Digital Applications in Archaeology and Cultural Heritage 28 (2023) e00262

Contents lists available at ScienceDirect

Digital Applications in Archaeology and Cultural Heritage

journal homepage: www.elsevier.com/locate/daach

Geographic features recognition for heritage landscape mapping – Case study: The Banda Islands, Maluku, Indonesia

Muhamad Iko Kersapati *, Josep Grau-Bové

Data Science for Cultural Heritage, Institute for Sustainable Heritage, University College London, Central House, 14 Upper Woburn Pl, London, WC1H 0NN, UK

ARTICLE INFO

Keywords:
CNN
Computer vision
Historic maps
Machine learning
OBIA

ABSTRACT

This study examines methods of geographic features recognition from historic maps using CNN and OBIA. These two methods are compared to reveal which one is most suitable to be applied to the historic maps dataset of the Banda Islands, Indonesia. The characteristics of cartographic images become the main challenge in this study. The geographic features are divided into buildings, coastline, and fortress. The results show that CNN is superior to OBIA in terms of statistical performance. Buildings and coastline give excellent results for CNN analysis, while fortress is harder to be interpreted by the model. On the other hand, OBIA reveals a very satisfying result is very depending on the maps' scales. In the aspect of technical procedure, OBIA offers easier steps in pre-processing, in-process and post-processing/finalisation which can be an advantage for a wide range of users over CNN.

1. Introduction

The advancement of landscape mapping technology and methods has been an evolution over a very long time starting with the first attempt at visualisation by the European map makers. The oldest record may refer to Pliny's work (77 CE), *Naturalis Historia*. This "world inventory" was noticeable in the raising of geographical knowledge, mapmaking and cartography in particular the Roman Empire in the Flavian age (Bianchetti, 2020). In the next phase, the development of numerical approaches increased the precision of the map-making procedure. This invention can be found in some monumental works done by mathematicians, the specific one was Eratosthenes who correlated Pytheas' "Thule" (astronomical information) and distance theory from Borysthenes to integrate data of travel and geometrical measurements of the sphere (Carman and Evans, 2015).

Several centuries later, colonisation by the Europeans contributed to the depiction of the "other world" outside the continent. For instance, the first map of the Caribbean Islands by a European cartographer, Juan de la Cosa, conveyed the imagined space such as a gold mine including the physical attributes and the local people such as the African rulers, the gold traders, and the slaves. Similar situations were depicted in Southeast Asia, particularly Indonesia and Myanmar. In general, historians mention that the maps produced by the Dutch in the golden age constituted intrinsic agents of the European powers and were often the symbols of acquisition, academics, and gifts to interfere with political views or investments (Sutton and Yingling, 2020). The Dutch colonial maps of Indonesia shaped the territorialisation of land and resources, intertwined with administration, rent-seeking, violence and responsibility for geospatial knowledge. Meanwhile, mapping by the British became a major initiative in shaping the image of the regions in Burma (Faxon, 2022).

In the context of modern research, these colonial maps had transformed into a part of heritage data sources to depict human civilisation in the past through map investigation. Many recent studies occupy map archives as the main source of their analysis. Andrade and Fernandes (2020) interpreted the maps as an irreplaceable primary source of geographical and political information from the past. Old maps are moreover a source of data in Historic Landscape Characterisation (HLC) that has been developed largely in Britain (Turner, 2018). Ekim et al. (2021) extracted valuable information on transportation infrastructures and spatial distribution of settlements for quantitative and geometrical analysis. This spatial information is critical to facilitating decision-making in environmental management (Iosifescu et al., 2010). The historic maps can be utilised for environmental policy for example to calculate the environmental changes such as elevation and land use change as conducted by Tortora et al. (2015). They worked on a dataset of historical maps of Basilicata, Italy, produced in 1848, 1877, and 1953. The analysis resulted in a mutual exchange between the areas of agriculture and crops, which was reduced by almost half, giving more space to the natural areas. This analysis became a fundamental aspect to

* Corresponding author.
E-mail address: muhamad.kersapati.21@ucl.ac.uk (M.I. Kersapati).

https://doi.org/10.1016/j.daach.2023.e00262
Received 22 October 2022; Received in revised form 2 January 2023; Accepted 25 January 2023
Available online 26 January 2023

develop appropriate policies for territorial management and implement actions to safeguard the territory.

