & Gayaka, S. (2018). Deep neural networks for texture classification—A theoretical analysis. *Neural Networks*, *97*, 173–182.

Buscombe, D., & Ritchie, A. C. (2018). Landscape classification with deep neural networks. *Geosciences (Switzerland)*, *8*(7), 1–23.

Cao, J., Liu, L., Wang, P., Huang, Z., Shen, C., & Shen, H. T. (2016). *Where to Focus: Query Adaptive Matching for Instance Retrieval Using Convolutional Feature Maps*. 1–10.

Chen, J., Zhou, Z., Pan, Z., & Yang, C. (2019). Instance Retrieval Using Region of Interest Based CNN Features. *Journal of New Media*, *1*(2), 87–99.

Cimpoi, M., Maji, S., Kokkinos, I., & Vedaldi, A. (2016). Deep Filter Banks for Texture Recognition, Description, and Segmentation. *International Journal of Computer Vision*, *118*(1), 65–94.

Comber, A., & Wulder, M. (2019). Considering spatiotemporal processes in big data analysis: Insights from remote sensing of land cover and land use. *Transactions in GIS*, *23*(5), 879–891.

Coops, N. C., & Wulder, M. A. (2019). Breaking the Habit(at). *Trends in Ecology & Evolution*, *34*(7), 585–587.

Dandois, J. P., & Ellis, E. C. (2013). High spatial resolution three-dimensional mapping of vegetation spectral dynamics using computer vision. *Remote Sensing of Environment*, *136*, 259–276.

Deniz, O., Bueno, G., Salido, J., & De La Torre, F. (2011). Face recognition using Histograms of Oriented Gradients. *Pattern Recognition Letters*, *32*(12), 1598–1603.

El Amin, A. M., Liu, Q., & Wang, Y. (2016). Convolutional neural network features based change detection in satellite images. *First International Workshop on Pattern Recognition. International Society for Optics and Photonics*, *10011*, 100110W.

Flores, C. F., Gonzalez-Garcia, A., van de Weijer, J., & Raducanu, B. (2019). Saliency for fine-grained object recognition in domains with scarce training data. *Pattern Recognition*, *94*, 62–73.

Fu, G., Liu, C., Zhou, R., Sun, T., & Zhang, Q. (2017a). Classification for high resolution remote sensing imagery using a fully convolutional network. *Remote Sensing*, *9*(5), 1–21.

Gahegan, M. (2020). Fourth paradigm GIScience? Prospects for automated discovery and explanation from data. *International Journal of Geographical Information Science*, *34*(1),

1–21.

Gao, Q., Lim, S., & Jia, X. (2018). Hyperspectral image classification using convolutional neural networks and multiple feature learning. *Remote Sensing*, *10*(2).

Gatys, L. A., Ecker, A. S., & Bethge, M. (2017). Texture and art with deep neural networks. *Current Opinion in Neurobiology*, *46*, 178–186. https://doi.org/10.1016/j.conb.2017.08.019

Ghorbanzadeh, O., Blaschke, T., Gholamnia, K., Meena, S. R., Tiede, D., & Aryal, J. (2019). Evaluation of different machine learning methods and deep-learning convolutional neural networks for landslide detection. *Remote Sensing*, *11*(2).

Girdhar, R., & Ramanan, D. (2017). Attentional pooling for action recognition. *Advances in Neural Information Processing Systems, 2017-Decem*, 34–45.

Gong, X., Xie, Z., Liu, Y., Shi, X., & Zheng, Z. (2018). Deep salient feature based anti-noise transfer network for scene classification of remote sensing imagery. *Remote Sensing*, *10*(3).

Gong, Y., Wang, L., Guo, R., & Lazebnik, S. (2014). Multi-Scale Orderless Pooling of Deep Convolutional Activation Features. *In European Conference on Computer Vision*, 392–407.

Grinblat, G. L., Uzal, L. C., Larese, M. G., & Granitto, P. M. (2016). Deep learning for plant identification using vein morphological patterns. *Computers and Electronics in Agriculture*, *127*, 418–424.

Gu, Y., Wang, Y., & Li, Y. (2019). A survey on deep learning-driven remote sensing image scene understanding: Scene classification, scene retrieval and scene-guided object detection. *Applied Sciences (Switzerland)*, *9*(10), 2110.

Hinton, G. E., Srivastava, N., Krizhevsky, A., Sutskever, I., & Salakhutdinov, R. R. (2012). Improving neural networks by preventing co-adaptation of feature detectors. 1–18.

Huang, H., & Xu, K. (2019). Combing triple-part features of convolutional neural networks for

scene classification in remote sensing. *Remote Sensing*, *11*(14), 1687.

Janowicz, K., Gao, S., McKenzie, G., Hu, Y., & Bhaduri, B. (2020). GeoAI: spatially explicit artificial intelligence techniques for geographic knowledge discovery and beyond. *International Journal of Geographical Information Science*, *34*(4), 625–636.

Jasiewicz, J., Netzel, P., & Stepinski, T. F. (2014). Landscape similarity, retrieval, and machine mapping of physiographic units. *Geomorphology*, *221*, 104–112.

Kalantar, B., Ueda, N., Al-Najjar, H. A. H., & Halin, A. A. (2020). Assessment of convolutional neural network architectures for earthquake-induced building damage detection based on pre-and post-event orthophoto images. *Remote Sensing*, *12*(21), 1–20.

Kang, M., Ji, K., Leng, X., & Lin, Z. (2017). Contextual region-based convolutional neural network with multilayer fusion for SAR ship detection. *Remote Sensing*, *9*(8).

Kingma, D. P., & Ba, J. L. (2015). Adam: A method for stochastic optimization. *3rd International Conference on Learning Representations, ICLR 2015 - Conference Track Proceedings*, 1–15.

Lettry, L., Perdoch, M., Vanhoey, K., & Van Gool, L. (2017). Repeated pattern detection using CNN activations. *Proceedings - 2017 IEEE Winter Conference on Applications of Computer Vision, WACV 2017*, 47–55.

Li, H., Ellis, J. G., Zhang, L., & Chang, S. F. (2018). PatternNet: Visual pattern mining with deep neural network. *ICMR 2018 - Proceedings of the 2018 ACM International Conference on Multimedia Retrieval*, 291–299.

Li, Z., White, J. C., Wulder, M. A., Hermosilla, T., Davidson, A. M., & Comber, A. J. (2020 a). Land cover harmonization using Latent Dirichlet Allocation. *International Journal of Geographical Information Science*, *00*(00), 1–27.

Lim, L. A., & Keles, H. Y. (2020). Learning multi-scale features for foreground segmentation. *Pattern Analysis and Applications*, *23*(3), 1369–1380.

Liu, L., Shen, C., & Van Den Hengel, A. (2015). The treasure beneath convolutional layers: Cross-convolutional-layer pooling for image classification. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 4749–4757.

Liu, Yanfei, Zhong, Y., & Qin, Q. (2018b). Scene classification based on multiscale convolutional neural network. *IEEE Transactions on Geoscience and Remote Sensing*, *56*(12), 7109–7121.

Liu, Yanfei, Zhong, Y., Fei, F., Zhu, Q., & Qin, Q. (2018c). Scene classification based on a deep random-scale stretched convolutional neural network. *Remote Sensing*, *10*(3).

Liu, Yazhou, Cao, G., Sun, Q., & Siegel, M. (2015). Hyperspectral classification via deep networks and superpixel segmentation. *International Journal of Remote Sensing*, *36*(13), 3459–3482.

Liu, Yishu, Han, Z., Chen, C., DIng, L., & Liu, Y. (2020). Eagle-Eyed Multitask CNNs for Aerial Image Retrieval and Scene Classification. *IEEE Transactions on Geoscience and Remote Sensing*, *58*(9), 6699–6721.

Long, J., & Robertson, C. (2018). Comparing spatial patterns. *Geography Compass*, *12*(2), 1–18.

Luo, J. H., Zhang, H., Zhou, H. Y., Xie, C. W., Wu, J., & Lin, W. (2018). ThiNet: Pruning CNN Filters for a Thinner Net. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *41*(10), 2525–2538.

Mahendran, A., & Vedaldi, A. (2014). *Understanding Deep Image Representations by Inverting Them*. 2014.

Miller, H. J., & Goodchild, M. F. (2015). Data-driven geography. *GeoJournal*, *80*(4), 449–461.

Murabito, F., Spampinato, C., Palazzo, S., Giordano, D., Pogorelov, K., & Riegler, M. (2018). Top-down saliency detection driven by visual classification. *Computer Vision and Image Understanding*, *172*, 67–76.

Nogueira, K., Penatti, O. A. B., & dos Santos, J. A. (2017). Towards better exploiting convolutional neural networks for remote sensing scene classification. *Pattern Recognition*, *61*, 539–556.

Omeiza, D., Speakman, S., Cintas, C., & Weldermariam, K. (2019). Smooth Grad-CAM++: An Enhanced Inference Level Visualization Technique for Deep Convolutional Neural Network Models. 1–10.

Peng, F., Wang, L., Zou, S., Luo, J., Gong, S., & Li, X. (2019). Content-based search of earth observation data archives using open-access multitemporal land cover and terrain products. *International Journal of Applied Earth Observation and Geoinformation*, *81*, 13–26.

Petrovska, B., Zdravevski, E., Lameski, P., Corizzo, R., Štajduhar, I., & Lerga, J. (2020). Deep learning for feature extraction in remote sensing: A case-study of aerial scene classification. *Sensors (Switzerland)*, *20*(14), 1–22.

Qi, X., Li, C. G., Zhao, G., Hong, X., & Pietikäinen, M. (2016). Dynamic texture and scene classification by transferring deep image features. *Neurocomputing*, *171*, 1230–1241.

Quoc Bao, T., Tan Kiet, N. T., Quoc Dinh, T., & Hiep, H. X. (2020). Plant species identification from leaf patterns using histogram of oriented gradients feature space and convolution neural networks. *Journal of Information and Telecommunication*, *4*(2), 140–150.

Reichstein, M., Camps-Valls, G., Stevens, B., Jung, M., Denzler, J., Carvalhais, N., & Prabhat. (2019). Deep learning and process understanding for data-driven Earth system science. *Nature*, *566*(7743), 195–204.

Rubner, Y., Tomasi, C., & Guibas, L. J. (2000). Earth mover's distance as a metric for image retrieval. *International Journal of Computer Vision*, *40*(2), 99–121.

Rui, T., Zou, J., Zhou, Y., Fei, J., & Yang, C. (2017). Convolutional neural network feature maps selection based on LDA. *Multimedia Tools and Applications*, *1*, 1–15.

Selvaraju, R. R., Cogswell, M., Das, A., Vedantam, R., Parikh, D., & Batra, D. (2017). Grad-CAM: Visual Explanations from Deep Networks via Gradient-Based Localization. *In Proceedings of the IEEE International Conference on Computer Vision*, 618–626.

Shi, X., & Qian, X. (2019). Exploring spatial and channel contribution for object based image retrieval. *Knowledge-Based Systems*, *186*, 104955.

Simonyan, K., Vedaldi, A., & Zisserman, A. (2013). *Deep Inside Convolutional Networks: Visualising Image Classification Models and Saliency Maps*. 1–8.

Simonyan, K., & Zisserman, A. (2015). Very Deep Convolutional Networks for Large-Scale Image Recognition. *ICLR 2015*, 1–14.

Song, F., Yang, Z., Gao, X., Dan, T., Yang, Y., Zhao, W., & Yu, R. (2018). Multi-scale feature based land cover change detection in mountainous terrain using multi-temporal and multi-sensor remote sensing images. *IEEE Access*, *6*, 77494–77508.

Springenberg, J. T., Dosovitskiy, A., Brox, T., & Riedmiller, M. (2014). *Striving for Simplicity: The All Convolutional Net*. 1–14.

Townshend, J. R., Masek, J. G., Huang, C., Vermote, E. F., Gao, F., Channan, S., Sexton, J. O., Feng, M., Narasimhan, R., Kim, D., Song, K., Song, D., Song, X. P., Noojipady, P., Tan, B., Hansen, M. C., Li, M., & Wolfe, R. E. (2012). Global characterization and monitoring of forest cover using Landsat data: Opportunities and challenges. *International Journal of Digital Earth*, *5*(5), 373–397.

Tracewski, L., Bastin, L., & Fonte, C. C. (2017). Repurposing a deep learning network to filter and classify volunteered photographs for land cover and land use characterization. *Geo-Spatial Information Science*, *20*(3), 252–268.

Turner, M. G. (1989a). Landscape Ecology : The Effect of Pattern on Process. *Annual Review of Ecology and Systematics*, *20*, 171–197.

Unar, S., Wang, X., Wang, C., & Wang, Y. (2019). A decisive content based image retrieval approach for feature fusion in visual and textual images. *Knowledge-Based Systems*, *179*, 8–20.

Ustyuzhaninov, I., Brendel, W., Gatys, L. A., & Bethge, M. (2016). *Texture Synthesis Using Shallow Convolutional Networks with Random Filters*. 1–9.

Wang, A., Wang, Y., & Chen, Y. (2019). Hyperspectral image classification based on convolutional neural network and random forest. *Remote Sensing Letters*, *10*(11), 1086–1094.

Wulder, M. A., Coops, N. C., Roy, D. P., White, J. C., & Hermosilla, T. (2018). Land cover 2.0. *International Journal of Remote Sensing*, *39*(12), 4254–4284.

Xia, G., Hu, J., Hu, F., & Shi, B. (2016). AID : A Benchmark Dataset for Performance Evaluation of Aerial Scene Classification. *IEEE Transactions on Geoscience and Remote Sensing*, *55*(7), 3965–3981.

Xie, X., Han, X., Liao, Q., & Shi, G. (2017). Visualization and pruning of SSD with the base network VGG16. *ACM International Conference Proceeding Series*, 90–94.

Xu, K., Huang, H., Li, Y., & Shi, G. (2020). Multilayer Feature Fusion Network for Scene Classification in Remote Sensing. *IEEE Geoscience and Remote Sensing Letters*, *17*(11), 1894–1898.

Yandex, A. B., & Lempitsky, V. (2015). Aggregating Deep Convolutional Features for Image Retrieval. *Proceedings of the IEEE International Conference on Computer Vision*, 1269–1277.

Yang, B., Yan, J., Lei, Z., & Li, S. Z. (2015). Convolutional channel features. *In Proceedings of the IEEE International Conference on Computer Vision*, 82–90.

Ye, L., Wang, L., Sun, Y., Zhao, L., & Wei, Y. (2018b). Parallel multi-stage features fusion of deep convolutional neural networks for aerial scene classification. *Remote Sensing Letters*, *9*(3), 294–303.

Yu, Y., & Liu, F. (2018). Aerial Scene Classification via Multilevel Fusion Based on Deep Convolutional Neural Networks. *IEEE Geoscience and Remote Sensing Letters*, *15*(2), 287–291.

Zagoruyko, S., & Komodakis, N. (2017). Paying more attention to attention: Improving the performance of convolutional neural networks via attention transfer. *5th International Conference on Learning Representations*, 1–13.

Zeiler, M. D., & Fergus, R. (2014). Visualizing and understanding convolutional networks. *In European Conference on Computer Vision*, 818–833.

Zeng, D., Chen, S., Chen, B., & Li, S. (2018). Improving remote sensing scene classification by integrating global-context and local-object features. *Remote Sensing*, *10*(5), 1–19.

Zhang, H., Zhang, T., Pedrycz, W., Zhao, C., & Miao, D. (2019). Improved adaptive image retrieval with the use of shadowed sets. *Pattern Recognition*, *90*, 390–403.

Zhou, B., Khosla, A., Lapedriza, A., Oliva, A., & Torralba, A. (2016). Learning Deep Features for Discriminative Localization. *In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2921–2929.

Zhu, Q., Zhong, Y., Liu, Y., Zhang, L., & Li, D. (2018). A Deep-local-global feature fusion framework for high spatial resolution imagery scene classification. *Remote Sensing*, *10*(4).

Zhuang, S., Wang, P., Jiang, B., Wang, G., & Wang, C. (2019). A single shot framework with multi-scale feature fusion for geospatial object detection. *Remote Sensing*, *11*(5).

Chapter 5: U-Net convolutional neural network models for detecting and quantifying placer mining disturbances at watershed scale

5.1 Abstract

Placer mining is a mineral extraction method in floodplains that involves the removal of earth material to access mineral-laden sediments, a process that can have significant and long-term impacts on aquatic ecosystems. Given the widespread nature of mining, new tools are required to monitor the potential watershed-scale ecological impacts of placer mining. This study adapted and evaluated a deep learning model – a U-Net convolution neural network – and compared it to a traditional image classification method – random forests (RF) – to detect and quantify the area of post-placer mining disturbance at the watershed scale. Both modelling frameworks achieved at least 75% accuracy in the classification of digitized (i.e., labelled placer pixels) placer samples in 7 out of 12 modelling scenarios. Misclassification of non-placer pixels as placer was highly variable among different models, data configurations, study sites, and time periods. The addition of vegetation index channels – Normalized Difference Vegetation Index and Soil Adjusted Vegetation Index as model inputs tended to be more important for U-Net models than RF models. In general, U-Net models performed better in terms of minimizing misclassification errors, whereas RF models performed slightly better in classifying known placer pixels. We conclude with discussions on the advantages of deploying U-Net and RF models for placer detection, and identifying outstanding issues which need to be addressed in future placer modelling studies.

Keywords: Placer mining, placer disturbance detection, U-Net, semantic segmentation, Random Forest

5.2 Introduction

The Earth's water resources are increasingly being exposed to human-induced disturbances from global to local scales. Human disturbances of fluvial ecosystems can degrade water quality, altering adjacent and downstream fish and aquatic invertebrate communities as water is delivered from headwaters to oceans (Matthews, 2016). Spatial-temporal changes in land use and the exposure of biotic communities to disturbances challenges our ability to monitor and manage these systems (Solana-Gutiérrez et al., 2017). For example, impacts from disturbance may attenuate downstream, accumulate where there are multiple disturbed sites, and impact how disturbed sites recover. Therefore, tools for quantifying the amount of human-induced disturbance are important for estimating the exposure of biotic communities to impacts which are crucial for understanding and managing these fluvial ecosystems.

There is an increasing interest in the use of remotely sensed data for monitoring mining areas as most available methods and techniques can achieve high land-use/landcover classification as well as improve accuracy in extracting environmental variables. Freely available datasets such as Landsat and Sentinel, and the advent of archives of analysis-ready data (Dwyer et al., 2018), are enabling unprecedented access to data, hence unlocking the potential to characterize, quantify and monitor human-induced disturbances. It has been estimated that most forms of disturbances such as mineral resource extraction have widespread spatially explicit ecological and environmental impacts due to land transformation, and quantifying these impacts is key to understanding the extent of impacts and developing mitigation measures (Werner et al., 2019).

Remote sensing has been successfully deployed to document changes in fluvial environments for decades (Nellis et al., 1998; Wright et al., 2000; Woldai, 2001). For example, Gilvear et al. (1995) applied image analysis to quantify water depths and instream habitats and showed that

placer mining contributed to reduced water depth and habitat diversity. Ghoshal et al. (2010) also proposed using topographic maps to quantify the impacts of historic placer mining on stream sediment load. Other studies have employed hyperspectral data (Marcus 2002), terrain data (Egidarev and Simonov 2015) and object-based image analysis (Whiteside and Bartolo 2015; Wang et al. 2016) to examine the impacts of mining on aquatic ecosystems. Using high spatial resolution imagery, Asner et al. (2013) quantified gold mining in the Amazon from 1999 – 2012 and showed that mining is expanding rapidly than has been previously reported. Similarly, Espejo et al. (2018) found that gold mining in Western Amazon increased by 21% between 1985 – 2017 compared to what has been reported. In related research, Classification and Regression Trees implementation in Google Earth Engine was employed to detect mining areas in Amazon; various type of minerals mines were discovered with an increased area of activity than previously known (Lobo et al., 2018). Obodai et al. (2019) applied Spectral Angle Mapper (SAM) on Landsat imagery and detected about 50% reduction in closed forest areas from 1991 – 2016 in the Ankobra River basin, Ghana. Mhangara et al. (2020) quantified changes in vegetation cover, bare soil, and mined open pits from 2014 – 2018 in South Africa using SPOT-6 satellite imagery and reported progressive decline in vegetation cover. The authors utilized morphological classification and SAM to map changes in vegetated areas and bare soil. A study conducted by Yu et al. (2018a) demonstrates the potential of remotely sensed data to provide insight into the dynamics of mining activities at global scales. Using multiple sensor datasets, the findings revealed that a substantial number of surface mines in North America decreased in size (i.e., were rehabilitated) from 1980 – 2013; conversely, South America and Asia experienced the highest expansion in mining areas.

Despite the above accomplishments, specific applications of modern image-based classification/mapping methods to placer mining are limited, and considerable work remains to realize disturbance mapping over large areas owing to the complexity of placer signature and its spatial/temporal properties. More importantly, the above approaches offer limited contextual information required to discriminate placer signals; thus, the need for contextual-aware detection methods.

5.2.1 Classification/detection of Placer Mining Disturbance

Placer mining requires several activities which leave a visible trace on the landscape; including but not limited to road building, dirt/earthen mounds, and ponding. For example, discharge of placer mine tailings change sediment concentrations, as well as downstream water turbidity, thus degrading benthic habitat quality, providing a detectable signal from moderate spatial resolution image data (Nelson and Church, 2012; Egidarev and Simonov, 2015). Figure 5.1 provides an air photo depicting a section of a watershed with active placer mining. Roads, mining ponds/water pools and placer mines occur in close proximity, while natural disturbances frequently appear near forested areas. The challenge for mapping at the watershed scale is to develop an image analysis scheme that can effectively recognize and discriminate placer signature. Clearly, the complexity of the placer mining disturbances on the landscape does not lend itself well to discrimination via pixel-based analysis.

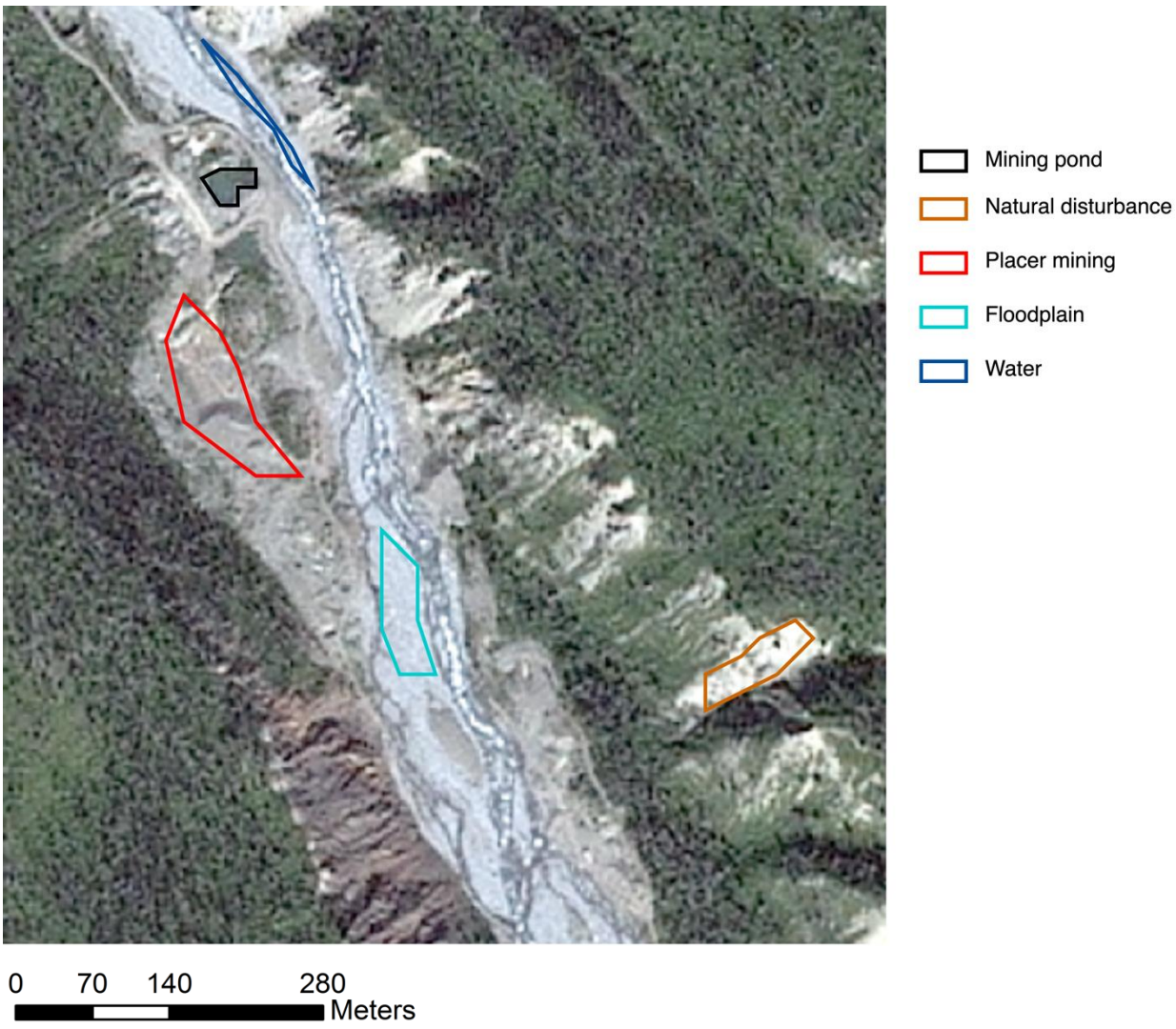


Figure 5.1 A section of Alsek watershed depicting the scale of placer disturbance.

5.2.2 Deep Learning Models in Environmental Mapping

Machine learning models are increasingly being deployed to map mineral mining related disturbances. For example, Ibrahim et al. (2020) identified placer mining areas in Colombia and estimated 35% vegetation loss between 2016 – 2019. The study demonstrates the effectiveness of support vector machine (SVM) and potential of Sentinel-2 data to detect placer disturbance. In related research, Dlamini and Xulu (2019) illustrated the potential of Google Earth Engine based

LandTrendr algorithm to map mine disturbance and recovery trajectory in South Africa. Support vector machine optimized with k-fold cross validation was found to improve landcover classification in open pit mines (Chen et al., 2020). Given the complexity of mining disturbances, other sophisticated modelling approaches have been employed to monitor mining activities. For instance, Abaidoo et al. (2019) applied Artificial Neural Networks to detect landcover changes as result of mining as well as vegetation response to land reclamation practices in abandoned mine areas.

Inspired by successes in classification, pattern recognition, and object detection (Krizhevsky et al., 2012), deep learning models are now frequently deployed to exploit spectral and contextual information available in remote sensing data to improve landcover classification (Zhang et al., 2020; Han et al., 2020). Convolutional neural networks (CNN) are spatially explicit models consisting of layers that process their input hierarchically using convolutional operations and non-linear mapping functions (Krizhevsky et al., 2012). Wang et al. (2020) applied Mask Region-based CNN (Mask R-CNN) to identify open-pit mines and evaluate environmental damages caused by mines; while Maxwell et al. (2020) proposed applying Mask R-CNN for topographic mapping of valley fill faces which is indicative of mountaintop removal due to coal mining. In related research, Madhuanand et al. (2021) detected surface coal mines from satellite images using Mask R-CNN, reporting 95+% overall accuracy on independent validation satellite images from different countries. U-Net is a modification of CNN, employing a U-shaped architecture. U-Net was first implemented in biomedical image segmentation and found to segment cells with high speed and accuracy while learning discriminative patterns on very few training images (Ronneberger et al., 2015). The U-Net model has been shown to be effective in environmental mapping as the model architecture enables the learning of texture and other

spatially discriminative features (Flood et al., 2019). Recently, Gallwey et al. (2020) developed a U-Net model to map and quantify small scale mining disturbances in Southern Ghana using Sentinel-2 data. The authors modified the model's output to yield classifications from the final layer as well as per landcover class prediction probability; the mean probability score was ultimately used to determine final pixel classes, and thus minimizing classification errors.

The studies noted above emphasize the potential of remotely sensed imagery as well as the capability of machine learning techniques to map mining related disturbances. RF on the one hand is an ensemble learning method that is increasingly been applied in land-use/landcover classification owing to its high performance. It has been proven to efficiently model complex relationships due to its effective weighting of contributing variables via majority voting mechanism and requires less training time (Syrris et al., 2019). Convolutional neural networks on the other hand are known for their inherent capacity to model complex processes using contextual information and a high number of parameters. U-Net and RF models are both non-parametric and have been proven to learn sophisticated patterns with low overfitting tendency when exposed to large number of input variables. Thus, these two modelling frameworks hold potential for learning the complex patterns of disturbance associated with placer mining activities.

The goal of this study was to characterize and quantify placer-induced disturbances at watershed scale using different configurations of high spatial resolution satellite imagery and machine learning models. More specifically, we focus on evaluating the potential and limitations of two modeling frameworks, U-Net and RF, in detecting placer mining activity. Through the use and comparison of multiple complimentary modelling approaches, and different combinations of satellite data (e.g., SPOT-6/7 band combination), we expect to get a clear

understanding of the potential for image-based placer disturbance monitoring. Operational mapping and monitoring of placer disturbances remains limited and significant potential exists to develop real-world tools capable of filling this gap in environmental monitoring systems. The contributions of this study are: (a) model parameterization issues in mapping placer disturbances are highlighted; (b) tools that hold potential to detect and characterize placer mining disturbance at watershed scale are proposed. We structure the remaining body of the paper as follows: first, our methodological approach is highlighted (e.g., study site and sampling, datasets and ancillary data derivation, and model design/architectures); next, the results of the modeling frameworks are presented followed by discussion and conclusions.

5.3 Materials and Methods

5.3.1 Study site selection

Our study is focused on a region with significant placer mining activity in Yukon Territory, Canada (Figure 5.2). The relationship between placer mining and its impacts are complex due to the long history of placer mining and vast areas it has taken place. The Alsek watershed was chosen for this study as mining has commenced relatively recently (early 1980s) in this watershed, therefore limiting the impact of noise introduced through remediated and/or semi-remediated historical mining sites. For example, in the Klondike watershed, the area has been continuously mined for over 100 years. We focused model development on recent placer mining activities to simplify the development of the initial models for placer classification. Figure 5.2 depicts the study site as well as the distribution of training/reference placer parcels and aquatic health monitoring test/reference sites. Placer areas are usually demarcated by monitoring

authorities using polygons (i.e., parcels) (e.g., see rectangles in Figure 5.2). Licenses are then assigned to parcels for individuals/companies interested in placer mining. Each parcel therefore has a unique license holder and thus forms the unit for disturbance monitoring relative to the level of compliance.

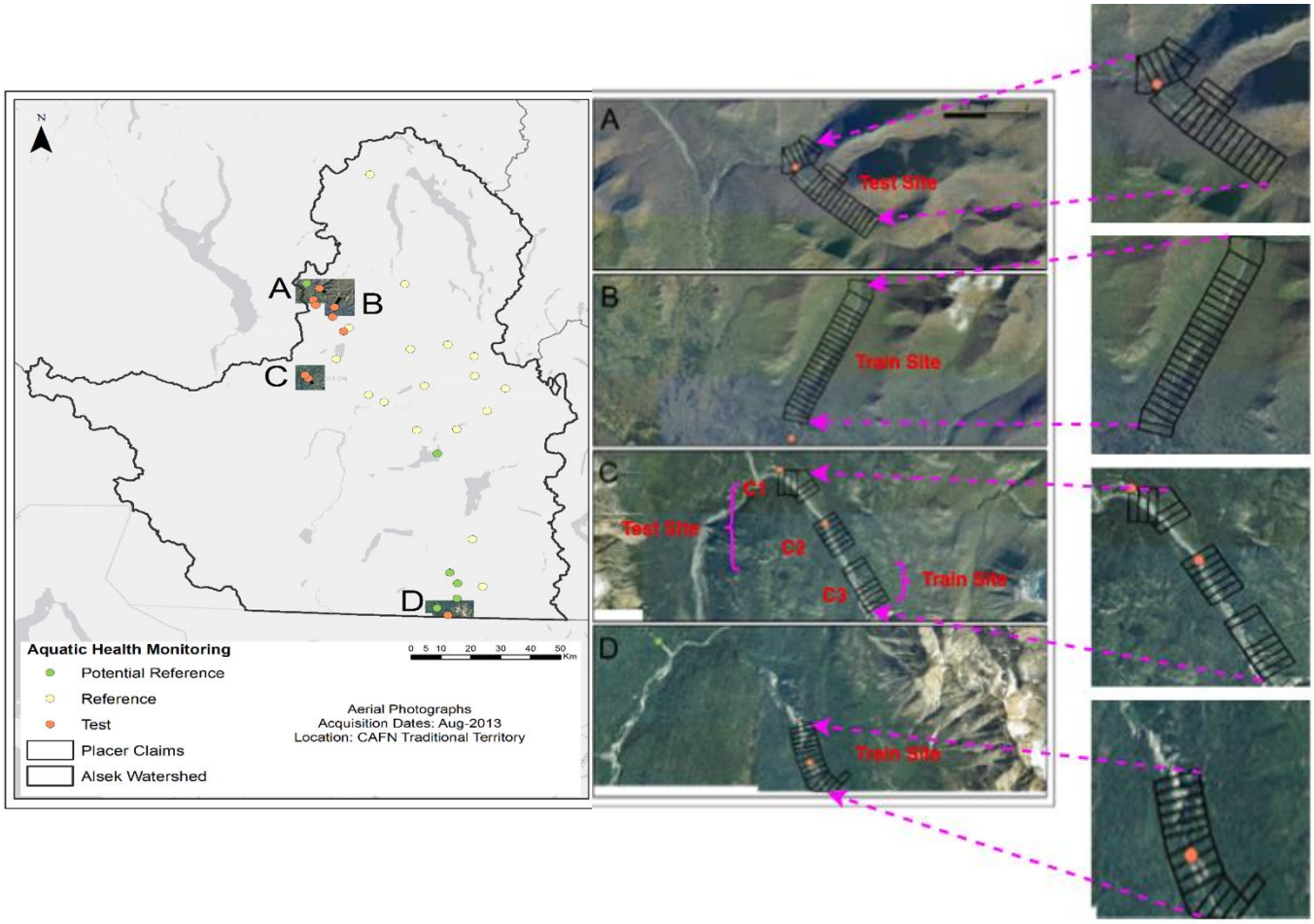


Figure 5.2 Map of distribution of training/reference placer parcels and aquatic health monitoring test/reference sites. The Panel on left shows a small-scale map of the Alsek watershed with letters A, B, C, D and aerial photographs showing locations of training/reference placer parcels. The four panels on

the right show a larger scale of the groupings of training/reference placer parcels (i.e., placer claims) with the letters in the top left corner corresponding with small scale map. Aerial photographs were taken in August 2013.

5.3.2 Placer Parcel Selection Criteria

Within the Alsek watershed, individual parcels of land were selected to create training and reference datasets for model development and validation. Parcels are the areal units of land which are used for managing placer claims. We used objective criteria to select these parcels, outlined in Table 5.1.

Table 5.1 Placer parcel selection criteria.

Parcel criterion	Justification
Major operations	Long-term, large-scale operations which have mandatory compliance monitoring and inspection water sampling
Upstream from monitoring stations	Monitoring stations data can be mapped to mining disturbance which has occurred upstream
Placer stake data < 20 years	Limits model development to parcels with only recent placer disturbance making it easier to detect disturbance events versus natural landscapes

Major mining operations were chosen as a criterion as these operations are generally larger scale and multi-year (i.e., small-scale exploration work and reclamation are typically completed within one year). Major operations require greater regulatory and environmental monitoring due to the increased environmental impacts. As such, these sites are potential test beds to evaluate remote sensing tools capability for disturbance intensity monitoring.

Within the past 40 years, substantial mining activity and subsequent reclamation/regeneration has occurred, thus significantly altering the landscape making it challenging to distinguish between naturally occurring landscapes and areas which have been disturbed but may be in the process of regeneration. Parcels polygon data were overlaid on aereal and satellite imagery to map identify placer disturbed areas. We relied on recently staked claims to visually discriminate between natural and disturbed areas by employing available high resolution multi-temporal aerial and satellite imagery. By using a multi-temporal/time-series approach to visually identify disturbance/regeneration events, it is possible to better understand the spectral response associated with short and long-term disturbance events and track stages of regeneration. The criteria rendered 75 parcels, 41 of which were used for training and 28 of which will be used as a reference dataset to validate the models. The remaining 6 parcels were not used due to challenges in visually classifying landscape features. Of the 75 parcels, 12 were staked between 2010 and 2019, 15 were staked between 2000 and 2009, 21 were staked between 1990 and 1999 and 27 between 1980 and 1989. Figure 5.3 shows the location and distribution of training and validation datasets. Sites A, C2 were selected for independent validation, while sites B, C3 and D were used for model training. Site C1 consists of 6 placer parcels, and these were not utilized due the challenges indicated above. The selection of both training and independent validation datasets is

based on the criteria outlined in table 5.1. Additionally, we found that placer disturbances can be visually identified at the selected sites with minimum errors using air photos.

5.3.3 Datasets

We explored both free (i.e., Landsat, and Sentinel), and commercial (i.e., World-View-2, IKONOS, SPOT-6/7 etc.) image sources to identify the data required to detect placer disturbance. After visually examining layers of free image sources and high-resolution base maps available at the GeoYukon website, we found that the low resolution of free imagery did not provide the image quality and spatial resolution required for accurate image classification. SPOT- 6/7 satellite imagery (60km x 60km scenes) and a high-resolution digital elevation model were the primary data sources used in the image analysis.

5.3.3.1 Satellite imagery pre-processing

High spatial resolution imagery is capable of providing detailed feature information which may capture potential signatures of placer disturbances. The SPOT-6/7 satellites capture four-band multispectral imagery (i.e., Red, Green, Blue and Near Infrared (NIR)) with a multispectral band spatial resolution of 6 m and panchromatic mode with a spatial resolution of 1.5 m. The four bands were pansharpened using the panchromatic band; this renders the satellite data suitable for finer detail habitat characterization (Whiteside and Bartolo, 2015). Available SPOT-6/7 products have been geometrically, radiometrically and atmospherically corrected, enabling indices used for disturbance detection, such as Normalized Difference Vegetation Index (NDVI) to be easily derived.

5.3.3.2 Ancillary data derivation

In addition to multi-spectral bands from SPOT-6/7 sensors, ancillary datasets including NDVI, and Soil-Adjusted-Vegetation Index (SAVI) were used. NDVI is readily used for assessing plant health and abundance and has been used to enhance detection of land cover changes driven by mining activities (Almeida-filho and Shimabukuro, 2000). The index is derived by using the Red and NIR bands from multi-spectral satellite imagery, as the spectral information captured by the Red band is related to chlorophyll absorption and the NIR band is related to plant mesophyll cellular reflectance (Hurcom and Harrison, 1998). Additionally, the normalization procedure is used to reduce the effect of atmospheric attenuation and seasonal sun angle differences (Rouse et al., 1974). NDVI is calculated by finding the ratio between the difference of the NIR band and Red band and the sum of the NIR band and Red band, rendering a scale of -1 to +1, with high values indicating healthy vegetation and low values being linked to disturbed vegetation or absence of vegetation (Rouse et al., 1974). Given the invasive nature of placer mining, it is anticipated that low values of NDVI will be an indicator of land degradation associated with areas of placer mining disturbance as a result of complete or partial removal of vegetation in the trenching process (Chen and Rao, 2008). SAVI, which is an extension of NDVI and corrects for the influence of soil brightness when vegetative cover is low, was also used. Equations 5.1 and 5.2 denote NDVI and SAVI, respectively. *L* denotes soil brightness correction factor which is set to 0.5 to accommodate most landcover types.

$$NDVI = \frac{NIR-R}{NIR+R} \qquad \textit{Equation 5.1}$$

$$SAVI = \frac{(1+L)(NIR-R)}{NIR+R+L} \qquad \textit{Equation 5.2}$$

5.3.3.3 Identifying placer patterns/signals

Placer features in imagery may vary with mining characteristics such as time-since-mining, intensity of mining, and type of extraction equipment deployed. Placer effects may manifest on watercourses as deposits of washed rocks or sand often seen as a light band along the rivers. The existence of small dams or pools, sometimes visible as black spots, are indicative placer signatures. Bare soil or barren lands adjacent to river reaches are prevalent features of placer effects originating from the removal of fertile soil layer along with vegetation. After visually exploring placer disturbed areas using air photos, we identified five land cover classes – placer, natural disturbance, floodplain, water, and forest. These classes may be distinguished from each other given their spectral, textural, and morphological differences. We defined the placer class as areas disturbed by human activities but are near the river reach, and sometimes adjacent to floodplains. Such areas are mostly characterized by bare soil, earthen mounds, and corrugated land surfaces in the river reach. Dark areas in the main river channels were identified as water. Also, pools of water that appear in placer mining areas were classified as water. The floodplain class is typically sand-rocks mixture deposited by water currents and are located adjacent to water in the river channels. The natural disturbance class is characterized by non-human disturbed areas, and may appear as bare soil, eroded land, or eroded mountain surfaces. Also, any bare soil located within forest is identified as natural disturbance since placer mining is restricted to mineral rich areas near the river reach. The forest category encompasses land surfaces covered with green or dry vegetation.

Placer disturbance polygons were digitized from high resolution air photos. We note that the placer polygon data demarcate the boundaries of each landowner, and therefore were only used

to map placer activity to aerial photos and SPOT-6/7 imagery. As mentioned earlier, the 41 parcels that met our selection criteria were used for model training (i.e., sites B, C3 and D). Twenty-eight placer parcels were digitized and not used for model training (i.e., sites A and C2), being held back for independent model validation. Given that training patches or pixels are not statistically independent from each other as they are split via bootstrapping into in-bag and out-of-bag (i.e., for RF model) (Cánovas-García et al., 2017), and training and validation (i.e., for U-Net), independent validation, we believed, was crucial for testing the models' generalization to watersheds in different locations. The validation sites were at Site A (2014 and 2018) (images were acquired in July 19, and September 16) and Site C2 (2013 and 2019) (images were acquired on July 12, and August 10). Other classes that were digitized were forest, floodplain, water, and natural/other disturbance. Figure 5.3 illustrates training and independent validation sample sites.

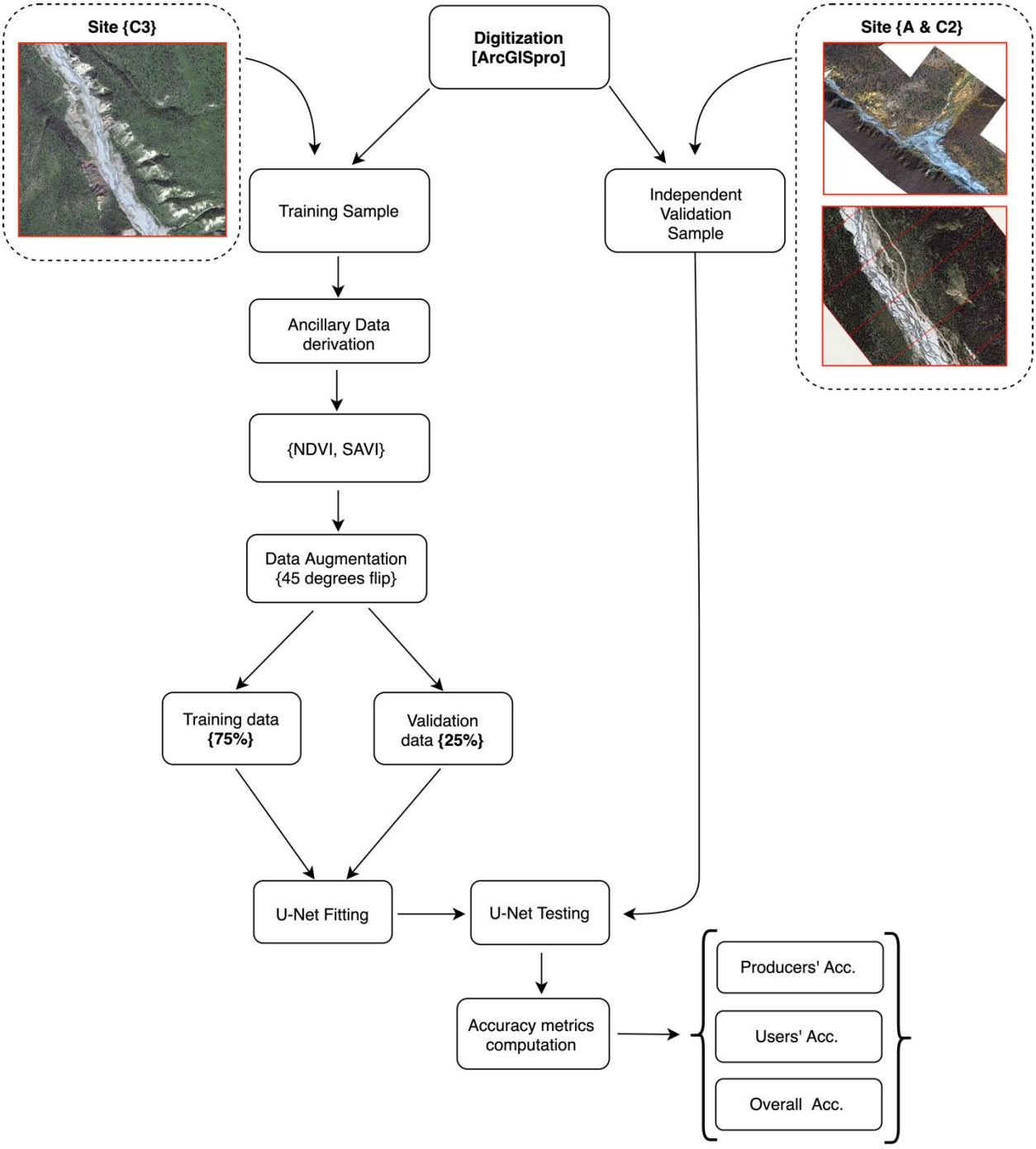


Figure 5.3 A flowchart for training/validation sample selection and model testing.

5.3.4 U-Net Model architecture and modeling framework

The proposed modelling framework employed classical approaches and deep learning models to detect and characterize placer signals. The U-Net is a deep feature extraction model that incorporates texture and spatial information to implement semantic learning and subsequent segmentation or classification of image pixels. The model architecture encompasses a feature encoding path and a decoding path. The encoding path extracts hierarchical features/patterns given labelled input data, while the feature decoding path learns spatial information required in order to reconstruct the original input (Ronneberger et al., 2015). The model has demonstrated outstanding performance in variety of semantic segmentation tasks (Gallwey et al., 2020). Image segmentation and/or classification problems are analogous to detecting or predicting placer signatures, though this was the first application of this modeling framework to placer disturbance monitoring that we are aware of.

5.3.4.1 Data augmentation

Training deep-learning models from scratch has overfitting consequences. To mitigate this limitation, data augmentation was applied to substantially increase the training sample size. ArcGIS-pro provides export training data for deep learning tool. Using this tool, we set the rotation parameter to 45 degrees. This flips each image-mask pair 45 degrees, thus resulting in in two-fold increase in the training sample size during image chip generation as each sample is doubled. In total, 2,058 of 256 × 256 patches were used for model training and testing. Figure 5.4 depicts samples of images and the corresponding augmented copies.

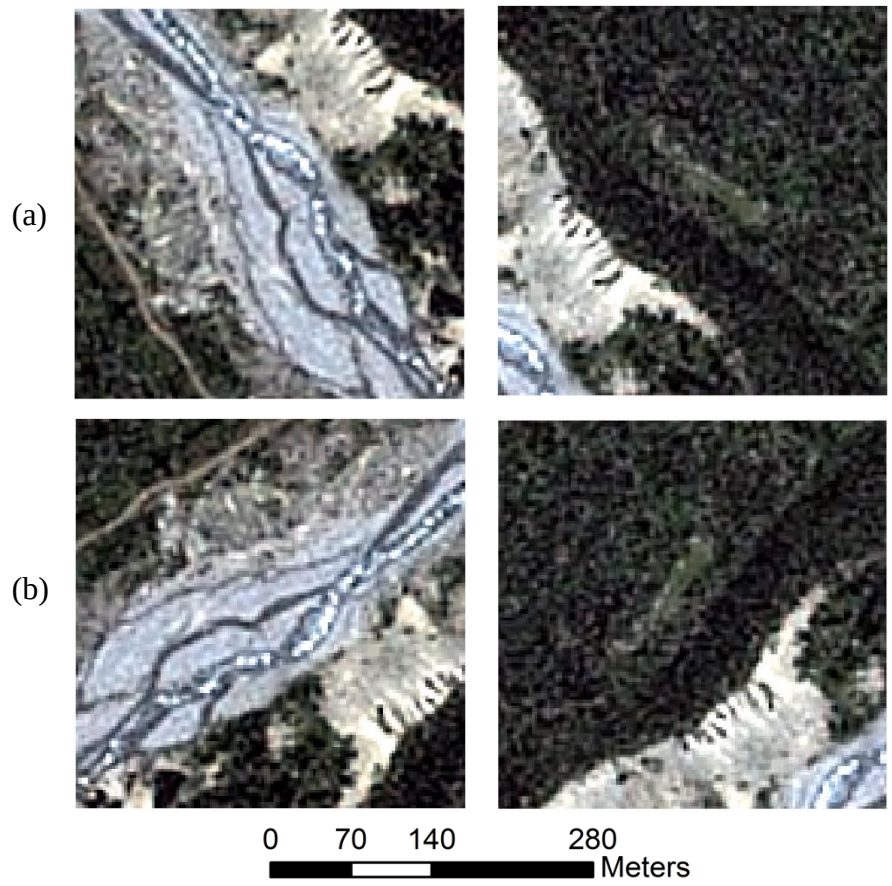


Figure 5.4 Samples of augmented SPOT-6/7 data from Site C3. Row a shows original data, while row b presents the corresponding augmented images at 45 degrees.

5.3.4.2 Model architecture

For operational purposes, we opted to develop a U-Net model from scratch with minimal parameters as possible. A computationally efficient and relatively small model is crucial for field operation. Moreover, given the limited size of our dataset, a large model is likely to result in over fitting. Three input data combinations were employed to evaluate the models' performance on placer detection: the original RGBNIR bands and the derived data (e.g., NDVI and SAVI). This

resulted in three data-model configurations – RGBNIR (i.e., U-Net RGBN model), RGBNIR+NDVI (i.e., U-Net NDVI model) and RGBNIR+SAVI (i.e., U-Net SAVI model). Given the band combinations, the RGBN model input remain four image channels, while the NDVI and SAVI models take five image channels as their input. We adopted the U-Net architecture proposed by Ronneberger et al. (2015), but reduced the number filters in the convolutional layers to develop a smaller model. To learn placer signals, a more flexible non-linear function was used to handle NDVI and SAVI bands where pixels range from small positive to negative values. We did this by applying elastic ReLU (EReLU) which randomly compresses and enlarges network activation within a moderate range at training time. EReLU has been proven to improve model training while minimizing overfitting (Jiang et al., 2018). To improve accuracy and reduce training time, Batch Normalization (BN) was applied to the down-sampling path of the model. Batch Normalization is a linear transformation of model input or activation to have mean of zero and unit variance; this reduces internal covariate shift and speeds up model convergence (Ioffe and Szegedy, 2015). Figure 5.5 shows the U-Net architecture adopted, while Table 5.2 summarizes the model architecture and parameters. The models were trained for 50 epochs but with early stopping parameter (i.e., patience set to 10) to stop training if there was no improvement after 10 epochs to avoid overfitting. Adam was used as the optimizer (Kingma and Ba, 2015). The learning rate was set to 0.001; other parameters were maintained at their default values. Keras-TensorFlow backend was used to support computations on computer (Intel(R) Xeon (R) CPU E5-2650 v2 @ 2.60GHz, RAM 64 GB) and NVIDIA enabled GPU (Quadro K4000) with 4G memory.

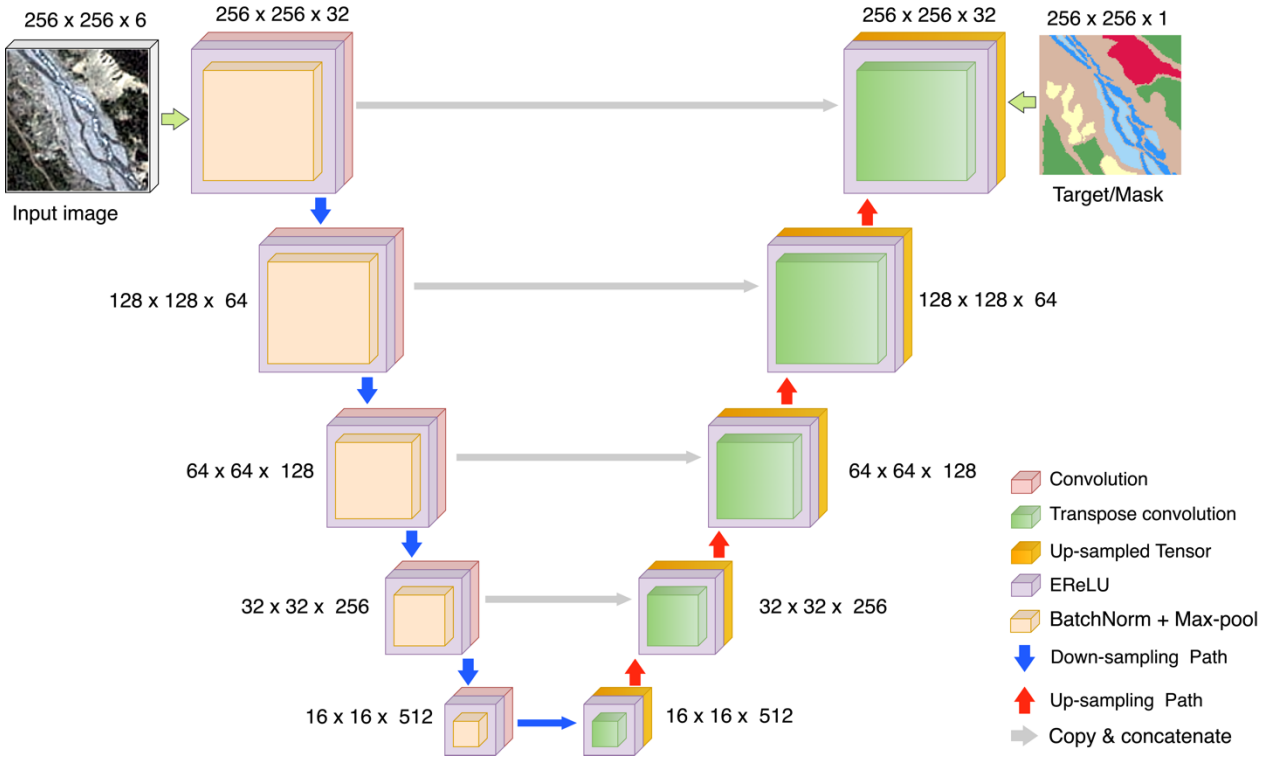


Figure 5.5. Architecture of placer disturbance detection U-Net. Down-sampling and up-sampling paths both lead to learning placer disturbances. Six classes are depicted in the mask, where the sixth class represents areas that were not digitized. Note that red pixels in the target/mask denote placer mining area.

Table 5.2 A summary of model architecture and parameters.