However, in practice, the limitation of human skills to draw the geographic features of the maps and integrate them into modern software (i. e., GIS) has become a prominent issue in heritage landscape analysis and visualisation. The researchers are unable to extract a bulk of data in any reasonable amount of time (Godfrey and Eveleth, 2015). For this reason, heritage experts and data scientists are investigating the use of technology to conceal the gap between manual works and automated recognition of spatial features on historic maps. There are two main groups of methods for the recognition and extraction of historic maps.

The first method is utilising neural networks. Uhl (2018) focused on the geospatial information extraction using Convolutional Neural Networks (CNN) for a scanned map sheet of Boulder, Colorado from 1966 at a scale of 1:24,000. The results indicated that although this method addresses the main issue of spatial granularity loss caused by the training labels at the patch level, the benchmark datasets are still not reaching the accurate and abundant training labels at the pixel level. A similar method was used by Petitpierre et al. (2021) on the dataset of 330 maps of Paris compared to 256 maps from all over the world. This work highlighted that neural networks are extremely robust in the face of figurative diversity. However, the cross-cultural validation revealed that greater ease of image segmentation occurred for the Western maps.

Another implementation of neural networks for image analysis was explored by Andrade and Fernandes (2020) to various styles of a historic map of Recife, Brazil, in 1808 using Conditional Generative Adversarial Networks (cGANs) or Pix2Pix to transfer and combine them with ortho-images (from satellite) into a new single image output. This method was also investigated by Christophe (2022) with multiple samples from simple-styled (Plan maps) to complex-styled maps (old Cassini, Etat-Major, or Scan50 BW). The results of these two related works indicated that this method is clearly not efficient for old complex-styled maps such as Cassini. This is related to the characteristics of the old-styled maps (more cluttered with dark brown and green colours and various textures of the papers). Furthermore, this method is more about evaluating the global similarity of styled images and integration to satellite imagery for composing neural image output rather than accomplishing the features recognition tasks.

The second group of methods is by integration of the Geographic Information System (GIS) which was experimented by Guirado (2021) in the given area of Cabo de Gata-Níjar Natural Park, Spain, utilising satellite images from Google Earth with three different spatial resolutions. Another experiment referred to Zatelli (2019) for a Historical Cadastral Map for the Province of Trento, Italy from 1859. He compared the method of Object-Based Image Analysis (OBIA) and Maximum Likelihood Classification (MLC) resulting in a surpassing performance of OBIA over the MLC with 98.43% vs 55.8% of accuracy. On the other hand, Gobbi (2019) identified several main classes of land use/land cover (LULC) using OBIA for three test maps. The output designated the values of kappa and overall accuracy ranged from 0.96 to 0.97 and 97%–99% consecutively.

Most of these previous works on geographic features recognition concluded that the level of success is highly affected by the characteristics of the old maps. In particular, the depiction of spatial elements on older cartographic images is very different from recent maps. Pchelov (2019) identified and systematised the semantics of symbols and emblems as the main characteristics of European maps from the 16th – 17th centuries. The territories were represented by imaginary heraldry, old symbolic designation of geographical phenomena (e. g., symbolic topography, buildings, roads, and geomorphological features). Compared to maps from the 20th century, maps from the previous centuries are more difficult to be interpreted by a computer system which can affect the accuracy of the built model in reading the graphics from images. This characteristic is related to the emergence of the map, which was originally closely related to artistry, especially painting, where many schools of landscape painting provide the course of cartography at the same time (Rees, 1980). On the other hand, the low quality of the archive papers also be affecting factors of the analysis performance. Several types of defective conditions (see Fig. 1) such as stains or torn edges may decrease the ability of the model to interpret the images.

Although there has been a plethora of research on computational image recognition from digital map archives, the predominant studies occupied the mainland Europe maps rather than the colonised regions. In reality, studying historical maps of the ex-colonies becomes more important in this global era. The essence is to uncover the salient spatial traces (such as buildings, areas, and place names) in the context of decolonisation (Knudsen, 2021). Moreover, the national institutions are able to reclaim them as heritage assets of the country. An example of colonised regions important to be analysed is the Banda Islands, Indonesia. The characteristics of tropical volcanic islands are the main difference between the Banda Islands and other regions in Europe that have been studied, most of which is land.