Layer name	Feature maps	No. of Filters	Max-Pooling	Activation	BN
Conv-1	256 × 256	32	2 × 2	EReLU	Yes
Conv-2	128 × 128	64	2 × 2	EReLU	Yes
Conv-3	64 × 64	128	2 × 2	EReLU	Yes
Conv-4	32 × 32	256	2 × 2	EReLU	Yes
Conv-5	16 × 16	512	2 × 2	EReLU	Yes
deConv-1	16 × 16	512	No	EReLU	No
deConv-2	32 × 32	256	No	EReLU	No
deConv-3	64 × 64	128	No	EReLU	No
deConv-4	128 × 128	64	No	EReLU	No
deConv-5	256 × 256	32	No	EReLU	No
deConv-6	256 × 256	6	No	Sigmoid	No

Although the U-Net architecture inherently incorporates fine-grain learning as well as spatial and contextual information, detecting placer disturbances interspersed with other landcover types having closely related signals could confound the task at hand. In this context, the models' filter size is one of the crucial parameters to consider, especially, for learning spatial dependency between different classes. Figure 5.6 illustrates a conceptual framework of the filter parameterization process to detect placer signals. We set the filter size to 3 × 3 (Figure 5.6b) after a close examination of high-resolution aerial imagery showed that most placer mining disturbances exceeded 5m × 5m. A pansharpened SPOT-6/7 imagery has spatial resolution of 1.5m × 1.5m pixel (Figure 5.6a). Note that the receptive field of the first convolutional layer filters is also equivalent to 3 × 3 (Figure 5.6c), and this translates to 4.5m × 4.5m in SPOT-6/7 data (Figure 5.6d). The filter size adopted, and receptive field size were sufficient to extract placer signals/signatures from the input data. Figure 5.6e shows original SPOT6/7 image and a sub-image depicting an area with intense placer disturbance. We assessed wider filter sizes and receptive fields (e.g., 5 × 5 and 7 × 7) but they did not improve the models' discriminative

performance. Also, large filters are likely to reduce detection accuracies in water areas and for natural disturbances which tended to be smaller than 3 × 3 pixels or 5m × 5m.

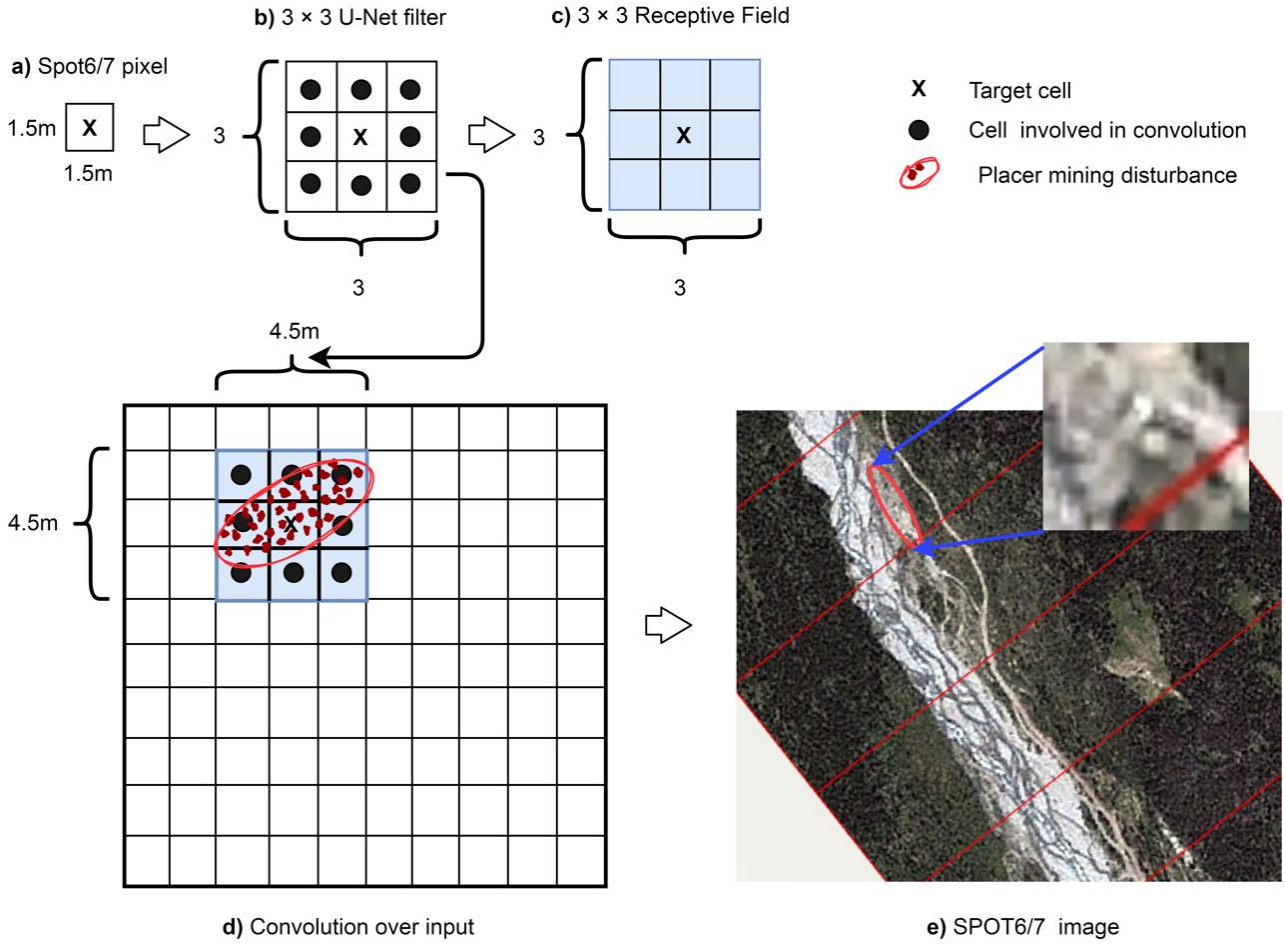


Figure 5.6 Schematics of U-Net filter selection to detect placer signatures. (a – c) show SPOT-6/7 pixel resolution, U-Net filter (kernel), and the first convolutional layer's filter receptive field size, respectively. (d) depicts convolutional operation over some input data using a 3 × 3 filter, an equivalence of 4.5m × 4.5m in SPOT-6/7 image. (e) denotes an original SPOT-6/7 image with a sub-image showing intense placer disturbance area.

5.3.4.3 Class weighting

During assessment of the augmented dataset, it became apparent that certain landcover classes (e.g., placer and natural disturbance areas) were underrepresented classes. Training models with such an imbalanced dataset could lead to lower accuracy on rare classes (Buschbacher et al., 2020). One possible approach to circumvent class imbalance problem is under-sampling technique which involves discarding images that contain the dominant classes (e.g., forest). Since landcover classes occur together, this solution could, diminish an already scarce dataset and reduce learning of discriminative patterns to distinguish classes. To surmount this limitation, oversampling which randomly increases the size of underrepresented minority class is recommended for improving deep learning models performance (Buda et al., 2018). Cost sensitive learning techniques that modify model architecture by introducing specific loss functions are also widely used (Khan et al., 2018). Dice loss is a potentially effective loss function for tackling class distribution problems. Dice score coefficient measures the extent of overlap between classes during segmentation, and is proven to be effective, especially when there is gold standard or ground truth. A variant of the Dice loss function which is modified to provide invariance to different class properties (Crum et al., 2006), has been shown to be effective for training deep learning models (Sudre et al., 2017). Class weighting is yet another effective approach proven to improve U-Net performance for an imbalanced class distribution (Ronneberger et al., 2015). We adopted class-sensitive learning by weighting class instances differently during model training. A custom weighted categorical cross-entropy (WCE) loss function was implemented to achieve this. With fewer classes, weights can be set and adjusted iteratively to obtain the best per-class accuracy. Forest and floodplain weights were set to 0.8 and 0.9 respectively, after a series of experiments; the remaining classes weights were maintained at

1.0. Since it was not feasible to digitize the entire training area, undigitized pixels were converted to the sixth landcover class. The weighting technique adopted allowed us to efficiently exclude these undigitized pixels at training time by setting the weight of the sixth class to a very negligible number. As shown by Buschbacher et al. (2020), WCE loss function can be formulated as follows:

$$L(\hat{y}, y) = -\sum_i w_i \hat{y}_i \, log(y_i) \qquad \textit{Equation 5.3}$$

Where w_i denotes the weight of class i, while $\hat{y}$ and y are the network's output (i.e., logits) and the corresponding true class label, respectively.

5.3.5 Random Forest model

Random forest classification offers a solution to the shortcomings of classification trees by employing bagging to construct an ensemble of independent, unpruned regression trees using a boot-strap sample of the dataset from which variable importance is decided based on a voting system (Breiman, 2001). Each tree of RF is based on a random subset of the dependent variable (on average 36.8% "out of the bag" not used for any individual tree), and each split within each tree is decided based on a random subset of independent variables (Breiman, 2001). Random forest has been found to be highly accurate in land cover classification, but RF are a-spatial and do not natively take advantage of spatial (i.e., contextual) information. In this modelling, 300 trees were utilized to create the RF model. To account for the class imbalance problem, 90% of the pixels in the forest class were dropped at training. Like the U-Net models, three RF models

were employed using similar image data combination – RGBNIR (i.e., RF RGBN model), RGBNIR+NDVI (i.e., RF NDVI model), and RGBNIR+SAVI (i.e., RF SAVI model).

5.3.6 Accuracy evaluation metrics

Accuracy of model predictions were determined by calculating confusion matrices over mapped classes as this approach is an intuitive and acceptable standard for evaluating remote sensing image classification accuracy (Conglaton, 1991). Producers' Accuracy (PA), Users' Accuracy and Overall Accuracy (OA) are intuitive and widely utilized accuracy assessment metrics. Placer detection accuracy was evaluated in terms of PA (aka recall) and UA (aka precision) for placer detection. Per class accuracy measures are suitable for dataset with no background class (Csurka et al., 2013). Omission errors (False Negative (FN)) are associated with PA (i.e., *FN* = 1-PA), while commission errors (False Positive (FP)) complements UA (i.e., *FP* = 1- UA). PA, given by $PA = \frac{TP}{TP+FN}$ is derived from True Positive (TP) and FN which are in the same column of the confusion matrix, hence, PA is unaffected by the data distribution. Contrarily, UA denoted by $UA = \frac{TP}{TP+FP}$, is influenced by imbalanced data (Tharwat, 2018). Similarly, OA given by $OA = \frac{TP+TN}{TP+FP+TN+FN}$, which is a commonly used metric for evaluating classification performance, is sensitive to data distribution, and hence is not suitable for imbalanced datasets (He and Garcia, 2009; Tharwat, 2018). PA can be thought of as how often the known features on the land are correctly classified on the map – how well are the features on the land represented by the map. For example, how often is a pixel that is truly placer classified as placer on the map. Alternatively, UA is how often the classifications on the map correctly represent the features on the land – how well does the map predict features on the land. For

example, how often is a pixel predicted by the model as placer is actually placer. While both accuracy metrics are useful information, for the purposes of detecting and quantifying placer activity more weight should be given to the UA results when evaluating model performance, since this would indicate the accuracy of using the output classification map. Additionally, since the aim is to quantify only placer disturbance in an applied earth-observation monitoring context, the effects of imbalanced data on placer detection is likely to be considerably low in areas where placer mining dominates compared to instances where all class types are of interest to the user. The above accuracy statistics should however be interpreted in conjunction with the total class area to provide further insight into class abundance. Other assessment metrics such as receiver operating characteristic curves, geometric mean, precision-recall and F measure are suitable for two-class imbalance problems (He and Garcia, 2009).

5.4 Results of modeling

5.4.1 Model performance assessment

As can be seen in Figure 7, the lowest loss was measured after 45 epochs. Also, it can be noticed that the training Dice loss remained consistently high and appears to not overlap with the validation dice loss. Contrarily, training and validation WCE losses overlap with training WCE loss being relatively low. Also, it can be observed that the model performance did not improve after 45 epochs.

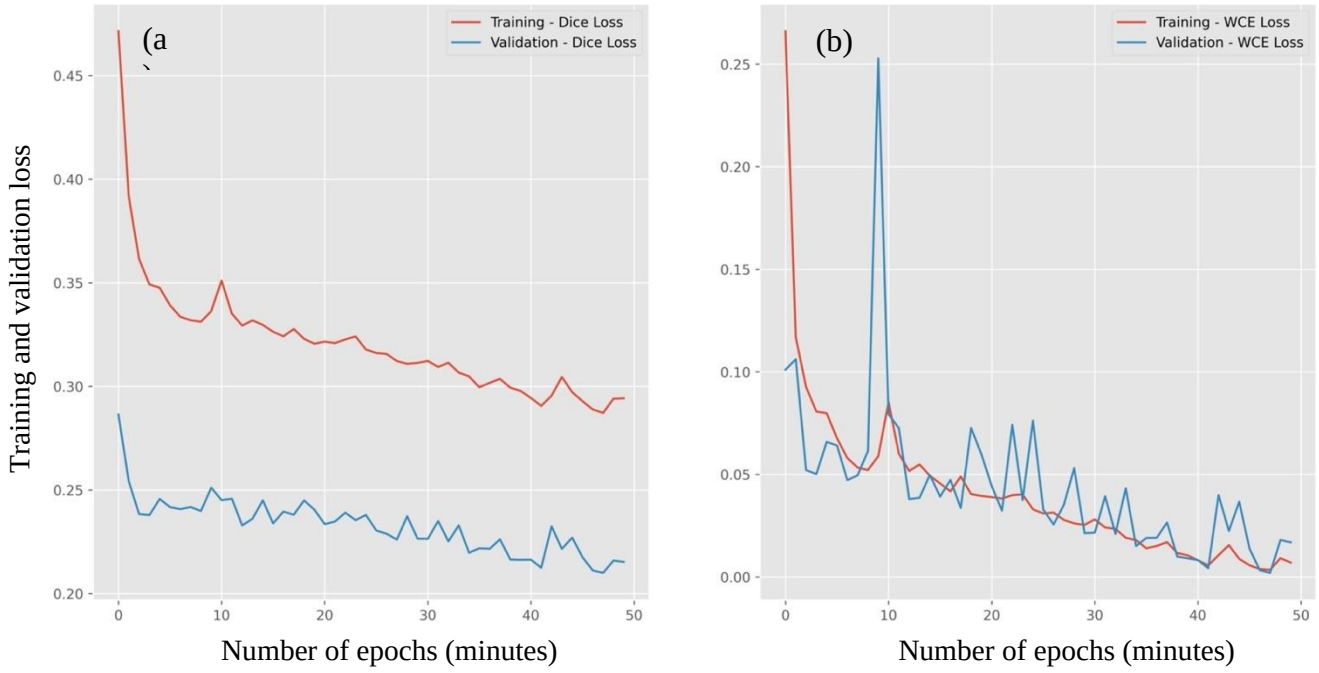


Figure 5.7 U-Net NDVI model training and validation losses. (a) training and validation Dice loss and (b) training and validation WCE loss. Dice losses assess model classification performance by computing the area of overlap between target pixels and predicted pixels, and the total area (i.e., intersection over union), while WCE applies custom weights to class instances.

5.4.2 Producers' accuracy

Overall, the models classified digitized placer sites well, with median producer accuracy (i.e., % of independent digitized placer samples) of 78.94% across all models (Table 5.3). Median for U-Net models was 77.31%, while median producer accuracy for RF models was 84.42%. The models' performance was relatively high in site C and site A-2018. Site A-2014 presented challenges to U-Net and RF as the models' accuracy tended to be lower (e.g., lowest accuracy for U-Net and RF were 45.52% and 71.73%, respectively).

Table 5.3 Classification accuracy assessment of placer disturbance mapping reported as Producer's accuracy.

Site-Year	Model	Placer Area (Ha)	U-Net Producer's Accuracy	RF Producer's Accuracy	Best Model Approach
C-2013	RGBN	0.71	64.79	91.89	RF
	NDVI	0.71	84.32	90.55	RF
	SAVI	0.71	75.21	90.36	RF
C-2019	RGBN	0.52	79.40	78.47	U-Net
	NDVI	0.52	82.00	77.06	U-Net
	SAVI	0.52	89.22	76.54	U-Net
A-2014	RGBN	0.77	45.52	71.73	RF
	NDVI	0.77	63.69	72.61	RF
	SAVI	0.77	47.43	73.20	RF
A-2018	RGBN	7.88	73.33	95.17	RF
	NDVI	7.88	92.36	95.78	RF
	SAVI	7.88	95.13	95.75	RF

5.4.3 Users Accuracy

The median user accuracy (i.e., % of sites classified as placer that were actually placer) was 65.47%. Overall, models were able to classify digitized samples well, but had also misclassified non-placer sites as placer (Table 5.4). On average, the U-Net model's User's accuracy tended to be higher than that of the RF.

Table 5.4 Classification accuracy assessment of placer disturbance mapping reported as User's accuracy. The best model approach is the model with highest accuracy.

Site-Year	Model	Placer (Ha)	U-Net	RF	Best Model Approach
C-2013	RGBN	0.71	84.98	61.94	U-Net
	NDVI	0.71	88.58	65.52	U-Net
	SAVI	0.71	86.98	65.43	U-Net
C-2019	RGBN	0.52	66.94	60.52	U-Net
	NDVI	0.52	61.20	58.62	U-Net
	SAVI	0.52	65.58	58.23	U-Net
A-2014	RGBN	0.77	5.72	27.98	RF
	NDVI	0.77	39.41	35.69	U-Net
	SAVI	0.77	61.74	35.67	U-Net
A-2018	RGBN	7.88	97.67	74.85	U-Net
	NDVI	7.88	80.10	83.25	RF
	SAVI	7.88	92.55	82.59	U-Net

Accuracies were highest for site C, especially the U-Net models (Figure 5.8). It is important to note that the sites they were obtained from were the same site used to train the model. As such, the site characteristics here for the training samples are more similar to the validation samples, however they were partitioned into training and validation parcels (Figure 5.3), so these accuracy statistics for site C still represent independent validation.

Figure 5.7 Comparison of U-Net and RF Producer's and User's accuracy for 2013 and 2019 sub-image at site C. Notice that lower illumination conditions in 2019 reduced classification accuracy.

5.4.4 Class Error Matrices

Tables 5.5 – 5.8 present the actual number of pixels in the validation data and class specific classification accuracies. The column sums are the total number of pixels in each class, while the row sums are the total number of pixels classified as the corresponding class. We note that since non-overlapping tiles were generated using the ArcGIS export training data for deep learning tool, column sum for pixels (i.e., labelled landcover classes) in the U-Net confusion matrix tables are less than RF due to dropping of tiles that were less than 256 × 256 at the edges during tiling of samples. For example, in tables 5.5 and 5.6, the number of labelled placer pixels for RF models exceed U-Net by 477 pixels. This corresponds to 14% reduction in actual placer pixels for U-Net at site A-2014. Similarly, in tables 5.7 and 5.8, placer pixels presented to RF exceed that of U-Net by 445 pixels. Again, this leads to 1.3% reduction in actual placer pixels for U-Net at site A-2018. Also, note that column sum for forest pixels in RF matrices are less than U-Net columns sums as 90% of forest pixels were dropped at training, testing and validation to reduce the class imbalance problem.

5.4.4.1 Site A class accuracy

Placer detection error rates vary remarkably between the models. Both models detect placer at reasonable accuracy but over 1500 actual pixels were detected to be Natural disturbance. There were a total of 2922 pixels in the placer class as determined by digitizing, and 1861 were correctly classified as placer by U-Net. Notably, U-Net error for placer misclassification as Water and Flood plain was zero pixels, while that of the RF exceeded 94 pixels (see Tables 5.5

and 5.6 second column). Interestingly, only 12 placer pixels were misclassified as forest by the RF, yet over 562 placer pixels were incorrectly detected to be forest by the U-Net.

Table 5.5 Class specific accuracy for A-2014 site using U-Net – NDVI model.

	Water	Placer	Natural disturb.	Forest	Floodplain	Total
Water	28	0	0	3	0	31
Placer	507	1861	224	2033	97	4722
Natural disturb.	182	499	68	24	321	1094
Forest	12	562	371	619765	14	620724
Floodplain	0	0	0	0	18	18
Total	729	2922	663	621825	450	626589

Table 5.6 Class specific accuracy for A-2014 using RF – NDVI model.

	Water	Placer	Natural disturb.	Forest	Floodplain	Total
Water	13	207	589	228	13	1050
Placer	819	2468	436	2853	339	6915
Natural disturb.	10	12	1	0	53	76
Forest	185	618	162	37958	0	38923
Floodplain	0	94	0	0	235	329
Total	1027	3399	1188	41039	640	47293

There was a total of about 35000 pixels in the placer class as determined during digitization. Placer detection was very high for both models (i.e., 31913 and 33532 pixels; for U-Net and RF, respectively). Distinguishing natural disturbance class from placer however tended to be challenging for the models. Over 1618 actual placer pixels were detected as natural disturbance by U-Net. Also, 596 and 25 placer pixels were incorrectly classified by RF and U-Net, respectively, to be Water (see column two of Tables 5.7 and 5.8).

Table 5.7 Class specific accuracy for A-2018 using U-Net – NDVI model.

	Water	Placer	Natural disturb.	Forest	Floodplain	Total
Water	0	25	0	23	0	48
Placer	323	31913	0	7453	154	39843
Natural disturb.	0	1618	0	16	0	1634
Forest	32	996	180	559416	0	560624
Floodplain	0	0	0	0	0	0
Total	355	34552	180	566908	154	602149

Table 5.8 Class specific accuracy for A-2018 site using RF – NDVI model.

	Water	Placer	Natural disturb.	Forest	Floodplain	Total
Water	37	598	21	0	0	656
Placer	465	33531	11	5981	289	40277
Natural disturb.	0	4	0	12	0	16
Forest	101	876	594	35730	8	37309
Floodplain	0	0	0	0	0	0
Total	603	35009	626	41723	297	78258

5.4.5 Visual assessment of placer detection performance

In Figure 5.9 row (a), it can be seen that A-2014 is the most difficult area to distinguish placer from other landcover types. The U-Net appears to detect only placer and forest, but unable to distinguish other feature types (e.g., Water/Flood-plain). The RF detected most of the different classes but tended to misclassify few forested areas as placers. This outcome aligns well with

per-class accuracy report presented in Table 5.7. In Figure 5.9 row (b), we noticed substantial placer detection. Again, this visual observation aligns with per-class accuracy report shown in the error matrix (see Table 5.8). The U-Net detected placers well, but with some error rates as certain placer pixels were identified to be natural disturbance (i.e., False negative). Similarly, the RF detected most of the digitized placer areas with acceptable accuracy levels. However, a large proportion of RF map errors can be observed in forested areas where reasonable amounts of actual forest pixels were identified as placer (i.e., False positive).

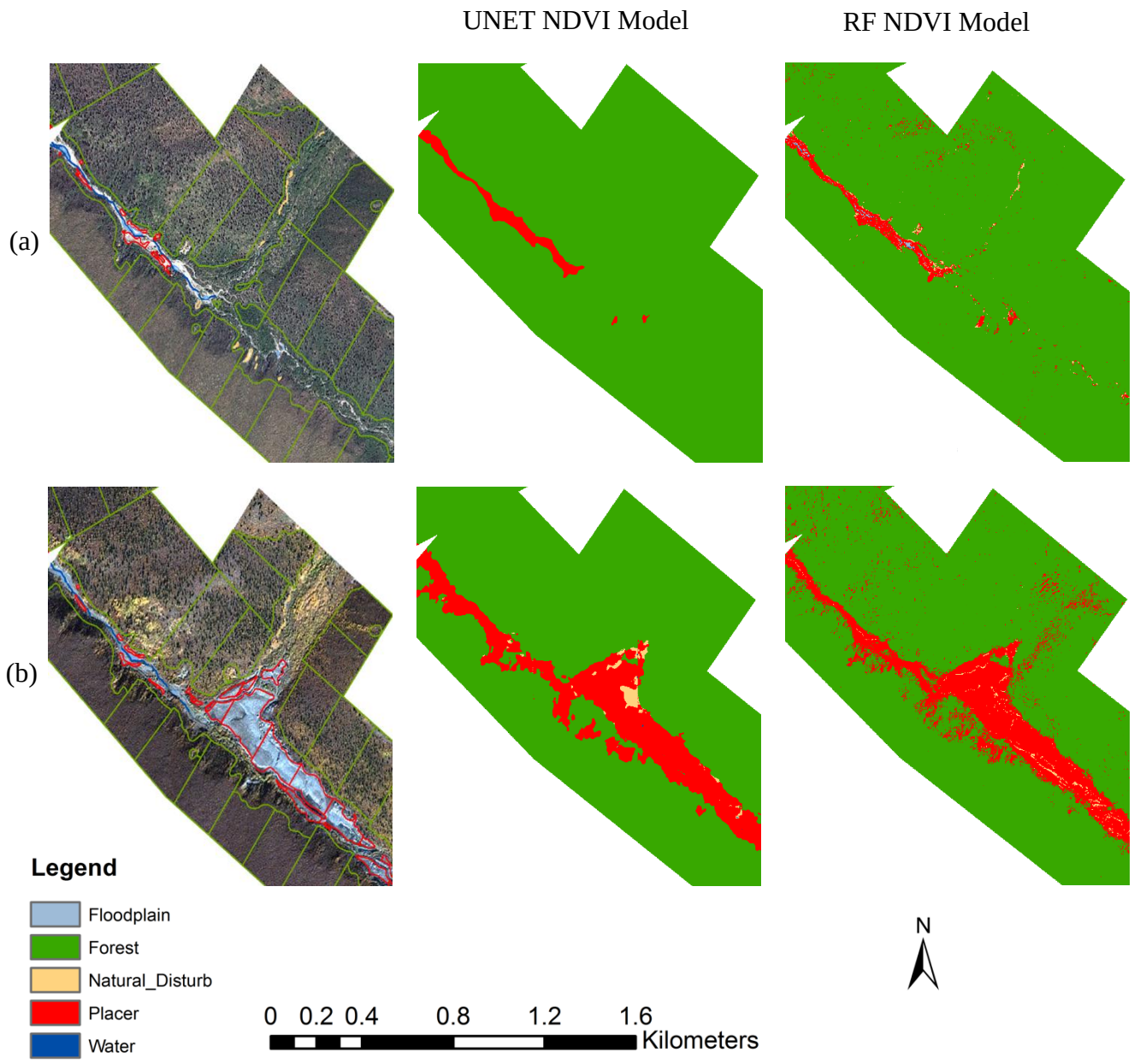


Figure 5.8 U-Net NDVI and RF NDVI models' placer detection maps using in 2014 and 2018 images. Rows (a) and (b) denote site A-2014 and site A-2018, respectively. Columns one shows original images with digitized independent validation samples. Columns two and three present classification maps using U-Net and RF models.

Placer detection maps are shown in Figure 5.10 for validation site C2-2013 and C2-2019. Row (a) shows C2-2013 image, and row (b) denote C2-2019 images and their corresponding placer detection maps using U-Net and RF. On average, the models performed satisfactorily in detecting placer areas that were digitized for validation. While U-Net may be under detecting placer, the RF appears to over detect placer as the maps show substantial placer pixels in forested areas away from the river reach. Natural disturbances occurring away from the reach and within forested areas tended to be well discriminated by U-Net. Contrarily, most naturally disturbed areas were detected by RF to be placer pixels. It should be noted that the U-Net also detected certain pixels in the river reach to be natural disturbance, but these were actually either water or flood-plain, and in some instances were placer pixels (see Figure 5.10 column two). As mentioned earlier, the two images (C2-2013 and C2-2019) exhibit marked differences in illumination characteristics due to variation in image acquisition dates.

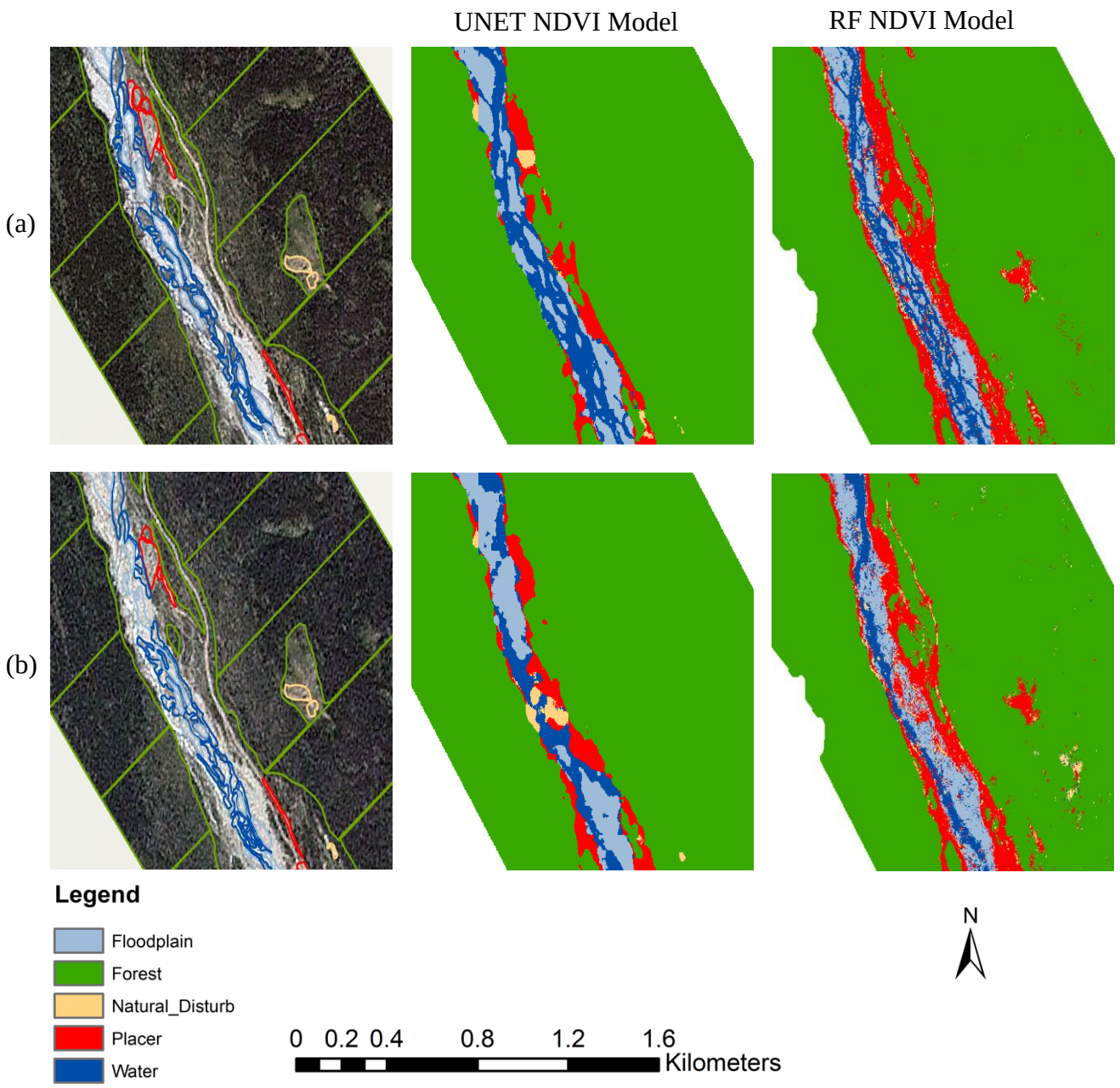


Figure 5.9 U-Net NDVI and RF NDVI models' placer detection maps using in 2013 and 2019 images. Row (a) site C2-2013, and row (b) site C2-2019. Column one shows original images with digitized validation samples. Columns two and three are classification maps.

5.5 Discussion

This study utilized SPOT-6/7 imagery to demonstrate the potential of machine learning methods in monitoring placer disturbance at watershed scales. A deep learning model – U-Net and RF models were fitted and used to detect placer disturbance on an independent validation dataset. U-Net models performance measured in terms of Dice losses and WCE losses varied but remained relatively low after few epochs (Figure 5.7), with the lowest loss determined after 45 epochs. While Dice loss suggests no overfitting (Figure 5.7a), WCE loss provides evidence of some potential overfitting. Maxwell et al. (2020) reported optimal model performance following few epochs while highlighting overfitting issues. Both modeling frameworks successfully detected placer with reasonable accuracy and showed promising potentials for operational mapping of placer disturbances. A U-Net architecture with 3×3 convolutional filters performed best at discriminating placer signals from disturbances.

Random forest models tended to outperform on Producer Accuracy (Table 5.3) whereas the U-Net models tended to outperform on User Accuracy (i.e., reducing misclassifications) (see Table 5.4). As depicted in Figure 5.9, there are differences in placer detection performance between models and data configurations. While there were occurrences of false positives in placer detection for RF models, similarly, there exists false negative rates in placer areas detected by U-Net models. Most misclassification was due to water, in the 2014 image for site A (see Tables 5.5 – 5.8). However, as can be seen in Figure 5.9 row (a), there was large misclassification of forest sites as placer, occurring for both models. We note that although the number of placer pixels presented to RF and U-Net models differed between sites A-2014 and A-2018 as shown in tables 5.5 – 5.8, the difference is not large enough to cause a significant change in the accuracy reports at site A-2018 where there was only 1.3% reduction in placer pixels for

U-Net models. However, at site A-2014 wherein there was relatively low placer pixels (i.e., total of 3399 placer pixels), and given that the site posed challenges to both RF and U-Net models, 14% reduction in placer pixels may have some impact on U-Net's accuracy reports.

Two things are worth noting on the classification results from site A. Firstly, site A was not used for model training; as such, the characteristics of that particular site were not seen by the model. This appears to be more of a factor for the U-Net models (which had some difficult classifying this site) than for the RF models. Secondly, for both models there was a major difference in performance between the 2014 and 2018 images.

It has been shown that given small datasets, RF models can perform at par with deep-learning models containing few hidden layers (Benkendorf and Hawkins, 2020). Additionally, model architecture (e.g., layer depth), input tile size and training strategies adopted are known to influence performance (Ghorbanzadeh et al., 2019). However, for potential real-world/field deployment, it is essential to implement a fairly light and computationally effective U-Net model with reduced parameters (Gadosey et al., 2020). Though we applied data augmentation to increase the training sample size for U-Net model, independent validation datasets with different feature distributions such as in site A can prove challenging for deep-learning models. The attention cropping strategy proposed by Xiao et al. (2018) is one possible and effective data augmentation approach, but we did not investigate this alternative. Transfer learning in which pretrained models and parameters are employed in model development is one of the effective techniques to circumvent overfitting while improving model performance, especially where there is limited amount of training data (Oquab et al., 2014; Yosinski et al., 2014). For instance, the VGG19 model was utilized for transfer learning but modified to FCN suitable for semantic segmentation of slums (Nowakowski et al., 2021). Cui et al. (2020) also modified U-Net with

dense connections for landuse/landcover classification and found that transfer learning effectively deals with insufficient and imbalanced data problems.

The models' error rates on average were relatively low throughout site C. As mentioned earlier, sites C2-2013 and C2-2019 images were a sub-set of site C data which were not used for model training. The features distribution as well as spectral signatures are therefore closely similar to the training dataset. This suggests that so long as representative samples are obtained for model training, reasonable classifications can be obtained. Examining a sub area of site A in Figure 5.9 row (b), we can see that in addition to significant placer disturbance, the images differ markedly in illumination characteristics. Furthermore, profound phenological difference exists between the images probably due to the differences in image acquisition times. Similarly, we observed differences in placer classification accuracy between site C in 2013 and 2019, likely due to differences in illumination conditions (Figures 5.10) and Figure 5.A1 in Appendix 5A. These differences could be due to atmospheric constituents or solar illumination with respect to image acquisition (July versus September).

The RF model is pixel-wise and encodes no texture or contextual information (Stoian et al., 2019); hence false positives are expected, especially for spectrally related classes. This probably resulted in misclassification of some classes; for example, placer as flood-plain/natural disturbance. Also, certain areas further away from flood plains (i.e., forest areas) at sites A and C were detected as placer by the RF; a visual examination shows that such areas consist of dry vegetation interspersed with bare/exposed soil. The SAVI model (see Figure 5.A2 in Appendix 5A) which accounts for exposed surfaces/soil was able to reduce such false positive rates but overall, the accuracy was not substantially different from the NDVI models' performance. Addition of vegetation index channels as model inputs tended to be more important for U-Net

models than for RF models. RF error rates can be reduced by using RF with conditional random field graph model to derive the contextual information (Sun et al., 2017). For example, to remove classification errors and improve accuracy, Tian et al. (2016) suggest incorporating phenological difference and textural information from grey level co-occurrence matrix, while Fu et al. (2017b) emphasize adopting object-based modelling to incorporate contextual information. Izquierdo-Verdiguier and Zurita-Milla (2020) suggest using RF with guided regularization as it does not require setting a threshold on the number of features to be selected and yields non-redundant and representative features. Given that placer occurrences are closer to river reaches, we believe that the RF has potential for field deployment despite the observed false positive rates in forested areas.

The SPOT data used in this analysis was sufficient for the classification task. While it is shown that higher resolution data (e.g., World-View-2 data) would likely result in better accuracies, especially for the U-Net (Melville et al., 2018), this may come at a cost of geographic and historical archive. Data fusion (e.g., SPOT-6/7 and SAR data) may help improve accuracy (Liu et al., 2018c; Yu et al., 2018b), yet this may increase data processing requirements for end-users. Additional data preprocessing might however be needed to ensure image data are adequately prepared for model training. This is especially true when using a model to predict placer at sites beyond the sites used to fit the model as seen in sites A-2014 and A-2018.

Recently, U-Net has been applied to map small-scale mining using Sentinel-2 dataset and shown to effectively detect mined areas with high accuracy (Gallwey et al., 2020). Hence, the U-Net modelling framework has potential for mapping placer disturbance at watershed scale; nevertheless, we suggest developing both models further in parallel for the following reasons. Firstly, RF models are relatively easy to set-up and implement, and given input data resources,

can be used as a reference model for comparison. Secondly, unlike U-Net, RF models require less computational resources and technical expertise to run and parameterize, which may be critical factors in an operational setting (Stoian et al., 2019).

5.5.1 Limitations and future modeling improvement

We find image selection to be one of the crucial steps in the modelling framework; therefore, a set of guidelines on imagery selection could be specified for a given model, for example (e.g., image acquired in August/September are likely to improve classification accuracy). For instance, a study conducted by Noi Phan et al. (2020) finds that datasets with summer scenes covering June and September produced the best accuracy as such images reflect features with closely similar scene characteristics (e.g., illumination and vegetation phenology).

Given that we saw some differences in illumination and atmospheric conditions which affected our results, we should strive as much as possible to reduce atmospheric noise in imagery and radiometric differences in input imagery prior to model training. There are two possible approaches to do this; one is to preprocess and normalize input data as much as possible to reduce artefactual differences; the second is to capture enough variability in illumination differences within the training samples to capture these characteristics within model features.

One limitation during training and validation dataset generation is that digitizing specific land cover classes (e.g., placer, floodplain, and natural disturbance) was difficult for ~10% of the sites. This resulted in 6 out of 75 potential sites (i.e., placer parcels) being discarded and there is the potential for errors in the sites and landscape features that were classified. Since there was no ground truth data for wall-to-wall mapping, and the independent validation data was based on landcover types that were manually interpreted during digitizing process, the accuracy of our

assessment metrics is subject to potential errors from incorrectly identifying class types. Improving training and validation data using ground truth samples identified by field experts would likely reduce error rates for both models and improve confidence in derived outputs (i.e., classification maps).

One specific variable to incorporate into the models' input is digital elevation model (DEM) data. As placer mining occurs in close proximity to waterways, Height Above Nearest Drainage (HAND) (Rennó et al., 2008), which is essentially a topographic normalization of a drainage network using existing DEM, could be a potential variable for improving placer detection in future modelling.

Understanding the intensity of disturbance in each placer parcel would be useful for monitoring and decision-making. To this end, we suggest future modeling frameworks consider converting model outputs of pixels classified as placer to create an indicator of total area of placer per parcel.

5.6 Conclusion

Understanding the impacts of large-scale land alteration requires accurate estimates of the amount of land altered by human development. We used U-Net convolutional neural network and RF models to detect placer mining activity. We compared the accuracy of the two modeling approaches and variables (e.g., NDVI and SAVI) to gain insights into the feasibility of using these tools for monitoring placer mining activity in the Yukon. Our analysis demonstrated the utility of SPOT-6/7 satellite imagery for detecting placer activity. Overall, the user's accuracy was more variable (even only within site C) for both modelling frameworks. For areas with larger areas of placer activity (e.g., site A-2018), both models performed well. Given that the

user's accuracy for U-Net was higher than RF for most of the sites, we believe that the U-Net has potential to improve misclassification errors with additional training data and model development.

It is worth noting too that the technical requirements and computation time for these modelling frameworks are not equal. RF models behave in much the same way as other modelling tools for image data, employing ground truth samples and input image data, and generating classifications for outputs. RF modelling as was completed here is relatively easy, requiring use of GIS software for image processing and generating training samples, and statistical software or remote sensing software for fitting classification models. U-Net modelling requires more extensive parameterization in terms of defining a tile size, methods to reduce overfitting, and greater computation times. However, once U-Net models are fit they can be utilized in much the same way as RF models.

Authors' contributions

Karim Malik – data processing, conceptualization, and writing; Colin Robertson – conceptualization and writing; Douglas Braun – writing; Clara Greig – data processing.

Acknowledgements

The authors would like to thank the Natural Sciences and Engineering Research Council of Canada for funding Dr. Robertson to conduct this research. The authors also acknowledge funding provided by the Fisheries and Oceans Canada to obtain SPOT-6/7 data.

References

Abaidoo, C. A., Osei Jnr, E. M., Arko-Adjei, A., & Prah, B. E. K. (2019). Monitoring the Extent of Reclamation of Small Scale Mining Areas Using Artificial Neural Networks. *Heliyon*, *5*(4), e01445.

Almeida-filho, R., & Shimabukuro, Y. E. (2000). Detecting areas disturbed by gold mining activities through JERS-1 SAR images, Roraima State, Brazilian Amazon. *International Journal of Remote Sensing*, *21*(17), 3357–3362.

Asner, G. P., Llactayo, W., Tupayachi, R., & Luna, E. R. (2013). Elevated rates of gold mining in the Amazon revealed through high-resolution monitoring. *Proceedings of the National Academy of Sciences of the United States of America*, *110*(46), 18454–18459.

Benkendorf, D. J., & Hawkins, C. P. (2020). Effects of sample size and network depth on a deep learning approach to species distribution modeling. *Ecological Informatics*, *60*, 101137.

Breiman, L. (2001). Random forests. *Machine Learning*, 5–32.

Buda, M., Maki, A., & Mazurowski, M. A. (2018). A systematic study of the class imbalance problem in convolutional neural networks. *Neural Networks*, *106*, 249–259.

Buschbacher, K., Ahrens, D., Espeland, M., & Steinhage, V. (2020). Image-based species identification of wild bees using convolutional neural networks. *Ecological Informatics*, *55*, 101017.

Canovas-Garcia, F., Alonso-Sarria, F., Gomariz-Castillo, F., & Onate-Valdivieso, F. (2017). Modification of the random forest algorithm to avoid statistical dependence problems when classifying remote sensing imagery. *Computers and Geosciences*, *103*, 1–11.

Chen, S., & Rao, P. (2008). Land degradation monitoring using multi-temporal Landsat TM/ETM data in a transition zone between grassland and cropland of northeast China.

International Journal of Remote Sensing, *29*(7), 2055–2073.

Chen, W., Li, X., & Wang, L. (2020). Fine land cover classification in an open pit mining area using optimized support vector machine and world view-3 imagery. *Remote Sensing*, *12*(1), 12–14.

Conglaton, R. G. (1991). A Review of Assessing the Accuarcy of Classifications of Remotely Sensed Data. *Remote Sensing of Environment*, *37*(1), 35–46.

Crum, W. R., Camara, O., & Hill, D. L. G. (2006). Generalized overlap measures for evaluation and validation in medical image analysis. *IEEE Transactions on Medical Imaging*, *25*(11), 1451–1461.

Csurka, G., Larlus, D., & Perronnin, F. (2013). What is a good evaluation measure for semantic segmentation? *BMVC 2013 - Electronic Proceedings of the British Machine Vision Conference 2013*.

Cui, B., Chen, X., & Lu, Y. (2020). Semantic Segmentation of Remote Sensing Images Using Transfer Learning and Deep Convolutional Neural Network with Dense Connection. *IEEE Access*, *8*, 116744–116755.

Dlamini, L. Z. D., & Xulu, S. (2019). Monitoring mining disturbance and restoration over RBM site in South Africa using landtrendr algorithm and landsat data. *Sustainability (Switzerland)*, *11*(24).

Dwyer, J. L., Roy, D. P., Sauer, B., Jenkerson, C. B., Zhang, H. K., & Lymburner, L. (2018). Analysis ready data: Enabling analysis of the landsat archive. *Remote Sensing*, *10*(9), 1–19.

Egidarev, E. G., & Simonov, E. A. (2015). Assessment of the environmental effect of placer gold mining in the Amur river basin. *Water Resources*, *42*(7), 897–908.

Espejo, J. C., Messinger, M., Román-Dañobeytia, F., Ascorra, C., Fernandez, L. E., & Silman,

M. (2018). Deforestation and forest degradation due to gold mining in the Peruvian Amazon: A 34-year perspective. *Remote Sensing*, *10*(12), 1–17.

Flood, N., Watson, F., & Collett, L. (2019). Using a U-net convolutional neural network to map woody vegetation extent from high resolution satellite imagery across Queensland, Australia. *International Journal of Applied Earth Observation and Geoinformation*, *82*, 101897.

Fu, B., Wang, Y., Campbell, A., Li, Y., Zhang, B., Yin, S., Xing, Z., & Jin, X. (2017b). Comparison of object-based and pixel-based Random Forest algorithm for wetland vegetation mapping using high spatial resolution GF-1 and SAR data. *Ecological Indicators*, *73*, 105–117.

Gadosey, P. K., Li, Y., Agyekum, E. A., Zhang, T., Liu, Z., Yamak, P. T., & Essaf, F. (2020). SD-UNET: Stripping down U-net for segmentation of biomedical images on platforms with low computational budgets. *Diagnostics*, *10*(2).

Gallwey, J., Robiati, C., Coggan, J., Vogt, D., & Eyre, M. (2020). A Sentinel-2 based multispectral convolutional neural network for detecting artisanal small-scale mining in Ghana: Applying deep learning to shallow mining. *Remote Sensing of Environment*, *248*, 111970.

Ghorbanzadeh, O., Blaschke, T., Gholamnia, K., Meena, S. R., Tiede, D., & Aryal, J. (2019). Evaluation of different machine learning methods and deep-learning convolutional neural networks for landslide detection. *Remote Sensing*, *11*(2).

Ghoshal, S., James, L. A., Singer, M. B., & Aalto, R. (2010). Channel and floodplain change analysis over a 100-year period: Lower Yuba river, California. *Remote Sensing*, *2*(7), 1797–1825.

Gilvear, D. J., Waters, T. M., & Milner, A. M. (1995). Image analysis of aerial photography to quantify changes in channel morphology and instream habitat following placer mining in interior Alaska. *Freshwater Biology*, *34*(2), 389–398.

Han, Z., Dian, Y., Xia, H., Zhou, J., Jian, Y., Yao, C., Wang, X., & Li, Y. (2020). Comparing fully deep convolutional neural networks for land cover classification with high-spatial-resolution gaofen-2 images. *ISPRS International Journal of Geo-Information*, *9*(8).

He, H., & Garcia, E. A. (2009). Learning from imbalanced data. *Studies in Computational Intelligence*, *21*(9), 1263–1284.

Hurcom, S. J., & Harrison, A. R. (1998). The NDVI and spectral decomposition for semi-arid vegetation abundance estimation. *International Journal of Remote Sensing*, *19*(16), 3109–3125.

Ibrahim, E., Lema, L., Barnabé, P., Lacroix, P., & Pirard, E. (2020). Small-scale surface mining of gold placers: Detection, mapping, and temporal analysis through the use of free satellite imagery. *International Journal of Applied Earth Observation and Geoinformation*, *93*, 102194.

Ioffe, S., & Szegedy, C. (2015). Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift. *In: International Conference on Machine Learning, Lille, France.*, 448–456.

Izquierdo-Verdiguier, E., & Zurita-Milla, R. (2020). An evaluation of Guided Regularized Random Forest for classification and regression tasks in remote sensing. *International Journal of Applied Earth Observation and Geoinformation*, *88*, 102051.

Jiang, X., Pang, Y., Li, X., Pan, J., & Xie, Y. (2018). Deep neural networks with Elastic Rectified Linear Units for object recognition. *Neurocomputing*, *275*, 1132–1139.

Khan, S. H., Hayat, M., Bennamoun, M., Sohel, F. A., & Togneri, R. (2018). Cost-sensitive learning of deep feature representations from imbalanced data. *IEEE Transactions on Neural Networks and Learning Systems*, *29*(8), 3573–3587.

Kingma, D. P., & Ba, J. L. (2015). Adam: A method for stochastic optimization. *3rd International Conference on Learning Representations, ICLR 2015 - Conference Track Proceedings*, 1–15.

Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). ImageNet Classification with Deep Convolutional Neural Networks. *Advances In Neural Information Processing Systems*, 1–9.

Lobo, F. de L., Souza-Filho, P. W. M., Novo, E. M. L. de M., Carlos, F. M., & Barbosa, C. C. F. (2018). Mapping mining areas in the Brazilian amazon using MSI/Sentinel-2 imagery (2017). *Remote Sensing*, *10*(8).

Madhuanand, L., Sadavarte, P., Visschedijk, A. J. H., Denier Van Der Gon, H. A. C., Aben, I., & Osei, F. B. (2021). Deep convolutional neural networks for surface coal mines determination from sentinel-2 images. *European Journal of Remote Sensing*, *54*(1), 296–309.

Marcus, W. A. (2002). Mapping of stream microhabitats with high spatial resolution hyperspectral imagery. *Journal of Geographical Systems*, *4*(1), 113–126.

Matthews, N. (2016). People and Fresh Water Ecosystems: Pressures, Responses and Resilience. *Aquatic Procedia*, *6*, 99–105.

Maxwell, A. E., Pourmohammadi, P., & Poyner, J. D. (2020). Mapping the topographic features of mining-related valley fills using mask R-CNN deep learning and digital elevation data. *Remote Sensing*, *12*(3), 1–23.

Melville, B., Lucieer, A., & Aryal, J. (2018). Object-based random forest classification of

Landsat ETM+ and WorldView-2 satellite imagery for mapping lowland native grassland communities in Tasmania, Australia. *International Journal of Applied Earth Observation and Geoinformation*, *66*, 46–55.

Mhangara, P., Tsoeleng, L. T., & Mapurisa, W. (2020). Monitoring the development of artisanal mines in South Africa. *Journal of the Southern African Institute of Mining and Metallurgy*, *120*(4), 299–306.

Nellis, M. D., Harrington, J. A., & Wu, J. (1998). Remote sensing of temporal and spatial variations in pool size, suspended sediment, turbidity, and Secchi depth in Tuttle Creek Reservoir, Kansas: 1993. *Geomorphology*, *21*(3–4), 281–293.

Nelson, A. D., & Church, M. (2012). Placer mining along the Fraser River, British Columbia: The geomorphic impact. *Bulletin of the Geological Society of America*, *124*(7–8), 1212–1228.

Noi Phan, T., Kuch, V., & Lehnert, L. W. (2020). Land cover classification using google earth engine and random forest classifier-the role of image composition. *Remote Sensing*, *12*(15).

Nowakowski, A., Mrziglod, J., Spiller, D., Bonifacio, R., Ferrari, I., Mathieu, P. P., Garcia-Herranz, M., & Kim, D. H. (2021). Crop type mapping by using transfer learning. *International Journal of Applied Earth Observation and Geoinformation*, *98*, 102313.

Obodai, J., Adjei, K. A., Odai, S. N., & Lumor, M. (2019). Land use/land cover dynamics using landsat data in a gold mining basin-the Ankobra, Ghana. *Remote Sensing Applications: Society and Environment*, *13*, 247–256.

Oquab, M., Bottou, L., Laptev, I., & Sivic, J. (2014). Learning and Transferring Mid-Level Image Representations using Convolutional Neural Networks. *IEEE Conference on Computer Vision and Pattern Recognition*, 1717–1724.

Renno, C. D., Nobre, A. D., Cuartas, L. A., Soares, J. V., Hodnett, M. G., Tomasella, J., & Waterloo, M. J. (2008). HAND, a new terrain descriptor using SRTM-DEM: Mapping terra-firme rainforest environments in Amazonia. *Remote Sensing of Environment*, *112*(9), 3469–3481.

Ronneberger, O., Fischer, P., & Brox, T. (2015). U-Net: Convolutional Networks for Biomedical Image Segmentation. *In International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, Cham.*, *9351*, 234–241.

Solana-Gutiérrez, J., Rincon, G., Alonso, C., & Garcia-de-Jalon, D. (2017). Using fuzzy cognitive maps for predicting river management responses: A case study of the Esla River basin, Spain. *Ecological Modelling*, *360*, 260–269.

Stoian, A., Poulain, V., Inglada, J., Poughon, V., & Derksen, D. (2019). Land cover maps production with high resolution satellite image time series and convolutional neural networks: Adaptations and limits for operational systems. *Remote Sensing*, *11*(17), 1–26.

Sudre, C. H., Li, W., Vercauteren, T., Ourselin, S., & Jorge Cardoso, M. (2017). Generalised dice overlap as a deep learning loss function for highly unbalanced segmentations. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, *10553 LNCS*, 240–248.

Sun, X., Lin, X., Shen, S., & Hu, Z. (2017). High-resolution remote sensing data classification over urban areas using random forest ensemble and fully connected conditional random field. *ISPRS International Journal of Geo-Information*, *6*(8).

Syrris, V., Hasenohr, P., Delipetrev, B., Kotsev, A., Kempeneers, P., & Soille, P. (2019). Evaluation of the Potential of Convolutional Neural Networks and Random Forests for Multi-Class Segmentation of Sentinel-2 Imagery. *Remote Sensing*, *11*(8), 907.

Tharwat, A. (2018). Classification assessment methods. *Applied Computing and Informatics*, *17*(1), 168–192.

Tian, S., Zhang, X., Tian, J., & Sun, Q. (2016). Random forest classification of wetland landcovers from multi-sensor data in the arid region of Xinjiang, China. *Remote Sensing*, *8*(11), 1–14.

Wang, Caixia, Pavlowsky, R. T., Huang, Q., & Chang, C. (2016). Channel bar feature extraction for a mining-contaminated river using high-spatial multispectral remote-sensing imagery. *GIScience and Remote Sensing*, *53*(3), 283–302.

Wang, Chunsheng, Chang, L., Zhao, L., & Niu, R. (2020). Automatic identification and dynamic monitoring of open-pit mines based on improved mask R-CNN and transfer learning. *Remote Sensing*, *12*(21), 1–20.

Werner, T. T., Bebbington, A., & Gregory, G. (2019). Assessing impacts of mining: Recent contributions from GIS and remote sensing. *Extractive Industries and Society*, *6*(3), 993–1012.

Whiteside, T. G., & Bartolo, R. E. (2015). Use of WorldView-2 time series to establish a wetland monitoring program for potential offsite impacts of mine site rehabilitation. *International Journal of Applied Earth Observation and Geoinformation*, *42*, 24–37.

Woldai, T. (2001). Application of remotely sensed data and GIS in assessing the impact of mining activities on the environment. *Proceedings–17th International Mining Congress & Exhibition of Turkey*, 75–84.

Wright, A., Marcus, W. A., & Aspinall, R. (2000). Evaluation of multispectral, fine scale digital imagery as a tool for mapping stream morphology. *Geomorphology*, *33*(1–2), 107–120.

Xiao, Q., Li, G., Xie, L., & Chen, Q. (2018). Real-world plant species identification based on

deep convolutional neural networks and visual attention. *Ecological Informatics*, *48*, 117–124.