Lape (2002) utilised historic maps of the Banda Islands from 1599, 1602, and 1615 for retracing the heritage evidence of these islands. His invention was highlighting several main results: the shift of settlement locations, spatial organisation and the relationships between the settlements and resources for tracing cross-cultural contact and interaction. This finding became a critical source to depict the Banda Islands as the historical landscape and milestone of the start of the Dutch colonial period in the archipelago. At this point, the works of related experts (such as anthropologists, archaeologists, and historians) as technology users should meet adequate or even advanced tools to facilitate the investigation process. Hence, it is essential to examine and evaluate the methods of geographic features recognition, particularly in the case study of historic maps of the Banda Islands as an example of colonised territories.

This study aims to compare and evaluate the image analysis methods of CNN and OBIA in the limitation of historic maps of the Banda Islands as well as to give recommendations on the best method to be developed for heritage landscape mapping. The results of this comparison of methods indicated that in terms of statistical performance, CNN is superior to OBIA with an accuracy value of 0.9270 and 0.9049 respectively. However, from the technical procedure, CNN requires further complex processing to be implemented in heritage landscape mapping, hence the users of this method need a basis of computer and data science to proceed with this analysis. Meanwhile, OBIA has more convenient procedures of georeferentiation and training sample creation that make the integration with GIS easier for advanced mapping as well as more friendly to be used by experts from various related disciplines. These can be taken into consideration for users to choose according to the advantages of each method.

2. Materials and methods

2.1. Workflow and dataset

The scanned images of map archives of the Banda Islands are selected with a colour depth is 24 and a resolution is 96 dpi as input to data preprocessing such as image enhancements, augmentation, and georeferentiation to enhance the quality and quantity of the data. Passing this operation, the images will be processed and analysed separately for each procedure scheme of CNN and OBIA. The statistical output of performance from both methods will be compared and evaluated such as loss and accuracy function, precision, and recall displayed in confusion matrices. Four historic maps of the Banda Islands are collected from the Dutch National Archives for the Indonesian archipelago maps collection. These maps constitute the primary source to portray the islands as a part of the Dutch colonisation with different periods of making that represent various eras and conditions as a heritage landscape. Fig. 2 shows the steps and dataset of this research.