Yosinski, J., Clune, J., Bengio, Y., & Lipson, H. (2014). How transferable are features in deep neural networks? 1–9.

Yu, L., Xu, Y., Xue, Y., Li, X., Cheng, Y., Liu, X., Porwal, A., Holden, E. J., Yang, J., & Gong, P. (2018a). Monitoring surface mining belts using multiple remote sensing datasets: A global perspective. *Ore Geology Reviews*, *101*, 675–687.

Yu, Y., Li, M., & Fu, Y. (2018b). Forest type identification by random forest classification combined with SPOT and multitemporal SAR data. *Journal of Forestry Research*, *29*(5), 1407–1414.

Zhang, C., Yue, P., Tapete, D., Shangguan, B., Wang, M., & Wu, Z. (2020). A multi-level context-guided classification method with object-based convolutional neural network for land cover classification using very high resolution remote sensing images. *International Journal of Applied Earth Observation and Geoinformation*, *88*, 102086.

Appendix 5A

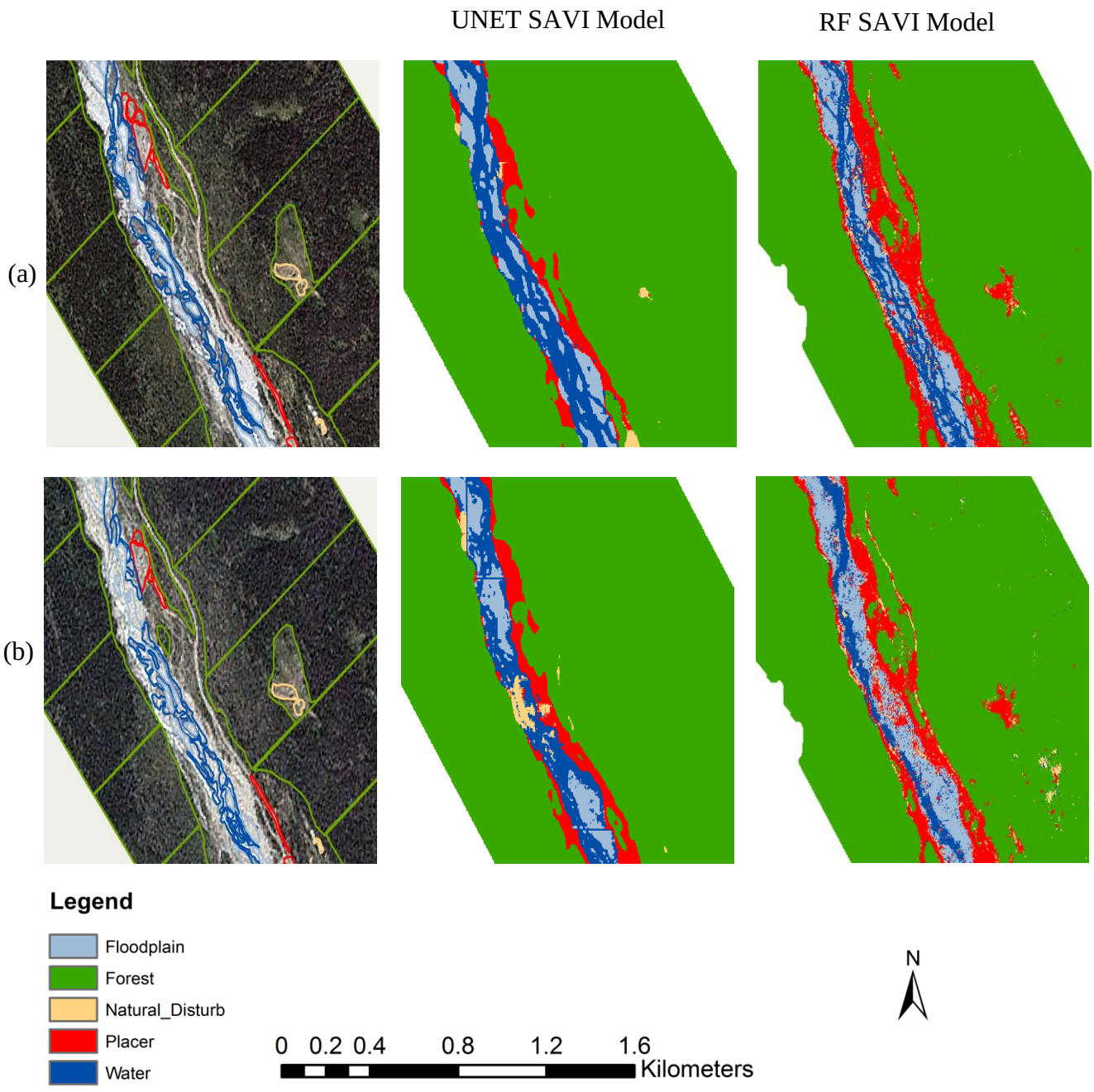


Figure A1. U-Net SAVI and RF SAVI models' placer detection maps using in 2013 and 2019. (a) site C2-2013, and (b) site C2-2019.

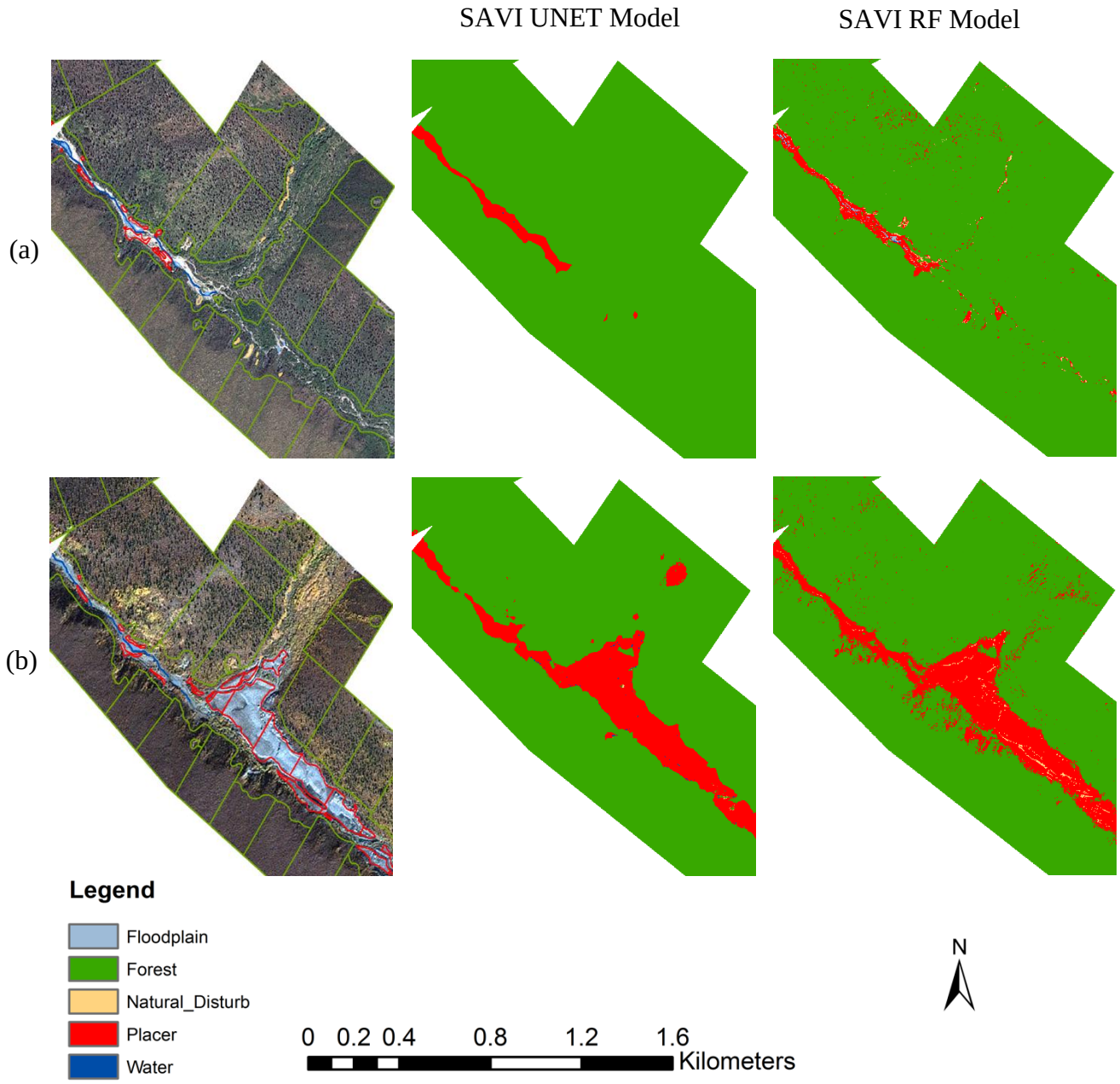


Fig. A2. SAVI U-Net and SAVI RF models' placer detection maps using in 2014 and 2018. (a) site A-2014, and (b) site A-2018.

Chapter 6: Conclusions

6.1 Overview

In this research, computer vision methods capability and potential adoption for geographical research are addressed using simulation-based data as well as real-word datasets. Specifically, the research provides an in-depth insight into both simple and complex computer vision tools, and how they can be employed to extract spatial pattern information, a crucial ingredient for quantifying and characterizing changes and similarities in patterns and processes of interest across space and time. Chapter 2 provides a compilation of literature and explores scale domains and how computer vision algorithms are developed to handle scale issues. As theories on scale and techniques to address scale challenges are well matured in geography, computer vision discipline's approach to resolve scale issues is not entirely new to geography, and in fact, geographers and GIScientist do not only have a fundamental role to play in adapting computer vision methods but can contribute to the development of scale and context-aware models.

Spatial structure is known to be the defining characteristics of patterns generated by spatial processes acting of a range of scales. While illumination and contrast are frequently occurring components in remotely sensed data, these are regarded as fundamental artifacts that substantially impede the performance of many classical pattern comparison metrics. Chapter 3 addresses this challenge by comparing patterns in simulated data and snow water equivalent maps using two structurally-sensitive computer vision algorithms (i.e., SSIM and CWSSIM). By accounting for contrast and illumination differences between image-reference pairs, the indices focus more on spatial structure when comparing patterns.

Chapter 4 attempts to avoid the challenges posed by illumination, contrast and other implicit data artifacts through the use of deep learning models. Extracting relevant pattern signatures from trained CNN models and comparing patterns using CNN feature maps precludes the need to deal with raw image data which is vulnerable to the above artifacts. Such a novel approach demonstrates the potential of CNN features for environmental monitoring. Deep learning models come with various architectures aimed at optimizing model performance. U-Net CNN which is known for effectively segmenting biomedical images is presented in Chapter 5 to detect placer disturbance signatures at watershed scales. U-Net model is a typical example of context-aware deep learning modelling, a novel and promising tool for detecting and monitoring placer mining disturbance.

6.2 Contributions of this research

This research points to the potential for computer vision methods and tools to be easily adopted to resolve pattern comparison challenges in geography and GIScience. There are well-known theories and methods for representing and analyzing spatial patterns in geography. For example, the awareness of spatial structure and correlation, spatial interpolation methods, and other geostatistical techniques that been employed to study patterns in geography and GIScience (Atkinson and Tate, 2000), are fundamental to sampling and processing data to develop computer vision tools. For instance, scaling image data to represent spatial objects at varying spatial resolutions is a common data augmentation technique in computer vision whose underlying objective can be well understood in geographic context.

The literature synthesized in Chapter 2 illuminates the potential of computer vision algorithms to be used in spatial pattern comparison applications, especially for continuous-value data. A

scale typology outlines fundamental challenges in parameterizing computer vision algorithms to capture the signals of pattern generating processes, and how to address scale issues as they emerge in pattern comparison tasks. As scale is a core geographical concept, this may also be an area where geographers and GIScientists can help to facilitate adoption and refinement of computer vision tools for environmental monitoring. For instance, in geography, it is widely recognized that remotely sensed data is not immune from MAUP and ecological fallacy (Openshaw, 1984; Marceau and Hay, 1999). Changing scale and aggregation level, for example, is proven to substantially reduce per-class classification accuracy (Marceau et al., 1994). Therefore, parametrizing models based on the possible effects of scale is likely to lead to generalizable models and more predictable results. To partly address scale issues when developing CNN models, an appropriate dimension of scale – pixel spatial resolution (i.e., grain) and input tile size (i.e., spatial extent) is crucial. Another viable approach is to develop texture encoded models in which fine-grained spatial information residing in lower layers is propagated across to augment coarser features in higher layers. Additionally, selecting the best model parameters given prior knowledge of process-pattern spatial and temporal resolutions is an effective technique. For example, while processes that manifest themselves as fine-grained patterns may require small-sized filters (e.g., 3×3), in order extract discriminative signatures, coarser pattern signatures may be best extracted by deploying large filters (e.g., 15×15) (Peng et al., 2017). Finally, multiscale CNNs (e.g., Zhao and Du, 2016; Liu et al., 2018b), can be used to mitigate the MAUP problem as such models can take tiles with varying dimensions or scaled (i.e., zoomed-in or zoomed-out images) data as their input. This approach aligns well with the multiscale Object-Specific framework suggested by Hay et al. (2001).

As discussed in Chapter 2, spatial patterns tend to exhibit unique structure that is linked to underlying processes; yet at varying scales, these patterns manifest themselves differently (Wu, 2004; Kulha et al., 2019), requiring methods and tools that are relatively robust to changes in scale. Aside artifacts that may be introduced due to the range of scale over which patterns are studied, illumination, contrast, image misregistration and shift in sensor platforms are amongst potential sources of errors in remotely sensed data that may impact the accuracy of pattern comparison (Townshend et al., 1992; Coppin et al., 2004). Additionally, given that patterns are shaped by complex dynamic processes acting at various scales, classical pattern metrics that quantify static, discrete overall pattern or individual patch properties may not be suitable for continuous-valued data or data derived from model simulations (Bolliger et al., 2007). Structurally sensitive metrics (e.g., SSIM and CWSSIM) are key to resolving some of the foregone challenges. By examining exemplar computer vision algorithms, Chapter 3 revealed the potential of SSIM and CWSSIM to discover changes in patterns. The research elucidated the geometric invariant property and robustness of CWSSIM to frequently encountered errors such as shifts in data acquisition instruments. When it comes to pattern comparison in continuous-valued data, structurally sensitive tools remain somewhat limited in geography. Therefore, this finding is likely to pave a way for adoption of existing as well as emerging computer vision tools for spatial pattern comparison in geographic research.

Chapter 4 extends the work in Chapter 3 by developing a texture-based CNN model to extract feature maps for landscape similarity analysis. Feature maps represent the most discriminative pattern signatures learned by a deep learning model and thus may contain vital information pertaining to patterns generated by spatial processes (Amirshahi et al., 2017). The study demonstrated the utility of deep learning models in a landscape comparison context. More

importantly, the novelty of the approach stems from combining multiple techniques: encoding texture information, the utilization of CNN feature maps, and deriving feature map histograms to compare landscapes. Histograms of a given landscape derived from its corresponding CNN feature maps also retain the distribution of the discriminative features. Such distributions hold potential to represent the complexity and uniqueness of a given landscape, and hence could help resolve spatial pattern comparison challenges. By employing this approach, Chapter 4 demonstrates that similarity distributions could be established for any landscape type to aid in comparison and to effectively distinguish disparate landscapes. This technique would be essential for developing tools aimed at discovering spatial processes driving for example, a particular land degradation pattern across a variety of geographical locations.

The analysis carried out in Chapter 5 builds on the texture-based model presented in Chapter 4 by introducing a more spatially explicit and contextual-aware variant of CNN models – U-Net, and comparing the model's performing with RF. The study revealed the potential of both RF and U-Net for watershed scale monitoring in real-world scenarios. By combing SPOT-6/7 data and derived data (e.g., RGBN + SAVI and RGBN + NDVI), this research further illuminates the data requirements for developing models to detect Placer mining disturbance and showed that SAVI and NDVI may be helpful for improving the discriminative performance of U-Net models. The findings of this research bring to light the potential complementarity of both modelling frameworks. While the U-Net models reduced misclassification errors, the RF models were found to be easily trainable with less modelling and computational demands.

6.3 Adopting computer vision methods in geography: challenges and opportunities

In this study, both the potentials and drawbacks of computer vision methods in spatial pattern comparison are highlighted. The applicability of computer vision methods in geographic context, however, requires further scrutiny with experimental approaches that examine process-pattern relationships at varying scales, and across multimodal datasets (e.g., vector, raster, radar and lidar). It is also worth noting that, like classical approaches to pattern comparison, computer vision methods do not provide adequate explanations of the relationships between patterns and processes (Reichstein et al., 2019; Gahegan, 2020). Although neural networks learn underlying data generating processes, the models are unable to explain how processes operate to create patterns, and process-pattern linkages learned by models may not resemble how real-world processes and patterns interact within a given problem domain (Gahegan, 2020). Integrating a process-based framework for pattern comparison proposed in Csillag and Boots (2005) with current techniques in computer vision may offer a more valuable insight into understanding process-pattern interactions, though this may involve extensive and statistically defensible evaluation of several candidate pattern generating processes.

Chapter 3 demonstrated the potential of CWSSIM and SSIM indices for pattern comparison in continuous-valued data; however, there are caveats underlying their applicability, especially for geographic phenomena comparison. The simulation-based approach presented using GMR fields sheds light on the challenges likely to be encountered when using simulation to model real geographic processes. For example, mvr3 and mvr8 images were generated by disparate spatial dependence parameter; yet the metrics, in some instances, could not distinguish these patterns. This highlights limits to the process-pattern framework as a whole and what can be learned by pattern analysis alone. Also, as shown in this study, both CWSSIM and SSIM are sensitive to

change in scale (i.e., window size). In geographic phenomena comparison, selecting an appropriate scale at which processes operate, and interact with patterns is fundamental to uncovering changes or similarity in spatial patterns (Frate et al., 2014). Thus, the use of CWSSIM and SSIM requires thorough examination of the window size to effectively compare patterns. For instance, CWSSIM's relative resilience to shift in scale is only valid for "small" changes in scale relative to the size of the wavelet filter (Sampat et al. 2009). Given that geometric errors in geographic datasets can span scales ranging from sub-meters to kilometers, the utility of CWSSIM's geometric invariance property cannot be guaranteed at large scales (e.g., datum shift between NAD27 versus WGS84). Moreover, extensive evaluation of the performance of wavelet domain indices such as CWSSIM has been conducted on generic images (Zhang et al., 2007; Yang et al., 2008; Gao et al., 2011). The robustness of CWSSIM therefore remains to be extensively assessed in datasets representing real-world geographic phenomena. Unlike CWSSIM, several authors have demonstrated the superior performance of SSIM in comparing geographic patterns (e.g., Robertson et al., 2014; Jones et al., 2016; Wiederholt et al., 2019). However, there is no empirical assessment of CWSSIM and SSIM perceptual similarity property against human observers' judgment of pattern similarity in the context of patterns generated by geographic processes.

Convolutional neural networks are a class of spatial-aware models which have attained high accuracy in solving a variety of geographic related problems such as scene classification, object detection and image segmentation (Krizhevsky et al., 2012; Lecun et al., 2015). The potential of CNNs to model complex non-linear relationships peculiar to geographic patterns and processes makes them useful tools for spatial data analysis (Fischer, 2001). Convolutional neural networks do not rely on pixel level information but leverage contextual information using filters to detect

and classify patterns. The adoption of CNNs for analyzing patterns is thus likely to improve the characterization of spatial and temporal patterns leading to the discovery of processes that govern pattern formation over a range of scales (Brodrick et al., 2019). Despite this promising potential, a widespread adoption and deployment of CNNs for geographic problem solving will not be as straightforward as found in computer vision. For instance, deep learning algorithms require a huge amount of labelled data for training, validation, and testing, and this motivates deep learning researchers to turn to benchmark datasets (Zhu et al., 2017; Reichstein et al., 2019). Benchmark datasets are generally limited in geography, and this can impair the ability of geographers to rigorously validate models in geographic context. To add to this challenge, satellite imagery which is one of the primary sources of geospatial data, comes in configurations that are often more complex than the generic images used to develop many deep learning models. This may weaken the generalizability and transferability of models to different locations and datasets. In Chapter 4 for example, the CNN models performed better on AID than on Sentinel-2 data; surprisingly, retraining them with a combination of AID and Sentinnel-2 data substantially improved their performance. This observation underlines the complexity inherent in satellite imagery compared to high spatial resolution aerial images which are relatively simple to classify. In geographic context, datasets such as AID have inherent limitations in terms of class definitions. For example, the definition of what constitutes mountainous pattern versus forested landscape is not mutually exclusive for some training examples but turns out to be a function of scale in geographic context. Clearly, this problem manifests in the comparison of feature map distributions between forest and mountains wherein the distributions tend to overlap. Such an outcome is largely driven by the fact that feature maps from mountains that contain forest are likely to be similar to actual forest landscapes at high spatial scales. Unlike the computer vision

discipline, generating benchmark datasets in geography can be challenging. The perception of similarity in geographic constructs and patterns may be influenced by the level of expertise and field of interest of a geographer (Csillag and Boots, 2005). In geographic context, creating a huge benchmark dataset in which class labels are mutually exclusive will require extensive crowdsourcing using audience from varied sub-fields. Citizen science is a potential platform that can be utilized to obtain labelled data (Robertson and Feick, 2017). However, to ensure labels are mutually exclusive, further filtering using a representative subset of audience in geography will be essential.

The U-Net model is a CNN architecture in which fine spatial resolution features in early layers of the network are utilized to improve the spatial resolution of corresponding higher layer features. The U-Net learning module is spatially explicit in that it applies only fully convolutional layers (i.e., layers with 2-dimensional feature maps). The model is widely utilized to classify biomedical images (e.g., Ronneberger et al., 2015; Chen et al., 2019), and in a variety of landcover mapping and change detection tasks (e.g., Flood et al., 2019; Gallwey et al., 2020). By comparing RF and U-Net, Chapter 5 sheds light on how competitive a non-spatial model – RF can be; suggesting that the most complicated modelling framework may not always be the best technique to detect placer mining disturbance. Chapter 5 further points to the need for careful examination of the approach used to obtain validation data for model testing. As illustrated in the modelling results, the correlation structure in samples from the same spatial location may be easily learned by deep learning models resulting in high performance, yet samples taken from disparate locations tend to be challenging for models to accurately classify.

To accurately assess the performance of deep learning models, large and spatially explicit data are recommended (Elmes et al., 2020). Unlike the computer vision discipline, the limited nature

of labelled data in geography compels researchers to evaluate models on few available samples. In Chapter 5 for example, U-Net and RF performance was assessed based on limited validation data. More often, training and validation samples are assumed to accurately represent the underlying ground features or phenomena; yet errors may result from sampling design and data collection techniques (Elmes et al., 2020). Assessing the accuracy of training and validation samples using expert knowledge is therefore crucial to minimizing data errors. The U-Net and RF models need to be subjected to further scrutiny with substantial amounts of independent validation data in which classes are identified, interpreted, and labelled by experts in geography.

6.4 Key limitations in this research

One fundamental limitation of the review presented in Chapter 2 is that the complexity of computer vision models and the challenges related to explaining and linking models' decisions and/or predictions to underlying input variables was not addressed. Although there has been substantial progress towards explaining the mechanics underlying deep neural networks decisions (e.g., Omeiza et al., 2019; Srinivas and Fleuret, 2019), much remains to be seen when it comes to linking model input variables to outputs. Explaining the linkages between deep learning models inputs and outputs will not only go a long way to improve adoption and deployment, but could potentially lead to effective model parameterization in the geographic context.

In Chapter 3, the assessment of CWSSIM and SSIM capability to compare spatial patterns excluded human subjects' judgement of pattern similarity; an experimental design to compare both metrics similarity values and human observations decisions as to which image pairs are more similar would be crucial to increasing confidence in the use of the metrics in real-life

decision-making systems. Additionally, CWSSIM and SSIM evaluation in simulation data was based on moving average fields with two disparate sets of spatial dependence parameter settings – high spatial dependence (mvr3) and low spatial dependence (mvr8). While these two sets of data were helpful for demonstrating the capability of the metrics to expose subtle differences in patterns between reference-images pairs, examining the metrics sensitivity over a more varying range of spatial dependence parameters would be essential, especially, for deployment in real-world pattern comparison tasks.

The Tex-CNN developed in Chapter 4 though proved to be more discriminative, did not outperform the classical CNN significantly except for mountain landscapes. This leads to many open questions regarding the effectiveness of our texture learning module. Other approaches that retain spatially discriminative features via residual network architecture may warrant future consideration. We also found that the models, trained using high resolution AID imagery, misclassified over 90% of agriculture landscapes as forest in Sentinel dataset. While this outcome may not be surprising owing to the significant difference in spatial resolutions of the two datasets, we anticipated that the Tex-CNN would be relatively robust to varying image resolutions. This observation again calls into question the potential robustness and generalizability of our Tex-CNN to lower resolution datasets. Retraining the models with a combination of AID and Sentinel however produced models with high classification accuracies, suggesting the need for inclusion of both AID and Sentinel datasets at model training phase. Additional limitation of our feature-based approach stems from the fact that when extracting feature maps, we considered only layer-two of the CNN; while this layer tended to yield more visually meaningful and interpretable feature maps, it would be worthy investigating the

discriminative potential of other layer feature maps, especially the fourth layer since most of the models' decisions are made using information from this layer.

In Chapter 5, we maintained the originality of SPOT-6/7 data and therefore did not apply preprocessing techniques, however, it turned out that due to varying image acquisition dates, differences in scene characteristics (e.g., varying illumination and vegetation phenology) likely reduced models' performance, especially at site A-2018. Another limitation and a possible source of error stems from the lack of ground validation of the digitized Placer areas. It is probable that some areas that were identified to be Placer during digitization actually belong to other landcover types or certain actual Placer disturbed areas were not digitized at all and hence were excluded from the data. While omitting actual Placer pixels only reduces sample size, incorrectly identifying Placer pixels could significantly impact results derived from accuracy metrics.

6.5 Future research directions

In this research, the caveats and limitations of computer vision methods were identified; such drawbacks could be harnessed to improve adoption and deployment of computer vision methods and tools for environmental resource monitoring. We suggest future research examines CWSSIM and SSIM on images with more complex spatial patterns such as images with distinct edges (e.g., urban environments) and images with visible objects (e.g., forests). The results of these metrics on the aforementioned datasets would provide vital information on the sensitivity and potential of the metrics to compare spatial patterns in real-world problems.

The current data deluge offers enormous opportunity to explore and discover the potentials of computer vision methods in geographical and GIScience research endeavors. Unlike the

disciplines of geography and GIScience, the computer vision discipline has utilized the recent data-rich environment and computational resources to advance the development and deployment of pattern detection and recognition algorithms. With the abundance of raw sensor data as well as analysis ready datasets, testbeds for parametrizing computer vision models in geographical context could be designed. To this end, it is crucial that future research focuses on curating and compiling labeled environmental data to create benchmark datasets as shown in (e.g., Basu et al., 2015 and Xia et al., 2016). Given the complex nature of patterns geographers are interested in, such novel datasets will serve as essential testbeds to effectively evaluate computer vision model's potential to compare real-world patterns driven by the complex landscapes processes as well as their interactions at varying scales. For example, labeled datasets with healthy versus disease forest (e.g., mountain pine beetle infestation), mining (e.g., placer disturbance) versus natural disturbance (e.g., landslide) and burned forest versus harvested forest present complex pattern-process interactions whose signatures could be learned by deep learning models for classification and monitoring Earth's resources; yet compilations of such novel datasets are either limited or non-existent in geography.

Chapter 4 investigated the potential utilization of deep learning features to compare landscapes. We believe that CNN feature maps possess useful discriminant features and could be utilized in change detection as well. Thus, future research in this direction is warranted. Furthermore, the CNN feature-based approach can be considered as an important starting point for future studies towards explaining the decisions and/or predictions of deep CNNs in a more geographically relevant context. It is also essential that further research focuses on linking models' inputs to outputs at varying scales to improve confidence and acceptability of models' decisions, and in the utilization of features maps for further tasks (e.g., change detection).

The Placer modelling results in Chapter 5 hold potential for future improvement in many key areas. For example, training and validation datasets could be improved and increased by using ground truth samples identified by field experts to reduce digitization errors. Furthermore, given that differences in illumination and atmospheric conditions impacted the results, future modelling should focus on using images that are acquired with a minimum of at least one month interval or apply preprocessing reduce atmospheric noise and radiometric differences in input imagery prior to model training. For operational mapping and monitoring mining activities in each Placer parcel, we suggest that future research strives to obtain the magnitude of disturbance per parcel by converting areas detected as placers to create disturbance intensity over an entire area of interest.

As CNNs are "data hungry" models, it is worth investigating the utility of fine-tuning (i.e., transfer-learning) and using pretrained networks from other domains (e.g., medical imaging and graphical image processing) to extract discriminative features and pattern signatures. Pretrained models may not only obviate the need for large training sample size, but could alleviate the relatively high computational resources and time required to parameterize and develop deep learning models. However, there are challenging dilemmas in the choice of pretrained CNNs, and we emphasize that models trained on datasets that contain related features of interest to the problem being investigated by geographers and GIScientists should be considered.

Finally, with the advent of cloud computing services, model training and deployment workflow is envisioned to scale-up with sophisticated algorithms been accessible to geographers. However, to increase credibility of such models, their performance must be validated against labeled geographic datasets. Additionally, establishing a library of signatures characterizing certain patterns, is likely to streamline models' performance validation workflows. It must be

stressed that building libraries of pattern signatures would be a convoluted task given the complexity of features/patterns and underlying processes often encountered in real-world.

References

Amirshahi, S. A., Pedersen, M., & Yu, S. X. (2017). Image quality assessment by comparing CNN features between images. *IS and T International Symposium on Electronic Imaging Science and Technology*, 42–51.

Atkinson, P. M., & Tate, N. J. (2000). Spatial Scale Problems and Geostatistical Solutions: A Review. *Professional Geographer*, *52*(4), 607–623.

Basu, S., Ganguly, S., Mukhopadhyay, S., DiBiano, R., Karki, M., & Nemani, R. (2015). *DeepSat - A Learning framework for Satellite Imagery*.

Bolliger, J., Wagner, H. H., & Turner, M. G. (2007). *Identifying and Quantifying Landscape Patterns in Space and Time*. 177–194.

Brodrick, P. G., Davies, A. B., & Asner, G. P. (2019). Uncovering Ecological Patterns with Convolutional Neural Networks. *Trends in Ecology and Evolution*, *34*(8), 734–745.

Chen, Y., Wang, K., Liao, X., Qian, Y., Wang, Q., Yuan, Z., & Heng, P. A. (2019). Channel-Unet: A Spatial Channel-Wise Convolutional Neural Network for Liver and Tumors Segmentation. *Frontiers in Genetics*, *10*, 1–13.

Coppin, P., Jonckheere, I., Nackaerts, K., Muys, B., & Lambin, E. (2004). Digital change detection methods in ecosystem monitoring: a review. *International Journal of Remote Sensing*, *25*(9), 1565–1596.

Csillag, F., & Boots, B. (2005). Toward Comparing Maps as Spatial Processes. *In Developments in Spatial Data Handling*, 641–652.

Elmes, A., Alemohammad, H., Avery, R., Caylor, K., Eastman, J. R., Fishgold, L., Friedl, M. A., Jain, M., Kohli, D., Bayas, J. C. L., Lunga, D., McCarty, J. L., Pontius, R. G., Reinmann, A. B., Rogan, J., Song, L., Stoynova, H., Ye, S., Yi, Z. F., & Estes, L. (2020). Accounting for

training data error in machine learning applied to earth observations. *Remote Sensing*, *12*(6), 1–39.

Fischer, M. M. (2001). Computational Neural Networks - Tools for Spatial Data Analysis. *GeoComputational Modelling*, 15–34.

Flood, N., Watson, F., & Collett, L. (2019). Using a U-net convolutional neural network to map woody vegetation extent from high resolution satellite imagery across Queensland, Australia. *International Journal of Applied Earth Observation and Geoinformation*, *82*, 101897.

Frate, L., Saura, S., Minotti, M., di Martino, P., Giancola, C., & Carranza, M. L. (2014). Quantifying forest spatial pattern trends at multiple extents: An approach to detect significant changes at different scales. *Remote Sensing*, *6*(10), 9298–9315.

Gahegan, M. (2020). Fourth paradigm GIScience? Prospects for automated discovery and explanation from data. *International Journal of Geographical Information Science*, *34*(1), 1–21.

Gallwey, J., Robiati, C., Coggan, J., Vogt, D., & Eyre, M. (2020). A Sentinel-2 based multispectral convolutional neural network for detecting artisanal small-scale mining in Ghana: Applying deep learning to shallow mining. *Remote Sensing of Environment*, *248*, 111970.

Gao, Y., Rehman, A., & Wang, Z. (2011). CW-SSIM Based Image Classification. *IEEE International Conference on Image Processing*, 1249–1252.

Hay, G. J., Marceau, D. J., Dube, P., & Bourchad, A. (2001). A Multiscale Framework for Landscape Analysis : Object-specific analysis and upscaling. *Landscape Ecology*, *16*(514), 1–49.

Jones, E. L., Rendell, L., Pirotta, E., & Long, J. A. (2016). Novel application of a quantitative spatial comparison tool to species distribution data. *Ecological Indicators*, *70*, 67–76.

Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). ImageNet Classification with Deep Convolutional Neural Networks. *Advances In Neural Information Processing Systems*, 1–9.

Kulha, N., Pasanen, L., Holmström, L., De Grandpré, L., Kuuluvainen, T., & Aakala, T. (2019). At What Scales and Why Does Forest Structure Vary in Naturally Dynamic Boreal Forests? An Analysis of Forest Landscapes on Two Continents. *Ecosystems*, *22*(4), 709–724.

Lecun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, *521*(7553), 436–444.

Liu, Y., Zhong, Y., & Qin, Q. (2018). Scene classification based on multiscale convolutional neural network. *IEEE Transactions on Geoscience and Remote Sensing*, *56*(12), 7109–7121.

Marceau, D. J., Gratton, D. J., Fournier, R. A., & Fortin, J. P. (1994). Remote sensing and the measurement of geographical entities in a forested environment. 2. The optimal spatial resolution. *Remote Sensing of Environment*, *49*(93), 105–117.

Marceau, D. J., & Hay, G. J. (1999). Remote sensing contributions to the scale issue. *Canadian Journal of Remote Sensing*, *25*(4), 357–366.

Omeiza, D., Speakman, S., Cintas, C., & Weldermariam, K. (2019). Smooth Grad-CAM++: An Enhanced Inference Level Visualization Technique for Deep Convolutional Neural Network Models. 1–10.

Openshaw, S. (1984). Ecological fallacies and the analysis of areal census data. *Environment & Planning A*, *16*(1), 17–31.

Peng, C., Zhang, X., Yu, G., Luo, G., & Sun, J. (2017). Large kernel matters - Improve semantic segmentation by global convolutional network. *Proceedings - 30th IEEE Conference on*

Computer Vision and Pattern Recognition, CVPR 2017, 1743–1751.

Reichstein, M., Camps-Valls, G., Stevens, B., Jung, M., Denzler, J., Carvalhais, N., & Prabhat. (2019). Deep learning and process understanding for data-driven Earth system science. *Nature*, *566*(7743), 195–204.

Robertson, C., Long, J. A., Nathoo, F. S., Nelson, T. A., & Plouffe, C. C. F. (2014). Assessing Quality of Spatial Models Using the Structural Similarity Index and Posterior Predictive Checks. *Geographical Analysis*, *46*, 53–74.

Robertson, C., & Feick, R. (2017). Defining Local Experts : Geographical Expertise as a Basis for Geographic Information Quality. In *13th International Conference on Spatial Information Theory (COSIT 2017)*. Schloss Dagstuhl-Leibniz-Zentrum fuer Informatik.

Ronneberger, O., Fischer, P., & Brox, T. (2015). U-Net: Convolutional Networks for Biomedical Image Segmentation. *In International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, Cham.*, *9351*, 234–241.

Sampat, M. P., Wang, Z., Gupta, S., Bovik, A. C., Markey, M. K., & Member, S. (2009). Complex Wavelet Structural Similarity : A New Image Similarity Index. *IEEE Transactions on Image Processing*, *18*(11), 2385–2401.

Srinivas, S., & Fleuret, F. (2019). Full-Gradient Representation for Neural Network Visualization. *33rd Conference on Neural Information Processing Systems (NeurIPS 2019), Vancouver, Canada.*, 1–10.

Townshend, J. R. G., Gurney, C., McManus, J., & Justice, C. O. (1992). The Impact of Misregistration on Change Detection. *IEEE Transactions on Geoscience and Remote Sensing*, *30*(5), 1054–1060.

Wiederholt, R., Paudel, R., Khare, Y., Davis, S. E., Melodie Naja, G., Romañach, S., Pearlstine,

L., & Van Lent, T. (2019). A multi-indicator spatial similarity approach for evaluating ecological restoration scenarios. *Landscape Ecology*, *34*(11), 2557–2574.

Wu, J. (2004). Effects of changing scale on landscape pattern analysis: Scaling relations. *Landscape Ecology*, *19*(2), 125–138.

Xia, G., Hu, J., Hu, F., & Shi, B. (2016). AID : A Benchmark Dataset for Performance Evaluation of Aerial Scene Classification. *IEEE Transactions on Geoscience and Remote Sensing*, *55*(7), 3965–3981.

Yang, C., Gao, W., & Po, L. (2008). Discrete Wavelet Transform-Based Structural Similarity. *In Image Processing, 2008. ICIP 2008. 15th IEEE International Conference on 2008 Oct 12*, 377–380.

Zhang, L., Guo, Z., Wang, Z., & Zhang, D. (2007). Palmprint verification using complex wavelet transform. *IEEE Transactions on Signal Processing*, 2007–2010.

Zhao, W., & Du, S. (2016). Learning multiscale and deep representations for classifying remotely sensed imagery. *ISPRS Journal of Photogrammetry and Remote Sensing*, *113*, 155–165.

Zhu, X. X., Tuia, D., Mou, L., Xia, G. S., Zhang, L., Xu, F., & Fraundorfer, F. (2017). Deep Learning in Remote Sensing: A Comprehensive Review and List of Resources. *IEEE Geoscience and Remote Sensing Magazine*, *5*(4), 8–36.

Chapter 7: References

Abaidoo, C. A., Osei Jnr, E. M., Arko-Adjei, A., & Prah, B. E. K. (2019). Monitoring the Extent of Reclamation of Small Scale Mining Areas Using Artificial Neural Networks. *Heliyon*, *5*(4), e01445.

Ahmed, K. T., Ummesafi, S., & Iqbal, A. (2019). Content based image retrieval using image features information fusion. *Information Fusion*, *51*, 76–99.

Albert, A., Kaur, J., & Gonzalez, M. C. (2017, August). Using convolutional networks and satellite imagery to identify patterns in urban environments at a large scale. In *Proceedings of the 23rd ACM SIGKDD international conference on knowledge discovery and data mining*, 1357–1366.

Almeida-filho, R., & Shimabukuro, Y. E. (2000). Detecting areas disturbed by gold mining activities through JERS-1 SAR images, Roraima State, Brazilian Amazon. *International Journal of Remote Sensing*, *21*(17), 3357–3362.

Altaf, F., Islam, S. M. S., Akhtar, N., & Janjua, N. K. (2019). Going deep in medical image analysis: Concepts, methods, challenges, and future directions. *IEEE Access*, *7*, 99540–99572.

Amirshahi, S. A., Pedersen, M., & Yu, S. X. (2017). Image quality assessment by comparing CNN features between images. *IS and T International Symposium on Electronic Imaging Science and Technology*, 42–51.

Andrearczyk, V., & Whelan, P. F. (2016). Using filter banks in Convolutional Neural Networks for texture classification. *Pattern Recognition Letters*, *84*, 63–69.

Anwer, R. M., Khan, F. S., van de Weijer, J., Molinier, M., & Laaksonen, J. (2018). Binary patterns encoded convolutional neural networks for texture recognition and remote sensing

scene classification. *ISPRS Journal of Photogrammetry and Remote Sensing*, *138*, 74–85.

Asner, G. P., Llactayo, W., Tupayachi, R., & Luna, E. R. (2013). Elevated rates of gold mining in the Amazon revealed through high-resolution monitoring. *Proceedings of the National Academy of Sciences of the United States of America*, *110*(46), 18454–18459.

Atkinson, P. M., & Tate, N. J. (2000). Spatial Scale Problems and Geostatistical Solutions: A Review. *Professional Geographer*, *52*(4), 607–623.

Bajocco, S., Salvati, L., & Ricotta, C. (2011). Land degradation versus fire: A spiral process? *Progress in Physical Geography*, *35*(1), 3–18.

Basaeed, E., Bhaskar, H., Hill, P., Al-Mualla, M., & Bull, D. (2016). A supervised hierarchical segmentation of remote-sensing images using a committee of multi-scale convolutional neural networks. *International Journal of Remote Sensing*, *37*(7), 1671–1691.

Basheer, I. A., & Hajmeer, M. (2000). Artificial neural networks: Fundamentals, computing, design, and application. *Journal of Microbiological Methods*, *43*(1), 3–31.

Basu, S., Ganguly, S., Mukhopadhyay, S., DiBiano, R., Karki, M., & Nemani, R. (2015). Deepsat: a learning framework for satellite imagery. In *Proceedings of the 23rd SIGSPATIAL international conference on advances in geographic information systems*, 1-10

Basu, S., Mukhopadhyay, S., Karki, M., DiBiano, R., Ganguly, S., Nemani, R., & Gayaka, S. (2018). Deep neural networks for texture classification—A theoretical analysis. *Neural Networks*, *97*, 173–182.

Benkendorf, D. J., & Hawkins, C. P. (2020). Effects of sample size and network depth on a deep learning approach to species distribution modeling. *Ecological Informatics*, *60*, 101137.

Bernal, J., Kushibar, K., Asfaw, D. S., Valverde, S., Oliver, A., Martí, R., & Llado, X. (2019).

Deep convolutional neural networks for brain image analysis on magnetic resonance imaging: a review. *Artificial Intelligence in Medicine*, *95*, 64–81.

Betts, M. G., Diamond, A. W., Forbes, G. J., Villard, M. A., & Gunn, J. S. (2006). The importance of spatial autocorrelation, extent and resolution in predicting forest bird occurrence. *Ecological Modelling*, *191*(2), 197–224.

Blaschke, T. (2010). Object based image analysis for remote sensing. *ISPRS Journal of Photogrammetry and Remote Sensing*, *65*(1), 2–16.

Bolliger, J., Wagner, H. H., & Turner, M. G. (2007). *Identifying and Quantifying Landscape Patterns in Space and Time*. 177–194.

Boots, B., & Csillag, F. (2006). Categorical maps, comparisons, and confidence. *Journal of Geographical Systems*, *8*(2), 109–118.

Boulent, J., Foucher, S., Théau, J., & St-Charles, P. L. (2019). Convolutional Neural Networks for the Automatic Identification of Plant Diseases. *Frontiers in Plant Science*, *10*.

Bowyer, K., Kranenburg, C., & Dougherty, S. (2001). Edge detector evaluation using empirical ROC curves. *Computer Vision and Image Understanding*, *84*(1), 77-103.

Boyd, D. S., & Danson, F. M. (2005). Satellite remote sensing of forest resources : three decades of research. *Progress in Physical Geography*, *1*, 1–26.

Breiman, L. (2001). Random forests. *Machine Learning*, 5–32.

Brodrick, P. G., Davies, A. B., & Asner, G. P. (2019). Uncovering Ecological Patterns with Convolutional Neural Networks. *Trends in Ecology and Evolution*, *34*(8), 734–745.

Bruzzone, L., & Prieto, D. F. (2000). Automatic Analysis of the Difference Image for Unsupervised Change Detection. *IEEE Transactions on Geoscience and Remote Sensing*, *38*(3), 1171–1182.

Buda, M., Maki, A., & Mazurowski, M. A. (2018). A systematic study of the class imbalance problem in convolutional neural networks. *Neural Networks*, *106*, 249–259.

Buschbacher, K., Ahrens, D., Espeland, M., & Steinhage, V. (2020). Image-based species identification of wild bees using convolutional neural networks. *Ecological Informatics*, *55*, 101017.

Buscombe, D., & Ritchie, A. C. (2018). Landscape classification with deep neural networks. *Geosciences (Switzerland)*, *8*(7), 1–23.

Camps-valls, G., Gomez-chova, L., Munoz-mari, J., Rojo-alvarez, J. L., & Martinez-ramon, M. (2008). Kernel-Based Framework for Multitemporal and Multisource Remote Sensing Data Classification and Change Detection. *IEE Transactions on Geoscience and Remote Sensing*, *46*(6), 1822–1835.

Cánovas-Garcia, F., Alonso-Sarria, F., Gomariz-Castillo, F., & Onate-Valdivieso, F. (2017). Modification of the random forest algorithm to avoid statistical dependence problems when classifying remote sensing imagery. *Computers and Geosciences*, *103*, 1–11.

Cao, C., Dragicevic, S., & Li, S. (2019). Land-Use Change Detection with Convolutional Neural Network Methods. *Environments*, *6*(2), 25.

Cao, J., Liu, L., Wang, P., Huang, Z., Shen, C., & Shen, H. T. (2016). *Where to Focus: Query Adaptive Matching for Instance Retrieval Using Convolutional Feature Maps*. 1–10.

Carl, G., & Kühn, I. (2007). Analyzing spatial autocorrelation in species distributions using Gaussian and logit models. *Ecological Modelling*, *207*(2–4), 159–170.

Carleer, A. P., Debeir, O., & Wolff, E. (2005). Assessment of very high spatial resolution satellite image segmentations. *Photogrammetric Engineering and Remote Sensing*, *71*(11), 1285–1294.

Carpenter, G. A., Gopal, S., Shock, B. M., & Woodcock, C. E. (2001). A Neural Network Method for Land Use Change Classification, with Application to the Nile River Delta. *BU/CNS Technical Report TR-2001-010*.

Carranza, M. L., Hoyos, L., Frate, L., Acosta, A. T. R., & Cabido, M. (2015). Measuring forest fragmentation using multitemporal forest cover maps: Forest loss and spatial pattern analysis in the Gran Chaco, central Argentina. *Landscape and Urban Planning*, *143*, 238–247.

Chen, G., Hay, G. J., Carvalho, L. M. T., & Wulder, M. A. (2012). Object-based change detection. *International Journal of Remote Sensing*, *33*(14), 4434–4457.

Chen, J., Zhou, Z., Pan, Z., & Yang, C. (2019). Instance Retrieval Using Region of Interest Based CNN Features. *Journal of New Media*, *1*(2), 87–99.

Chen, L. C., Papandreou, G., Schroff, F., & Adam, H. (2017). *Rethinking Atrous Convolution for Semantic Image Segmentation*.

Chen, S., & Rao, P. (2008). Land degradation monitoring using multi-temporal Landsat TM/ETM data in a transition zone between grassland and cropland of northeast China. *International Journal of Remote Sensing*, *29*(7), 2055–2073.

Chen, W., Li, X., & Wang, L. (2020). Fine land cover classification in an open pit mining area using optimized support vector machine and world view-3 imagery. *Remote Sensing*, *12*(1), 12–14.

Chen, Y., Wang, K., Liao, X., Qian, Y., Wang, Q., Yuan, Z., & Heng, P. A. (2019). Channel-Unet: A Spatial Channel-Wise Convolutional Neural Network for Liver and Tumors Segmentation. *Frontiers in Genetics*, *10*(November), 1–13.

Chen, Y., Jiang, H., Li, C., Jia, X., & Ghamisi, P. (2016). Deep Feature Extraction and

Classification of Hyperspectral Images Based on Convolutional Neural Networks. *IEEE Transactions on Geoscience and Remote Sensing*, *54*(10), 6232–6251.

Cimpoi, M., Maji, S., Kokkinos, I., & Vedaldi, A. (2016). Deep Filter Banks for Texture Recognition, Description, and Segmentation. *International Journal of Computer Vision*, *118*(1), 65–94.

Comber, A., & Wulder, M. (2019). Considering spatiotemporal processes in big data analysis: Insights from remote sensing of land cover and land use. *Transactions in GIS*, 879–891.

Conglaton, R. G. (1991). A Review of Assessing the Accuarcy of Classifications of Remotely Sensed Data. *Remote Sensing of Environment*, *37*(1), 35–46.

Coops, N. C., & Wulder, M. A. (2019). Breaking the Habit(at). *Trends in Ecology & Evolution*, *34*(7), 585–587.

Coppin, P., Jonckheere, I., Nackaerts, K., Muys, B., & Lambin, E. (2004). Digital change detection methods in ecosystem monitoring: a review. *International Journal of Remote Sensing*, *25*(9), 1565–1596.

Crum, W. R., Camara, O., & Hill, D. L. G. (2006). Generalized overlap measures for evaluation and validation in medical image analysis. *IEEE Transactions on Medical Imaging*, *25*(11), 1451–1461.

Csillag, Ferko, & Boots, B. (2005). Toward Comparing Maps as Spatial Processes. *In Developments in Spatial Data Handling*, 641–652.

Csurka, G., Larlus, D., & Perronnin, F. (2013). What is a good evaluation measure for semantic segmentation? *BMVC 2013 - Electronic Proceedings of the British Machine Vision Conference 2013*.

Cui, B., Chen, X., & Lu, Y. (2020). Semantic Segmentation of Remote Sensing Images Using

Transfer Learning and Deep Convolutional Neural Network with Dense Connection. *IEEE Access*, *8*, 116744–116755.

Dabiri, Z., & Blaschke, T. (2019). Scale matters: a survey of the concepts of scale used in spatial disciplines. *European Journal of Remote Sensing*, *52*(1), 419–435.

Dai, X., & Khorram, S. (1999). Remotely sensed change detection based on artificial neural networks. *Photogrammetric Engineering & Remote Sensing*, *65*(10), 1187–1194.

Dandois, J. P., & Ellis, E. C. (2010). Remote sensing of vegetation structure using computer vision. *Remote Sensing*, *2*(4), 1157–1176.

Dandois, J. P., & Ellis, E. C. (2013). High spatial resolution three-dimensional mapping of vegetation spectral dynamics using computer vision. *Remote Sensing of Environment*, *136*, 259–276.

Dark, S. J., & Bram, D. (2007). The modifiable areal unit problem (MAUP) in physical geography. *Progress in Physical Geography*, *31*(5), 471–479.

DeFries, R. S., Hansen, M. C., Townshend, J. R. G., Janetos, A. C., & Loveland, T. R. (2000). A new global 1-km dataset of percentage tree cover derived from remote sensing. *Global Change Biology*, *6*(2), 247–254.

Delmelle, E. C. (2019). Toward a More Socially Impactful Geographical Analysis. *Geographical Analysis*, 1–9.

Déniz, O., Bueno, G., Salido, J., & De La Torre, F. (2011). Face recognition using Histograms of Oriented Gradients. *Pattern Recognition Letters*, *32*(12), 1598–1603.

Dice, L. R. . (1945). Measures of the Amount of Ecologic Association Between Species. *Ecology*, *26*(3), 297–302.

Dlamini, L. Z. D., & Xulu, S. (2019). Monitoring mining disturbance and restoration over RBM

site in South Africa using landtrendr algorithm and landsat data. *Sustainability (Switzerland)*, *11*(24).

Dosselmann, R., & Xue, D. Y. (2005). Existing and emerging image quality metrics. *Canadian Conference on Electrical and Computer Engineering*, *2005*, 1906–1913.

Dosselmann, R., & Yang, X. D. (2011). A comprehensive assessment of the structural similarity index. *Signal, Image and Video Processing*, *5*(1), 81–91.

Dosselmann, R., & Yang, X. D. (2013). A rank-order comparison of image quality metrics. *Canadian Conference on Electrical and Computer Engineering*, 1–4.

Dumic, E., Grgic, S., & Grgic, M. (2014). IQM2: new image quality measure based on steerable pyramid wavelet transform and structural similarity index. *Signal, Image and Video Processing*, *8*, 1159–1168.

Dungan, J. L., Perry, J. N., Dale, M. R. T., Legendre, P., Citron-Pousty, S., Fortin, M. J., Jakomulska, A., Miriti, M., & Rosenberg, M. S. (2002). A balanced view of scale in spatial statistical analysis. *Ecography*, *25*(5), 626–640.

Dungan, Jennifer L. (2006). Focusing on feature-based differences in map comparison. *Journal of Geographical Systems*, *8*(2), 131–143.

Duro, D. C., Coops, N. C., Wulder, M. A., & Han, T. (2007). Development of a large area biodiversity monitoring system driven by remote sensing. *Progress in Physical Geography*, *31*(3), 235–260.

Dwyer, J. L., Roy, D. P., Sauer, B., Jenkerson, C. B., Zhang, H. K., & Lymburner, L. (2018). Analysis ready data: Enabling analysis of the landsat archive. *Remote Sensing*, *10*(9), 1–19.

Egidarev, E. G., & Simonov, E. A. (2015). Assessment of the environmental effect of placer gold mining in the Amur river basin. *Water Resources*, *42*(7), 897–908.

El Amin, A. M., Liu, Q., & Wang, Y. (2016). Convolutional neural network features based change detection in satellite images. *First International Workshop on Pattern Recognition. International Society for Optics and Photonics*, *10011*, 100110W.

Elmes, A., Alemohammad, H., Avery, R., Caylor, K., Eastman, J. R., Fishgold, L., Friedl, M. A., Jain, M., Kohli, D., Bayas, J. C. L., Lunga, D., McCarty, J. L., Pontius, R. G., Reinmann, A. B., Rogan, J., Song, L., Stoynova, H., Ye, S., Yi, Z. F., & Estes, L. (2020). Accounting for training data error in machine learning applied to earth observations. *Remote Sensing*, *12*(6), 1–39.

Espejo, J. C., Messinger, M., Roman-Dañobeytia, F., Ascorra, C., Fernandez, L. E., & Silman, M. (2018). Deforestation and forest degradation due to gold mining in the Peruvian Amazon: A 34-year perspective. *Remote Sensing*, *10*(12), 1–17.

Falkowski, M. J., Wulder, M. A., White, J. C., & Gillis, M. D. (2009). Supporting large-area, sample-based forest inventories with very high spatial resolution satellite imagery. *Progress in Physical Geography*, *33*(3), 403–423.

Feng, W., Sui, H., & Tu, J. (2016). Object-Oriented Change Detection for Remote Sensing Images Based on Multi-Scale Fusion. *ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, *XLI-B7*, 483–491.

Fischer, M. M. (2001). Computational Neural Networks - Tools for Spatial Data Analysis. *GeoComputational Modelling*, 15–34.

Flood, N., Watson, F., & Collett, L. (2019). Using a U-net convolutional neural network to map woody vegetation extent from high resolution satellite imagery across Queensland, Australia. *International Journal of Applied Earth Observation and Geoinformation*, *82*, 101897.

Flores, C. F., Gonzalez-Garcia, A., van de Weijer, J., & Raducanu, B. (2019). Saliency for fine-grained object recognition in domains with scarce training data. *Pattern Recognition*, *94*, 62–73.

Foody, G. M. (2004). Thematic Map Comparison : Evaluating the Statistical Significance of Differences in Classification Accuracy. *Photogrammetric Engineering & Remote Sensing*, *70*(5), 627–633.

Foody, G. M. (2007). Map comparison in GIS. *Progress in Physical Geography*, *31*(4), 439–445.

Franklin J. (1995). Predictive Vegetation Mappimg: geographic modelling of biospatial patterns in relation to environmental gradients. *Progress in Physical Geography*, *19*(4), 474–499.