Located in the Province of Maluku, Indonesia (S40 28–40 39 E129

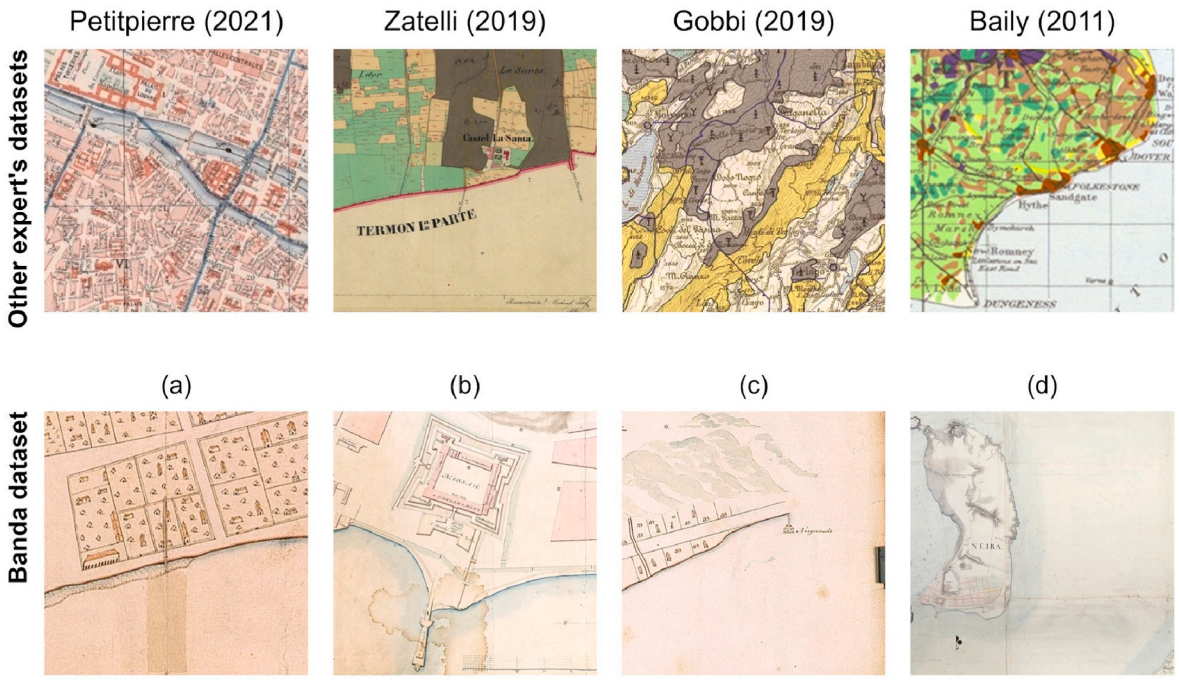

Fig. 1. Comparison between historic maps from previous studies and the Banda Islands. The Banda dataset has different cartographic characteristics of volcanic islands and some prominent issues of physical material conditions: (a) adhesive marks; (b) water drop stain; (c) torn edges and uneven colour pigment; (d) fold marks. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

39–130 04), these islands are also known under the name "Spice Islands" and consist of eleven small volcanic islands which include Neira, Rozengain, Pisang, Lonthoir (Banda Besar), Volcano, Rhun, and Ay. Has long been described in European historical records (such as Tomé Pires, Anna Forbes, Ridley, and Coolhaas) (Kersapati, 2021), making this region registered as a permanent delegation of the Republic of Indonesia to UNESCO as "The Historic and Marine Landscape of the Banda Islands" (UNESCO, 2015). The details of the dataset are described in Table 1.

2.2. Image enhancements

Digital archives, especially scanned images such as historic photographs, paintings, and maps are prone to experiencing problems in terms of visual attributes. The inferior conditions of historic documents need to be fixed by image enhancement to bring out specific features explicitly (Archana, 2016) before considering advanced processing and analysis. Some observations from Parkin (2018) postulate that the numerical properties of colour in digital images are requisite for computational measures and analysis such as hue, saturation, and sharpness.

A finding from Flachot and Gegenfurtner (2021) suggested that the kernels in CNNs tend to be mainly sensitive to changes in hues rather than changes in chroma or saturation, while sharpness is inversely related to blur which is typically characterised by the spread of edges (Yu, 2017) and very crucial in object identification using both CNN and OBIA. In this research, the image enhancements are using automatic enhancements in Microsoft Photos to cut the processing time by globally operating on the entire image without considering image content (Yan, 2016). This application enables users to edit and adjust the image automatically by selecting enhance button in the filter section (Microsoft, 2022). Kaufman (2012) in his research highlighted some features that were automatically adjusted by this method such as faces, sky as well as shadowed salient regions, and then applies a sequence of empirically determined steps for saturation, contrast as well as exposure adjustment.

2.3. Training data generation

The scientific value of old maps is reflected through the identification of geographical features such as land and anti-military map expression, elements of coastal defence, residential, administrative areas, etc (Jiang, 2017). However, in an archaic environment, the spatial elements in historic maps tend to be depicted in a non-uniform scheme of geographical space (Rees, 1980). There are no standards on what should be classified, in other words, this classification relies on the human's interpretation of the given maps.

For instance, Mahdianpari (2018) organised the land cover into nine classes for CNN processing in his study as he referred to Baily (2011) for the old cartographic classification of land use in Britain. Meanwhile, Skaloš & Engstová (2010) only classified land use into three classes in their research. Classification of geographical features is needed to give the labels on the trained data based on their appearance on historic maps. To uniform the maps produced in different years in this study, the classification of geographical features emphasises three main aspects: buildings, coastline, and fortress.

Training models for image classification involve a large number of image datasets. This is related to the process of data labelling to construct a higher quality performance in particular for CNN analysis. A large number of samples is required to extract robust features by minimising the error of estimating the true labels of training images (Yu, 2017). Generating synthetic pictures can be an alternative to increase the number of samples to be trained.

The creation of artificial images can be performed using data augmentation by employing geometric transformations: translations, rotations, cropping, flipping, resizing, etc (Wong, 2016). This method has proven effective in many fields, such as image processing and object recognition (Lashgari et al., 2020). 1440 samples of images are produced for this dataset through the augmentation process with the cropped patches dimensions of 42 × 42 pixels from the whole of selected historic maps as shown in Fig. 3.