Frate, L., Carranza, M. L., & Bit, T. (2013). Quantifying Landscape-Scale Patterns of Temperate Forests over Time by Means of Neutral Simulation Models. *ISPRS International Journal of Geo-Information*, *2*, 94–109.

Frate, L., Saura, S., Minotti, M., di Martino, P., Giancola, C., & Carranza, M. L. (2014). Quantifying forest spatial pattern trends at multiple extents: An approach to detect significant changes at different scales. *Remote Sensing*, *6*(10), 9298–9315.

Frazier, A. E. (2019). Emerging trajectories for spatial pattern analysis in landscape ecology. *Landscape Ecology*, *34*(9), 2073–2082.

Frey, K. E., & Smith, L. C. (2007). How well do we know northern land cover? Comparison of four global vegetation and wetland products with a new ground-truth database for West Siberia. *Global Biogeochemical Cycles*, *21*(1).

Fu, B., Wang, Y., Campbell, A., Li, Y., Zhang, B., Yin, S., Xing, Z., & Jin, X. (2017a). Comparison of object-based and pixel-based Random Forest algorithm for wetland vegetation mapping using high spatial resolution GF-1 and SAR data. *Ecological*

Indicators, *73*, 105–117.

Fu, G., Liu, C., Zhou, R., Sun, T., & Zhang, Q. (2017b). Classification for high resolution remote sensing imagery using a fully convolutional network. *Remote Sensing*, *9*(5), 1–21.

Gadosey, P. K., Li, Y., Agyekum, E. A., Zhang, T., Liu, Z., Yamak, P. T., & Essaf, F. (2020). SD-UNET: Stripping down U-net for segmentation of biomedical images on platforms with low computational budgets. *Diagnostics*, *10*(2),110.

Gahegan, M. (2020). Fourth paradigm GIScience? Prospects for automated discovery and explanation from data. *International Journal of Geographical Information Science*, *34*(1), 1–21.

Gallwey, J., Robiati, C., Coggan, J., Vogt, D., & Eyre, M. (2020). A Sentinel-2 based multispectral convolutional neural network for detecting artisanal small-scale mining in Ghana: Applying deep learning to shallow mining. *Remote Sensing of Environment*, *248*, 111970.

Gao, Q., Lim, S., & Jia, X. (2018). Hyperspectral image classification using convolutional neural networks and multiple feature learning. *Remote Sensing*, *10*(2).

Gao, Y., Rehman, A., & Wang, Z. (2011). CW-SSIM Based Image Classification. *IEEE International Conference on Image Processing*, 1249–1252.

Garry D., P. (2000). Scaling and ecological dynamics: Self-organizaton, hierachical structure, and ecological resilience. *Climate Change*, *44*, 291–309.

Gartner, G., Meng, L., & Peterson, M. P. (2008). Object-Based Image Analysis. https://doi.org/10.1007/978-3-540-77058-9

Gatys, L. A., Ecker, A. S., & Bethge, M. (2017). Texture and art with deep neural networks. *Current Opinion in Neurobiology*, *46*, 178–186.

Ghorbanzadeh, O., Blaschke, T., Gholamnia, K., Meena, S. R., Tiede, D., & Aryal, J. (2019). Evaluation of different machine learning methods and deep-learning convolutional neural networks for landslide detection. *Remote Sensing*, *11*(2).

Ghoshal, S., James, L. A., Singer, M. B., & Aalto, R. (2010). Channel and floodplain change analysis over a 100-year period: Lower Yuba river, California. *Remote Sensing*, *2*(7), 1797–1825.

Gillanders, S. N., Coops, N. C., Wulder, M. A., Gergel, S. E., & Nelson, T. (2008). Multitemporal remote sensing of landscape dynamics and pattern change: Describing natural and anthropogenic trends. *Progress in Physical Geography*, *32*(5), 503–528.

Gilvear, D. J., Waters, T. M., & Milner, A. M. (1995). Image analysis of aerial photography to quantify changes in channel morphology and instream habitat following placer mining in interior Alaska. *Freshwater Biology*, *34*(2), 389–398.

Girdhar, R., & Ramanan, D. (2017). Attentional pooling for action recognition. *Advances in Neural Information Processing Systems*, *2017*, 34–45.

Gong, X., Xie, Z., Liu, Y., Shi, X., & Zheng, Z. (2018). Deep salient feature based anti-noise transfer network for scene classification of remote sensing imagery. *Remote Sensing*, *10*(3).

Gong, Y., Wang, L., Guo, R., & Lazebnik, S. (2014). Multi-Scale Orderless Pooling of Deep Convolutional Activation Features. *In European Conference on Computer Vision*, 392–407.

Gopal, S. (2016). Artificial Neural Networks in Geospatial Analysis. *International Encyclopedia of Geography: People, the Earth, Environment and Technology*, 1–7.

Grinblat, G. L., Uzal, L. C., Larese, M. G., & Granitto, P. M. (2016). Deep learning for plant identification using vein morphological patterns. *Computers and Electronics in Agriculture*, *127*, 418–424.

Groth, E. J. (1986). A pattern-matching algorithm for two-dimensional coordinate lists. *The Astronomical Journal*, *91*(5), 1244–1248.

Gu, Y., Wang, Y., & Li, Y. (2019). A survey on deep learning-driven remote sensing image scene understanding: Scene classification, scene retrieval and scene-guided object detection. *Applied Sciences (Switzerland)*, *9*(10), 2110.

Guo, Y., Ji, J., Lu, X., Huo, H., Fang, T., & Li, D. (2019). Global-Local Attention Network for Aerial Scene Classification. *IEEE Access*, *7*, 67200–67212.

Gustafson, E. J. (1998). Quantifying Landscape Spatial Pattern : What Is the State of the Art ? *Ecosystems*, *1*(2), 143–156.

Hagen-Zanker, A., Straatman, B., & Uljee, I. (2005). Further developments of a fuzzy set map comparison approach. *International Journal of Geographical Information Science*, *19*(7), 769–785.

Hagen, A. (2002a). *Approaching human judgement in the automated comparison of categorical maps*. 1–6.

Hagen, A. (2002b). Multi-method assessment of map similarity. *5th AGILE Conference on Geographic Information Science*, 1–8.

Hamilton, S. E., & Friess, D. A. (2018). Global carbon stocks and potential emissions due to mangrove deforestation from 2000 to 2012. *Nature Climate Change*, *8*(3), 240–244.

Han, Z., Dian, Y., Xia, H., Zhou, J., Jian, Y., Yao, C., Wang, X., & Li, Y. (2020). Comparing fully deep convolutional neural networks for land cover classification with high-spatial-resolution gaofen-2 images. *ISPRS International Journal of Geo-Information*, *9*(8).

Hansen, M. C., & Loveland, T. R. (2012). A review of large area monitoring of land cover change using Landsat data. *Remote Sensing of Environment*, *122*, 66–74.

Hay, G. J., Dubé, P., Bouchard, A., & Marceau, D. J. (2002). A scale-space primer for exploring and quantifying complex landscapes. *Ecological Modelling*, *153*(1–2), 27–49.

Hay, G. J., Blaschke, T., Marceau, D. J., & Bouchard, A. (2003). A comparison of three image-object methods for the multiscale analysis of landscape structure. *ISPRS Journal of Photogrammetry and Remote Sensing*, *57*(5–6), 327–345.

Hay, G. J., Marceau, D. J., Dube, P., & Bourchad, A. (2001). A Multiscale Framework for Landscape Analysis : Object-specific analysis and upscaling. *Landscape Ecology*, *16*(514), 1–49.

He, H., & Garcia, E. A. (2009). Learning from imbalanced data. *Studies in Computational Intelligence*, *21*(9), 1263–1284.

He, H. S. (2008). Forest landscape models: Definitions, characterization, and classification. *Forest Ecology and Management*, *254*(3), 484–498. 2

Hengl, T. (2006). Finding the right pixel size. *Computers and Geosciences*, *32*(9), 1283–1298.

Hernandez, A. J., & Ramsey, R. D. (2013). A landscape similarity index: Multitemporal remote sensing to track changes in big sagebrush ecological sites. *Rangeland Ecology and Management*, *66*(1), 71–81.

Hinton, G. E., Srivastava, N., Krizhevsky, A., Sutskever, I., & Salakhutdinov, R. R. (2012). *Improving neural networks by preventing co-adaptation of feature detectors*. 1–18.

Huang, H., & Xu, K. (2019). Combing triple-part features of convolutional neural networks for scene classification in remote sensing. *Remote Sensing*, *11*(14), 1687.

Hurcom, S. J., & Harrison, A. R. (1998). The NDVI and spectral decomposition for semi-arid vegetation abundance estimation. *International Journal of Remote Sensing*, *19*(16), 3109–3125.

Hussain, M., Chen, D., Cheng, A., Wei, H., & Stanley, D. (2013). Change detection from remotely sensed images: From pixel-based to object-based approaches. *ISPRS Journal of Photogrammetry and Remote Sensing*, *80*, 91–106.

Ibrahim, E., Lema, L., Barnabé, P., Lacroix, P., & Pirard, E. (2020). Small-scale surface mining of gold placers: Detection, mapping, and temporal analysis through the use of free satellite imagery. *International Journal of Applied Earth Observation and Geoinformation*, *93*(August), 102194.

Im, J., Jensen, J. R., & Tullis, J. A. (2008). Object-based change detection using correlation image analysis and image segmentation. *International Journal of Remote Sensing*, *29*(2), 399–423.

Ioannidou, S., & Karathanassi, V. (2007). Investigation of the Dual-Tree Complex and Shift-Invariant Discrete Wavelet Transforms on Quickbird Image Fusion. *IEEE Geoscience and Remote Sensing Letters*, *4*(1), 166–170.

Ioffe, S., & Szegedy, C. (2015). Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift. *In: International Conference on Machine Learning, Lille, France.*, 448–456.

Izquierdo-Verdiguier, E., & Zurita-Milla, R. (2020). An evaluation of Guided Regularized Random Forest for classification and regression tasks in remote sensing. *International Journal of Applied Earth Observation and Geoinformation*, *88*, 102051.

Jain, A. K., & Mao, J. (1996). Artificial Neural Network: A Tutorial. *Communications*, *29*, 31–44.

Janowicz, K., Gao, S., McKenzie, G., Hu, Y., & Bhaduri, B. (2020). GeoAI: spatially explicit artificial intelligence techniques for geographic knowledge discovery and beyond.

International Journal of Geographical Information Science, *34*(4), 625–636.

Jasiewicz, J., Netzel, P., & Stepinski, T. F. (2014). Landscape similarity, retrieval, and machine mapping of physiographic units. *Geomorphology*, *221*, 104–112.

Jelinski, D. E., & Wu, J. (1996). The modifiable areal unit problem and implications for landscape ecology. *Landscape Ecology*, *11*(3), 129–140.

Jenerette, G. D., & Potere, D. (2010). Global analysis and simulation of land-use change associated with urbanization. *Landscape Ecology*, *25*(5), 657–670.

Jiang, X., Pang, Y., Li, X., Pan, J., & Xie, Y. (2018). Deep neural networks with Elastic Rectified Linear Units for object recognition. *Neurocomputing*, *275*, 1132–1139.

Jones, E. L., Rendell, L., Pirotta, E., & Long, J. A. (2016). Novel application of a quantitative spatial comparison tool to species distribution data. *Ecological Indicators*, *70*, 67–76.

Joseph, L. N., & Possingham, H. P. (2008). Grid-based monitoring methods for detecting population declines: Sensitivity to spatial scale and consequences of scale correction. *Biological Conservation*, *141*(7), 1868–1875.

Kalantar, B., Ueda, N., Al-Najjar, H. A. H., & Halin, A. A. (2020). Assessment of convolutional neural network architectures for earthquake-induced building damage detection based on pre-and post-event orthophoto images. *Remote Sensing*, *12*(21), 1–20.

Kang, M., Ji, K., Leng, X., & Lin, Z. (2017). Contextual region-based convolutional neural network with multilayer fusion for SAR ship detection. *Remote Sensing*, *9*(8).

Keane, R. E., Cary, G. J., Davies, I. D., Flannigan, M. D., Gardner, R. H., Lavorel, S., Lenihan, J. M., Li, C., & Rupp, T. S. (2004). A classification of landscape fire succession models: Spatial simulations of fire and vegetation dynamics. *Ecological Modelling*, *179*(1–2), 3–27.

Kennedy, R. E., Townsend, P. A., Gross, J. E., Cohen, W. B., Bolstad, P., Wang, Y. Q., &

Adams, P. (2009). Remote sensing change detection tools for natural resource managers: Understanding concepts and tradeoffs in the design of landscape monitoring projects. *Remote Sensing of Environment*, *113*(7), 1382–1396.

Khan, S.H., He, X., Porikli, F., & Bennamoun, M. (2017). Forest Change Detection in Incomplete Satellite Images with Deep Neural Networks. *IEEE Transactions on Geoscience and Remote Sensing*, *55*(9), 5407–5423.

Khan, Salman H., Hayat, M., Bennamoun, M., Sohel, F. A., & Togneri, R. (2018). Cost-sensitive learning of deep feature representations from imbalanced data. *IEEE Transactions on Neural Networks and Learning Systems*, *29*(8), 3573–3587.

Kingma, D. P., & Ba, J. L. (2015). Adam: A method for stochastic optimization. *3rd International Conference on Learning Representations, ICLR 2015 - Conference Track Proceedings*, 1–15.

Klapka, P., Halas, M., Netrdova, P., & Nosek, V. (2016). The efficiency of areal units in spatial analysis: Assessing the performance of functional and administrative regions. *Moravian Geographical Reports*, *24*(2), 47–59.

Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). ImageNet Classification with Deep Convolutional Neural Networks. *Advances In Neural Information Processing Systems*, 1–9.

Kroese, D. P., & Botev, Z. I. (2015). Spatial Process Simulation. In *Stochastic geometry, spatial statistics and Random Fields. Lecture Notes in Mathematics, vol 2120. Springer, Cham*, 369–404.

Kruger, N., Janssen, P., Kalkan, S., Lappe, M., Leonardis, A., Piater, J., Rodriguez-Sanchez, A. J., & Wiskott, L. (2013). Deep hierarchies in the primate visual cortex: What can we learn for computer vision? *IEEE Transactions on Pattern Analysis and Machine Intelligence*,

35(8), 1847–1871.

Kulha, N., Pasanen, L., Holmstrom, L., De Grandpré, L., Gauthier, S., Kuuluvainen, T., & Aakala, T. (2020). The structure of boreal old-growth forests changes at multiple spatial scales over decades. *Landscape Ecology*, *35*(4), 843–858.

Kulha, N., Pasanen, L., Holmström, L., De Grandpré, L., Kuuluvainen, T., & Aakala, T. (2019). At What Scales and Why Does Forest Structure Vary in Naturally Dynamic Boreal Forests? An Analysis of Forest Landscapes on Two Continents. *Ecosystems*, *22*(4), 709–724.

Law, S., Seresinhe, C. I., Shen, Y., & Gutierrez-Roig, M. (2020). Street-Frontage-Net: urban image classification using deep convolutional neural networks. *International Journal of Geographical Information Science*, *34*(4), 681–707.

Lecun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, *521*(7553), 436–444.

Lettry, L., Perdoch, M., Vanhoey, K., & Van Gool, L. (2017). Repeated pattern detection using CNN activations. *Proceedings - 2017 IEEE Winter Conference on Applications of Computer Vision, WACV 2017*, 47–55.

Levin, S. A. (1992). The Problem of Pattern and Scale in Ecology. *Ecology*, *73*(6), 1943–1967.

Li, Z. G., Xie, S. L., Yao, W., Rahardja, S. (2010). A Similarity Index in the Complex Wavelet Domain. *In Industrial Electronics and Applications (ICIEA) 5th IEEE Conference on 2010 Jun 15*, 1644–1649.

Li, F., & Zhang, L. (2007). Comparison of point pattern analysis methods for classifying the spatial distributions of spruce-fir stands in the north-east USA. *Forestry*, *80*(3), 337–349.

Li, H., Ellis, J. G., Zhang, L., & Chang, S. F. (2018). PatternNet: Visual pattern mining with deep neural network. *ICMR 2018 - Proceedings of the 2018 ACM International Conference on Multimedia Retrieval*, 291–299.

Li, Z., White, J. C., Wulder, M. A., Hermosilla, T., Davidson, A. M., & Comber, A. J. (2020a). Land cover harmonization using Latent Dirichlet Allocation. *International Journal of Geographical Information Science*, *00*(00), 1–27.

Li, J., Knapp, D. E., Fabina, N. S., Kennedy, E. V., Larsen, K., Lyons, M. B., Murray, N. J., Phinn, S. R., Roelfsema, C. M., & Asner, G. P. (2020b). A global coral reef probability map generated using convolutional neural networks. *Coral Reefs*, *39*(6), 1805–1815.

Li, W., & Hsu, C. Y. (2020). Automated terrain feature identification from remote sensing imagery: a deep learning approach. *International Journal of Geographical Information Science*, *34*(4), 637–660.

Lim, L. A., & Keles, H. Y. (2020). Learning multi-scale features for foreground segmentation. *Pattern Analysis and Applications*, *23*(3), 1369–1380.

Lindeberg, T. (1993). Lindeberg_Detecting salient blob-like image structures and their scales with a scale-space primal sketch-a method for focus-of-attention. *International Journal of Computer Vision*, *11*(3), 283–318.

Lindeberg, T. (1994). Scale-space theory: A basic tool for analyzing structures at different scales. *Journal of Applied Statistics*, *21*(1–2), 225–270.

Liu, L., Shen, C., & Van Den Hengel, A. (2015). The treasure beneath convolutional layers: Cross-convolutional-layer pooling for image classification. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 4749–4757.

Liu, T., Abd-elrahman, A., Morton, J., Wilhelm, V. L., Liu, T., Abd-elrahman, A., Morton, J., & Wilhelm, V. L. (2018a). Comparing fully convolutional networks , random forest , support vector machine , and patch-based deep convolutional neural networks for object- based wetland mapping using images from small unmanned aircraft system. *GIScience & Remote*

Sensing, *55*(2), 243–264.

Liu, Yanfei, Zhong, Y., & Qin, Q. (2018b). Scene classification based on multiscale convolutional neural network. *IEEE Transactions on Geoscience and Remote Sensing*, *56*(12), 7109–7121.

Liu, Z., Li, G., Mercier, G., He, Y., & Pan, Q. (2018c). Change Detection in Heterogenous Remote Sensing Images via Homogeneous Pixel Transformation. *IEEE Transactions on Image Processing*, *27*(4), 1822–1834.

Liu, Yanan, Gong, W., Hu, X., & Gong, J. (2018d). Forest type identification with random forest using Sentinel-1A, Sentinel-2A, multi-temporal Landsat-8 and DEM data. *Remote Sensing*, *10*(6), 1–25.

Liu, Yazhou, Cao, G., Sun, Q., & Siegel, M. (2015). Hyperspectral classification via deep networks and superpixel segmentation. *International Journal of Remote Sensing*, *36*(13), 3459–3482.

Liu, Yishu, Han, Z., Chen, C., DIng, L., & Liu, Y. (2020). Eagle-Eyed Multitask CNNs for Aerial Image Retrieval and Scene Classification. *IEEE Transactions on Geoscience and Remote Sensing*, *58*(9), 6699–6721.

Lobo, F. de L., Souza-Filho, P. W. M., Novo, E. M. L. de M., Carlos, F. M., & Barbosa, C. C. F. (2018). Mapping mining areas in the Brazilian amazon using MSI/Sentinel-2 imagery (2017). *Remote Sensing*, *10*(8).

Long, J., Nelson, T., & Wulder, M. (2010). Regionalization of Landscape Pattern Indices Using Multivariate Cluster Analysis. *Environmental Management*, *46*, 134–142.

Long, J., & Robertson, C. (2018). Comparing spatial patterns. *Geography Compass*, *12*(2), 1–18.

Lu, D., Mausel, P., Brondizio, E., Moran, E. (2003). Change detection techniques. *International*

Journal of Remote Sensing, *25*(12), 2365–2407.

Luders, E., Cherbuin, N., & Gaser, C. (2016). Estimating brain age using high-resolution pattern recognition: Younger brains in long-term meditation practitioners. *NeuroImage*, *134*, 508–513.

Luo, J. H., Zhang, H., Zhou, H. Y., Xie, C. W., Wu, J., & Lin, W. (2018). ThiNet: Pruning CNN Filters for a Thinner Net. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *41*(10), 2525–2538.

Madhuanand, L., Sadavarte, P., Visschedijk, A. J. H., Denier Van Der Gon, H. A. C., Aben, I., & Osei, F. B. (2021). Deep convolutional neural networks for surface coal mines determination from sentinel-2 images. *European Journal of Remote Sensing*, *54*(1), 296–309.

Mahendran, A., & Vedaldi, A. (2014). Understanding Deep Image Representations by Inverting Them. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 5188-5196.

Mairota, P., Cafarelli, B., Boccaccio, L., Leronni, V., Labadessa, R., Kosmidou, V., & Nagendra, H. (2013). Using landscape structure to develop quantitative baselines for protected area monitoring. *Ecological Indicators*, *33*, 82–95.

Makowiecki, W., & Alda, W. (2008). New sky pattern recognition algorithm. *In International Conference on Computational Science*, 749–758.

Marceau, D. J., Gratton, D. J., Fournier, R. A., & Fortin, J. P. (1994). Remote sensing and the measurement of geographical entities in a forested environment. 2. The optimal spatial resolution. *Remote Sensing of Environment*, *49*(93), 105–117.

Marceau, D. J., & Hay, G. J. (1999). Remote sensing contributions to the scale issue. *Canadian*

Journal of Remote Sensing, *25*(4), 357–366.

Marcus, W. A. (2002). Mapping of stream microhabitats with high spatial resolution hyperspectral imagery. *Journal of Geographical Systems*, *4*(1), 113–126.

Mas, J. F. (1999). Monitoring land-cover changes : A comparison of change detection techniques. *International Journal of Remote Sensing*, *20*(1), 139–152.

Matthews, N. (2016). People and Fresh Water Ecosystems: Pressures, Responses and Resilience. *Aquatic Procedia*, *6*, 99–105.

Maxwell, A. E., Pourmohammadi, P., & Poyner, J. D. (2020). Mapping the topographic features of mining-related valley fills using mask R-CNN deep learning and digital elevation data. *Remote Sensing*, *12*(3), 1–23.

McCallum, I., Obersteiner, M., Nilsson, S., & Shvidenko, A. (2006). A spatial comparison of four satellite derived 1 km global land cover datasets. *International Journal of Applied Earth Observation and Geoinformation*, *8*(4), 246–255.

McIntire, E. J. B., & Fajardo, A. (2009). Beyond description: The active and effective way to infer processes from spatial patterns. *Ecology*, *90*(1), 46–56.

McIntosh, A. R., Bookstein, F. L., Haxby, J. V., & Grady, C. L. (1996). Spatial pattern analysis of functional brain images using partial least squares. *NeuroImage*, *3*(3 I), 143–157.

Melville, B., Lucieer, A., & Aryal, J. (2018). Object-based random forest classification of Landsat ETM+ and WorldView-2 satellite imagery for mapping lowland native grassland communities in Tasmania, Australia. *International Journal of Applied Earth Observation and Geoinformation*, *66*(May 2017), 46–55.

Mhangara, P., Tsoeleng, L. T., & Mapurisa, W. (2020). Monitoring the development of artisanal mines in South Africa. *Journal of the Southern African Institute of Mining and Metallurgy*,

120(4), 299–306.

Miller, H. J., & Goodchild, M. F. (2015). Data-driven geography. *GeoJournal*, *80*(4), 449–461.

Mohammadimanesh, F., Salehi, B., Mahdianpari, M., Gill, E., & Molinier, M. (2019). A new fully convolutional neural network for semantic segmentation of polarimetric SAR imagery in complex land cover ecosystem. *ISPRS Journal of Photogrammetry and Remote Sensing*, *151*, 223–236.

Monserud, R. A., & Leemans, R. (1992). Comparing global vegetation maps with the Kappa statistic. *Ecological Modelling*, *62*(4), 275–293.

Murabito, F., Spampinato, C., Palazzo, S., Giordano, D., Pogorelov, K., & Riegler, M. (2018). Top-down saliency detection driven by visual classification. *Computer Vision and Image Understanding*, *172*, 67–76.

National Science Foundation. (2018). https://doi.org/https://www.nsf.gov/pubs/2016/nsf16595/nsf16595.pdf%0A%0A

Neil Adger, W., Arnell, N. W., & Tompkins, E. L. (2005). Successful adaptation to climate change across scales. *Global Environmental Change*, *15*(2), 77–86.

Nellis, M. D., Harrington, J. A., & Wu, J. (1998). Remote sensing of temporal and spatial variations in pool size, suspended sediment, turbidity, and Secchi depth in Tuttle Creek Reservoir, Kansas: 1993. *Geomorphology*, *21*(3–4), 281–293.

Nelson, A. D., & Church, M. (2012). Placer mining along the Fraser River, British Columbia: The geomorphic impact. *Bulletin of the Geological Society of America*, *124*(7–8), 1212–1228.

Nelson, T. A., & Robertson, C. (2012). Refining spatial neighbourhoods to capture terrain effects. *Ecological Prosesses*, *1*(1), 1–11.

Niesterowicz, J., Stepinski, T. F., & Jasiewicz, J. (2016). Unsupervised regionalization of the United States into landscape pattern types. *International Journal of Geographical Information Science*, *30*(7), 1450–1468.

Niesterowicz, J., & Stepinski, T. F. (2016). On using landscape metrics for landscape similarity search. *Ecological Indicators*, *64*, 20–30.

Nogueira, K., Penatti, O. A. B., & dos Santos, J. A. (2017). Towards better exploiting convolutional neural networks for remote sensing scene classification. *Pattern Recognition*, *61*, 539–556.

Noi Phan, T., Kuch, V., & Lehnert, L. W. (2020). Land cover classification using google earth engine and random forest classifier-the role of image composition. *Remote Sensing*, *12*(15).

Nowakowski, A., Mrziglod, J., Spiller, D., Bonifacio, R., Ferrari, I., Mathieu, P. P., Garcia-Herranz, M., & Kim, D. H. (2021). Crop type mapping by using transfer learning. *International Journal of Applied Earth Observation and Geoinformation*, *98*, 102313.

Nowosad, J., & Stepinski, T. F. (2018a). Global inventory of landscape patterns and latent variables of landscape spatial configuration. *Ecological Indicators*, *89*, 159–167.

Nowosad, J., & Stepinski, T. F. (2018b). Towards machine ecoregionalization of Earth's landmass using pattern segmentation method. *International Journal of Applied Earth Observation and Geoinformation*, *69*, 110–118.

Nowosad, J., & Stepinski, T. F. (2018c). Spatial association between regionalizations using the information-theoretical V-measure. *International Journal of Geographical Information Science*, *32*(12), 2386–2401.

Nowosad, J., & Stepinski, T. F. (2019). Information theory as a consistent framework for quantification and classification of landscape patterns. *Landscape Ecology*, *34*(9), 2091–

2101.

Nowosad, J., & Stepinski, T. F. (2021). Pattern-based identification and mapping of landscape types using multi-thematic data. *International Journal of Geographical Information Science*, *00*(00), 1–16.

Nowosad, J., Stepinski, T. F., & Netzel, P. (2019). Global assessment and mapping of changes in mesoscale landscapes: 1992–2015. *International Journal of Applied Earth Observation and Geoinformation*, *78*, 332–340.

Obodai, J., Adjei, K. A., Odai, S. N., & Lumor, M. (2019). Land use/land cover dynamics using landsat data in a gold mining basin-the Ankobra, Ghana. *Remote Sensing Applications: Society and Environment*, *13*, 247–256.

Olson, D. M., Dinerstein, E., Wikramanayake, E. D., Burgess, N. D., Powell, G. V. N., Underwood, E. C., D'amico, J. A., Itoua, I., Strand, H. E., Morrison, J. C., Loucks, C. J., Allnutt, T. F., Ricketts, T. H., Kura, Y., Lamoreux, J. F., Wettengel, W. W., Hedao, P., & Kassem, K. R. (2001). Terrestrial Ecoregions of the World: A New Map of Life on Earth. *BioScience*, *51*(11), 933.

Omeiza, D., Speakman, S., Cintas, C., & Weldermariam, K. (2019). *Smooth Grad-CAM++: An Enhanced Inference Level Visualization Technique for Deep Convolutional Neural Network Models*. 1–10.

Openshaw, S. (1984). Ecological fallacies and the analysis of areal census data. *Environment & Planning A*, *16*(1), 17–31.

Oppenheim, A. V., & Lim, J. S. (1981). The Importance of Phase in Signals. *Proceedings of the IEEE*, *69*(5), 529–541.

Oquab, M., Bottou, L., Laptev, I., & Sivic, J. (2014). Learning and Transferring Mid-Level

Image Representations using Convolutional Neural Networks. *IEEE Conference on Computer Vision and Pattern Recognition*, 1717–1724.

Papaodysseus, C., Exarhos, M., Panagopoulos, M., Rousopoulos, P., Triantafillou, C., & Panagopoulos, T. (2008). Image and pattern analysis of 1650 B.C. wall paintings and reconstruction. *IEEE Transactions on Systems, Man, and Cybernetics Part A:Systems and Humans*, *38*(4), 958–965.

Peng, C., Zhang, X., Yu, G., Luo, G., & Sun, J. (2017). Large kernel matters - Improve semantic segmentation by global convolutional network. *Proceedings - 30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017*, 1743–1751.

Peng, F., Wang, L., Zou, S., Luo, J., Gong, S., & Li, X. (2019). Content-based search of earth observation data archives using open-access multitemporal land cover and terrain products. *International Journal of Applied Earth Observation and Geoinformation*, *81*, 13–26.

Petrovska, B., Zdravevski, E., Lameski, P., Corizzo, R., Štajduhar, I., & Lerga, J. (2020). Deep learning for feature extraction in remote sensing: A case-study of aerial scene classification. *Sensors (Switzerland)*, *20*(14), 1–22.

Ping, J. L., Green, C. J., Zartman, R. E., & Bronson, K. F. (2004). Exploring spatial dependence of cotton yield using global and local autocorrelation statistics. *Field Crops Research*, *89*(2–3), 219–236.

Pontius Jr, R. G., Millones, M. (2011). Death to Kappa: birth of quantity disagreement and allocation disagreement for accuracy assessment. *International Journal of Remote Sensing*, *32*(15), 4407–4429.

Pontius Jr, R. G., & Schneider, L. C. (2001). Land-cover change model validation by an ROC method for the Ipswich watershed, Massachusetts, USA. *Agriculture, Ecosystems and*

Environment, *85*(1), 191–203.

Pontius, R. G. (2000). Quantification error versus location error in comparison of categorical maps. In *Photogrammetric Engineering and Remote Sensing*, *66*(8), 1011–1016.

Pontius, Robert Gilmore, Huffaker, D., & Denman, K. (2004). Useful techniques of validation for spatially explicit land-change models. *Ecological Modelling*, *179*(4), 445–461.

Potter, C., Tan, P. N., Steinbach, M., Klooster, S., Kumar, V., Myneni, R., & Genovese, V. (2003). Major disturbance events in terrestrial ecosystems detected using global satellite data sets. *Global Change Biology*, *9*(7), 1005–1021.

Power, C., Simms, A., & White, R. (2001). Hierarchical fuzzy pattern matching for the regional comparison of land use maps. *International Journal of Geographical Information Science*, *15*(1), 77–100.

Prieto, A., Prieto, B., Ortigosa, E. M., Ros, E., Pelayo, F., Ortega, J., & Rojas, I. (2016). Neural networks: An overview of early research, current frameworks and new challenges. *Neurocomputing*, *214*, 242–268.

Prieto, M. S., & Allen, A. R. (2003). A similarity metric for edge images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *25*(10), 1265–1273.

Qi, X., Li, C. G., Zhao, G., Hong, X., & Pietikäinen, M. (2016). Dynamic texture and scene classification by transferring deep image features. *Neurocomputing*, *171*, 1230–1241.

Quoc Bao, T., Tan Kiet, N. T., Quoc Dinh, T., & Hiep, H. X. (2020). Plant species identification from leaf patterns using histogram of oriented gradients feature space and convolution neural networks. *Journal of Information and Telecommunication*, *4*(2), 140–150.

Rehman, A., Gao, Y., Wang, J., Wang, Z. (2013). Signal Processing : Image Communication structural similarity. *Signal Processing : Image Communication*, *28*(8), 984–992.

Reichstein, M., Camps-Valls, G., Stevens, B., Jung, M., Denzler, J., Carvalhais, N., & Prabhat. (2019). Deep learning and process understanding for data-driven Earth system science. *Nature*, *566*(7743), 195–204.

Remmel, T.K, & Csillag, F. (2003). When are two landscape pattern indices significantly different ? 331–351.

Remmel, T. K. (2020). Distributions of Hyper-Local Configuration Elements to Characterize, Compare, and Assess Landscape-Level Spatial Patterns. *Entropy*, *420*(22), 1–13.

Remmel, T. K., & Fortin, M. J. (2013). Categorical, class-focused map patterns: Characterization and comparison. *Landscape Ecology*, *28*(8), 1587–1599.

Renno, C. D., Nobre, A. D., Cuartas, L. A., Soares, J. V., Hodnett, M. G., Tomasella, J., & Waterloo, M. J. (2008). HAND, a new terrain descriptor using SRTM-DEM: Mapping terra-firme rainforest environments in Amazonia. *Remote Sensing of Environment*, *112*(9), 3469–3481.

Robertson, C., & Feick, R. (2017). Defining Local Experts : Geographical Expertise as a Basis for Geographic Information Quality. In *13th International Conference on Spatial Information Theory (COSIT 2017)*. Schloss Dagstuhl-Leibniz-Zentrum fuer Informatik.

Robertson, C., Long, J. A., Nathoo, F. S., Nelson, T. A., & Plouffe, C. C. F. (2014). Assessing Quality of Spatial Models Using the Structural Similarity Index and Posterior Predictive Checks. *Geographical Analysis*, *46*, 53–74.

Robertson, C., Nelson, T. A., Boots, B., & Wulder, M. A. (2007). *STAMP : spatial – temporal analysis of moving polygons*. 207–227.

Rogan, J., & Chen, D. (2004). Remote sensing technology for mapping and monitoring land-cover and land-use change. *Progress in Planning*, *61*(4), 301–325.

Ronneberger, O., Fischer, P., & Brox, T. (2015). U-Net: Convolutional Networks for Biomedical Image Segmentation. *In International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, Cham.*, *9351*, 234–241.

Rubner, Y., Tomasi, C., & Guibas, L. J. (2000). Earth mover's distance as a metric for image retrieval. *International Journal of Computer Vision*, *40*(2), 99–121.

Rui, T., Zou, J., Zhou, Y., Fei, J., & Yang, C. (2017). Convolutional neural network feature maps selection based on LDA. *Multimedia Tools and Applications*, *1*, 1–15.

Sampat, M. P., Wang, Z., Gupta, S., Bovik, A. C., Markey, M. K., & Member, S. (2009). Complex Wavelet Structural Similarity : A New Image Similarity Index. *IEEE Transactions on Image Processing*, *18*(11), 2385–2401.

Selvaraju, R. R., Cogswell, M., Das, A., Vedantam, R., Parikh, D., & Batra, D. (2017). Grad-CAM: Visual Explanations from Deep Networks via Gradient-Based Localization. *In Proceedings of the IEEE International Conference on Computer Vision*, 618–626.

Sermanet, P., & Lecun, Y. (2011). Traffic sign recognition with multi-scale convolutional networks. *Proceedings of the International Joint Conference on Neural Networks*, *June*, 2809–2813.

Sharma, A., Liu, X., Yang, X., & Shi, D. (2017). A patch-based convolutional neural network for remote sensing image classification. *Neural Networks*, *95*, 19–28.

Shekhar, S., Evans, M. R., Kang, J. M., & Mohan, P. (2011). Identifying patterns in spatial information: A survey of methods. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, *1*(3), 193–214.

Shi, Xiaodan, Ma, G., Chen, F., & Ma, Y. (2016). A Kernel-Based Similarity Measuring for Change Detection in Remote Sensing Images. *ISPRS - International Archives of the*

Photogrammetry, Remote Sensing and Spatial Information Sciences, *XLI-B7*, 999–1006.

Shi, Xiaoxia, & Qian, X. (2019). Exploring spatial and channel contribution for object based image retrieval. *Knowledge-Based Systems*, *186*, 104955.

Shin, H. C., Roth, H. R., Gao, M., Lu, L., Xu, Z., Nogues, I., Yao, J., Mollura, D., & Summers, R. M. (2016). Deep Convolutional Neural Networks for Computer-Aided Detection: CNN Architectures, Dataset Characteristics and Transfer Learning. *IEEE Transactions on Medical Imaging*, *35*(5), 1285–1298.

Simonyan, K., Vedaldi, A., & Zisserman, A. (2013). *Deep Inside Convolutional Networks: Visualising Image Classification Models and Saliency Maps*. 1–8.

Simonyan, K., & Zisserman, A. (2015). Very Deep Convolutional Networks for Large-Scale Image Recognition. *ICLR 2015*, 1–14.

Solana-Gutierrez, J., Rincon, G., Alonso, C., & Garcia-de-Jalon, D. (2017). Using fuzzy cognitive maps for predicting river management responses: A case study of the Esla River basin, Spain. *Ecological Modelling*, *360*, 260–269.

Song, F., Yang, Z., Gao, X., Dan, T., Yang, Y., Zhao, W., & Yu, R. (2018). Multi-scale feature based land cover change detection in mountainous terrain using multi-temporal and multi-sensor remote sensing images. *IEEE Access*, *6*, 77494–77508.

Springenberg, J. T., Dosovitskiy, A., Brox, T., & Riedmiller, M. (2014). *Striving for Simplicity: The All Convolutional Net*. 1–14.

Srinivas, S., & Fleuret, F. (2019). Full-Gradient Representation for Neural Network Visualization. *33rd Conference on Neural Information Processing Systems (NeurIPS 2019), Vancouver, Canada.*, 1–10.

Srivastava, S., Vargas Munoz, J. E., Lobry, S., & Tuia, D. (2020). Fine-grained landuse

characterization using ground-based pictures: a deep learning solution based on globally available data. *International Journal of Geographical Information Science*, *34*(6), 1117–1136.

Stoian, A., Poulain, V., Inglada, J., Poughon, V., & Derksen, D. (2019). Land cover maps production with high resolution satellite image time series and convolutional neural networks: Adaptations and limits for operational systems. *Remote Sensing*, *11*(17), 1–26.

Sudre, C. H., Li, W., Vercauteren, T., Ourselin, S., & Jorge Cardoso, M. (2017). Generalised dice overlap as a deep learning loss function for highly unbalanced segmentations. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, *10553 LNCS*, 240–248.

Sun, X., Lin, X., Shen, S., & Hu, Z. (2017). High-resolution remote sensing data classification over urban areas using random forest ensemble and fully connected conditional random field. *ISPRS International Journal of Geo-Information*, *6*(8).

Sylvain, J. D., Drolet, G., & Brown, N. (2019). Mapping dead forest cover using a deep convolutional neural network and digital aerial photography. *ISPRS Journal of Photogrammetry and Remote Sensing*, *156*(August), 14–26.

Syrris, V., Hasenohr, P., Delipetrev, B., Kotsev, A., Kempeneers, P., & Soille, P. (2019). Evaluation of the Potential of Convolutional Neural Networks and Random Forests for Multi-Class Segmentation of Sentinel-2 Imagery. *Remote Sensing*, *11*(8), 907.

Szilassi, P., Bata, T., Szabo, S., Czucz, B., Molnár, Z., & Mezosi, G. (2017). The link between landscape pattern and vegetation naturalness on a regional scale. *Ecological Indicators*, *81*, 252–259.

Takala, M., Luojus, K., Pulliainen, J., Derksen, C., Lemmetyinen, J., Karna, J. P., Koskinen, J.,

& Bojkov, B. (2011). Estimating northern hemisphere snow water equivalent for climate research through assimilation of space-borne radiometer data and ground-based measurements. *Remote Sensing of Environment*, *115*(12), 3517–3529.

Tewkesbury, A. P., Comber, A. J., Tate, N. J., Lamb, A., & Fisher, P. F. (2015). A critical synthesis of remotely sensed optical image change detection techniques. *Remote Sensing of Environment*, *160*, 1–14.

Tharwat, A. (2018). Classification assessment methods. *Applied Computing and Informatics*, *17*(1), 168–192.

Tian, S., Zhang, X., Tian, J., & Sun, Q. (2016). Random forest classification of wetland landcovers from multi-sensor data in the arid region of Xinjiang, China. *Remote Sensing*, *8*(11), 1–14.

Tisseyre, B., Leroux, C., Pichon, L., Geraudie, V., & Sari, T. (2018). How to define the optimal grid size to map high resolution spatial data? *Precision Agriculture*, *19*(5), 957–971.

Townsend, P. A., Lookingbill, T. R., Kingdon, C. C., & Gardner, R. H. (2009). Spatial pattern analysis for monitoring protected areas. *Remote Sensing of Environment*, *113*(7), 1410–1420.

Townshend, J. R. G., Gurney, C., McManus, J., & Justice, C. O. (1992). The Impact of Misregistration on Change Detection. *IEEE Transactions on Geoscience and Remote Sensing*, *30*(5), 1054–1060.

Townshend, J. R., Masek, J. G., Huang, C., Vermote, E. F., Gao, F., Channan, S., Sexton, J. O., Feng, M., Narasimhan, R., Kim, D., Song, K., Song, D., Song, X. P., Noojipady, P., Tan, B., Hansen, M. C., Li, M., & Wolfe, R. E. (2012). Global characterization and monitoring of forest cover using Landsat data: Opportunities and challenges. *International Journal of*

Digital Earth, *5*(5), 373–397.

Tracewski, L., Bastin, L., & Fonte, C. C. (2017). Repurposing a deep learning network to filter and classify volunteered photographs for land cover and land use characterization. *Geo-Spatial Information Science*, *20*(3), 252–268.

Turner, Monica G., Dale, V. H., & Gardner, R. H. (1989). Predicting across scales: Theory development and testing. *Landscape Ecology*, *3*(3–4), 245–252.

Turner, Monica Goigel. (1989). Landscape Ecology : The Effect of Pattern on Process. *Annual Review of Ecology and Systematics*, *20*, 171–197.

Unar, S., Wang, X., Wang, C., & Wang, Y. (2019). A decisive content based image retrieval approach for feature fusion in visual and textual images. *Knowledge-Based Systems*, *179*, 8–20.

Urban, D. L., O'Neill, R. V, & Shugart, H. H. J. (1987). Landscape Ecology: A hierarchical perspective can help scientists understand spatial patterns. In *BioScience*, 37(2), 119–127.

Ustyuzhaninov, I., Brendel, W., Gatys, L. A., & Bethge, M. (2016). *Texture Synthesis Using Shallow Convolutional Networks with Random Filters*. 1–9.

van Vliet, J., Bregt, A. K., & Hagen-Zanker, A. (2011). Revisiting Kappa to account for change in the accuracy assessment of land-use change models. *Ecological Modelling*, *222*(8), 1367–1375.

Visser, H., & De Nijs, T. (2006). The map comparison kit. *Environmental Modelling and Software*, *21*(3), 346–358.

Volpi, M., Tuia, D., Bovolo, F., Kanevski, M., & Bruzzone, L. (2013). Supervised change detection in VHR images using contextual information and support vector machines. *International Journal of Applied Earth Observation and Geoinformation*, *20*(1), 77–85.

Vranken, I., Baudry, J., Aubinet, M., Visser, M., & Bogaert, J. (2015). A review on the use of entropy in landscape ecology: heterogeneity, unpredictability, scale dependence and their links with thermodynamics. *Landscape Ecology*, *30*(1), 51–65.

Wagner, H. H., & Fortin, M. J. (2005). Spatial analysis of landscapes: Concepts and statistics. *Ecology*, *86*(8), 1975–1987.

Walker, B. B., & Schuurman, N. (2015). The Pen or the Sword: A Situated Spatial Analysis of Graffiti and Violent Injury in Vancouver, British Columbia. *Professional Geographer*, *67*(4), 608–619.

Wang, A., Wang, Y., & Chen, Y. (2019). Hyperspectral image classification based on convolutional neural network and random forest. *Remote Sensing Letters*, *10*(11), 1086–1094.

Wang, Caixia, Pavlowsky, R. T., Huang, Q., & Chang, C. (2016). Channel bar feature extraction for a mining-contaminated river using high-spatial multispectral remote-sensing imagery. *GIScience and Remote Sensing*, *53*(3), 283–302.

Wang, Chunsheng, Chang, L., Zhao, L., & Niu, R. (2020). Automatic identification and dynamic monitoring of open-pit mines based on improved mask R-CNN and transfer learning. *Remote Sensing*, *12*(21), 1–20.

Wang, Z., & Bovik, A. C. (2006). Modern image quality assessment. *Synthesis Lectures on Image, Video, and Multimedia Processing*, *2*(1), 1-156.

Wang, Z., Bovik, A. C., Sheikh, H. R., & Simmoncelli, E. P. (2004). Image quality assessment: from error visibility to structural similarity. *IEEE Transactions on Image Processing*, *13*(4), 600–612.

Warner, T. (2011). Kernel-Based Texture in Remote Sensing Image Classification. *Geography*

Compass, *5*(10), 781–798.

Werner, T. T., Bebbington, A., & Gregory, G. (2019). Assessing impacts of mining: Recent contributions from GIS and remote sensing. *Extractive Industries and Society*, *6*(3), 993–1012.

Wheatley, M. (2010). Domains of scale in forest-landscape metrics: Implications for species-habitat modeling. *Acta Oecologica*, *36*(2), 259–267.

Whiteside, T. G., & Bartolo, R. E. (2015). Use of WorldView-2 time series to establish a wetland monitoring program for potential offsite impacts of mine site rehabilitation. *International Journal of Applied Earth Observation and Geoinformation*, *42*, 24–37.

Wickham, J., & Riitters, K. H. (2019). Influence of high-resolution data on the assessment of forest fragmentation. *Landscape Ecology*, *34*(9), 2169–2182.

Wiederholt, R., Paudel, R., Khare, Y., Davis, S. E., Melodie Naja, G., Romañach, S., Pearlstine, L., & Van Lent, T. (2019). A multi-indicator spatial similarity approach for evaluating ecological restoration scenarios. *Landscape Ecology*, *34*(11), 2557–2574.

Willis, K. S. (2015). Remote sensing change detection for ecological monitoring in United States protected areas. *Biological Conservation*, *182*, 233–242.

Witkin, A. P. (1984). Scale-space filtering: A new approach to multi-scale description. *In ICASSP'84. IEEE International Conference on Acoustics, Speech, and Signal Processing*, 150–153.

Woldai, T. (2001). Application of remotely sensed data and GIS in assessing the impact of mining activities on the environment. *Proceedings–17th International Mining Congress & Exhibition of Turkey*, 75–84.

Wright, A., Marcus, W. A., & Aspinall, R. (2000). Evaluation of multispectral, fine scale digital

imagery as a tool for mapping stream morphology. *Geomorphology*, *33*(1–2), 107–120.

Wu, H., & Li, Z. L. (2009). Scale issues in remote sensing: A review on analysis, processing and modeling. *Sensors*, *9*(3), 1768–1793.

Wu, J. (1999). Hierarchy and scaling: Extrapolating information along a scaling ladder. *Canadian Journal of Remote Sensing*, *25*(4), 367–380.

Wu, J. (2004). Effects of changing scale on landscape pattern analysis: Scaling relations. *Landscape Ecology*, *19*(2), 125–138.

Wu, J., & Li, H. (2006). Concepts of scale and scaling. In *Scaling and uncertainty analysis in ecology*, Springer, Dordrecht, 3–15

Wu, J., & Qi, Y. (2000). Dealing with scale in landscape analysis: An overview. *Geographic Information Sciences*, *6*(1), 1–5.

Wulder, M. A., Coops, N. C., Roy, D. P., White, J. C., & Hermosilla, T. (2018). Land cover 2.0. *International Journal of Remote Sensing*, *39*(12), 4254–4284.

Wulder, M. A., Masek, J. G., Cohen, W. B., Loveland, T. R., & Woodcock, C. E. (2012). Opening the archive: How free data has enabled the science and monitoring promise of Landsat. *Remote Sensing of Environment*, *122*, 2–10.

Wulder, M. A., White, J. C., Coops, N. C., Nelson, T., & Boots, B. (2007). Using local spatial autocorrelation to compare outputs from a forest growth model. *Ecological Modelling*, *209*(2–4), 264–276.

Xia, G., Hu, J., Hu, F., & Shi, B. (2016). AID : A Benchmark Dataset for Performance Evaluation of Aerial Scene Classification. *IEEE Transactions on Geoscience and Remote Sensing*, *55*(7), 3965–3981.

Xiao, Q., Li, G., Xie, L., & Chen, Q. (2018). Real-world plant species identification based on

deep convolutional neural networks and visual attention. *Ecological Informatics*, *48*, 117–124.

Xiaolong Dai, & Khorram, S. (1998). Quantification of the impact of misregistration on the accuracy of remotely sensed change detection. *IEEE Transactions on Geoscience and Remote Sensing*, *36*(5), 1566–1577.

Xie, X., Han, X., Liao, Q., & Shi, G. (2017). Visualization and pruning of SSD with the base network VGG16. *ACM International Conference Proceeding Series*, 90–94.

Xie, Y., Cai, J., Bhojwani, R., Shekhar, S., & Knight, J. (2020). A locally-constrained YOLO framework for detecting small and densely-distributed building footprints. *International Journal of Geographical Information Science*, *34*(4), 777–801.

Xu, K., Huang, H., Li, Y., & Shi, G. (2020). Multilayer Feature Fusion Network for Scene Classification in Remote Sensing. *IEEE Geoscience and Remote Sensing Letters*, *17*(11), 1894–1898.

Yandex, A. B., & Lempitsky, V. (2015). Aggregating Deep Convolutional Features for Image Retrieval. *Proceedings of the IEEE International Conference on Computer Vision*, 1269–1277.

Yang, B., Yan, J., Lei, Z., & Li, S. Z. (2015). Convolutional channel features. *In Proceedings of the IEEE International Conference on Computer Vision*, 82–90.

Yang, C., Gao, W., & Po, L. (2008). Discrete Wavelet Transform-Based Structural Similarity. *In Image Processing, 2008. ICIP 2008. 15th IEEE International Conference on 2008 Oct 12*, 377–380.

Ye, L., Wang, L., Sun, Y., Zhao, L., & Wei, Y. (2018a). Parallel multi-stage features fusion of deep convolutional neural networks for aerial scene classification. *Remote Sensing Letters*,

9(3), 294–303.

Ye, S., Pontius, R. G., & Rakshit, R. (2018b). A review of accuracy assessment for object-based image analysis: From per-pixel to per-polygon approaches. *ISPRS Journal of Photogrammetry and Remote Sensing*, *141*, 137–147.

Yosinski, J., Clune, J., Bengio, Y., & Lipson, H. (2014). How transferable are features in deep neural networks? 1–9.

Yu, L., Xu, Y., Xue, Y., Li, X., Cheng, Y., Liu, X., Porwal, A., Holden, E. J., Yang, J., & Gong, P. (2018a). Monitoring surface mining belts using multiple remote sensing datasets: A global perspective. *Ore Geology Reviews*, *101*, 675–687.

Yu, Y., Li, M., & Fu, Y. (2018b). Forest type identification by random forest classification combined with SPOT and multitemporal SAR data. *Journal of Forestry Research*, *29*(5), 1407–1414.

Yu, Y., & Liu, F. (2018). Aerial Scene Classification via Multilevel Fusion Based on Deep Convolutional Neural Networks. *IEEE Geoscience and Remote Sensing Letters*, *15*(2), 287–291.

Zagoruyko, S., & Komodakis, N. (2017). Paying more attention to attention: Improving the performance of convolutional neural networks via attention transfer. *5th International Conference on Learning Representations*, 1–13.

Zeiler, M. D., & Fergus, R. (2014). Visualizing and understanding convolutional networks. *In European Conference on Computer Vision*, 818–833.

Zeng, D., Chen, S., Chen, B., & Li, S. (2018). Improving remote sensing scene classification by integrating global-context and local-object features. *Remote Sensing*, *10*(5), 1–19.

Zhang, C., Yue, P., Tapete, D., Shangguan, B., Wang, M., & Wu, Z. (2020). A multi-level

context-guided classification method with object-based convolutional neural network for land cover classification using very high resolution remote sensing images. *International Journal of Applied Earth Observation and Geoinformation*, *88*, 102086.

Zhang, H., Li, Y., Zhang, Y., & Shen, Q. (2017). Spectral-spatial classification of hyperspectral imagery using a dual-channel convolutional neural network. *Remote Sensing Letters*, *8*(5), 438–447.

Zhang, Hongyun, Zhang, T., Pedrycz, W., Zhao, C., & Miao, D. (2019). Improved adaptive image retrieval with the use of shadowed sets. *Pattern Recognition*, *90*, 390–403.

Zhang, L., Guo, Z., Wang, Z., & Zhang, D. (2007). Palmprint verification using complex wavelet transform. *IEEE Transactions on Signal Processing*, 2007–2010.

Zhao, R., Pang, M., & Wang, J. (2018). Classifying airborne LiDAR point clouds via deep features learned by a multi-scale convolutional neural network. *International Journal of Geographical Information Science*, *32*(5), 960–979.

Zhao, W., & Du, S. (2016). Learning multiscale and deep representations for classifying remotely sensed imagery. *ISPRS Journal of Photogrammetry and Remote Sensing*, *113*, 155–165.

Zhao, Y., & Zhang, X. (2019). Calculating spatial configurational entropy of a landscape mosaic based on the Wasserstein metric. *Landscape Ecology*, *34*(8), 1849–1858.

Zhou, B., Khosla, A., Lapedriza, A., Oliva, A., & Torralba, A. (2016). Learning Deep Features for Discriminative Localization. *In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2921–2929.

Zhou, J., Yu, H., Smith, K., Wilder, C., Yu, H., & Wang, S. (2017). Identifying designs from incomplete, fragmented cultural heritage objects by curve-pattern matching. *Journal of*

Electronic Imaging, *26*, 011022.

Zhou, W., & Alan C., B. (2002). A universal image quality index. *IEEE, Signal Processing Letters*, *9*(3), 81–84.

Zhou, Y., Narumalani, S., Waltman, W. J., Waltman, S. W., & Palecki, M. A. (2003). A GIS-based spatial pattern analysis model for eco-region mapping and characterization. *International Journal of Geographical Information Science*, *17*(5), 445–462.

Zhu, Q., Zhong, Y., Liu, Y., Zhang, L., & Li, D. (2018). A Deep-local-global feature fusion framework for high spatial resolution imagery scene classification. *Remote Sensing*, *10*(4).

Zhu, X. X., Tuia, D., Mou, L., Xia, G. S., Zhang, L., Xu, F., & Fraundorfer, F. (2017). Deep Learning in Remote Sensing: A Comprehensive Review and List of Resources. *IEEE Geoscience and Remote Sensing Magazine*, *5*(4), 8–36.

Zhu, Z., & Woodcock, C. E. (2014). Continuous change detection and classi fi cation of land cover using all available Landsat data. *Remote Sensing of Environment*, *144*, 152–171.

Zhuang, S., Wang, P., Jiang, B., Wang, G., & Wang, C. (2019). A single shot framework with multi-scale feature fusion for geospatial object detection. *Remote Sensing*, *11*(5).

Zou, J., Li, W., Chen, C., & Du, Q. (2016). Scene classification using local and global features with collaborative representation fusion. *Information Sciences*, *348*, 209–226.