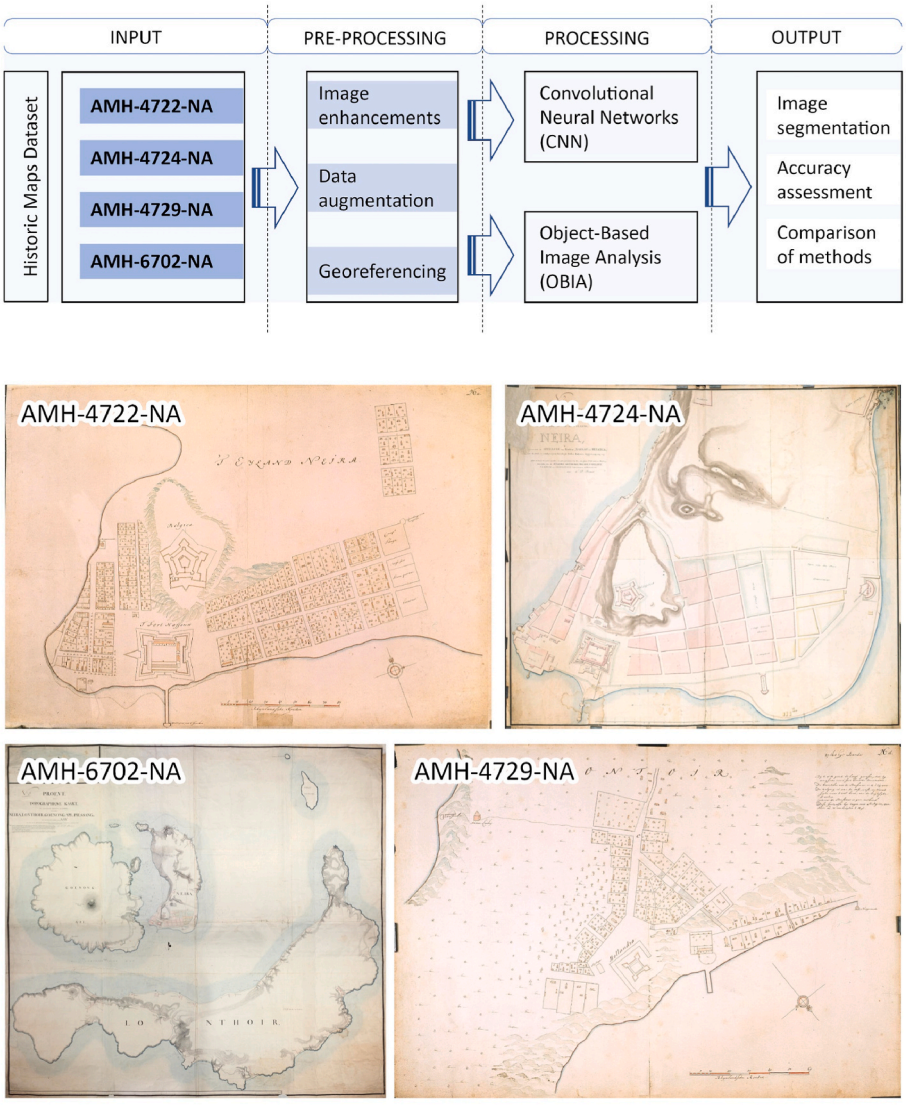

Fig. 2. Workflow and selected colonial maps of the Banda Islands.

Table 1
Dataset description.

Name	Description
AMH-4722-NA	Title: Map of part of the island of Neira, showing the forts Nassouw and Belgica. Produced: between 1690 and 1743. Dimension: 29.1 in (height); 20.8 in (width). Scale: 80 Rhynlandsche roeden (1:3500).
AMH-4724-NA	Title: Map of the island of Neira showing the Nassouw and Belgica forts. Produced: 1791. Dimension: 57.8 in (height); 50.1 in (width). Scale: unidentified clearly, estimated 80 Rhynlandsche roeden (1:3500).
AMH-4729-NA	Title: Fort Hollandia on Lonthor. Produced: between 1690 and 1705. Dimension: 29.5 in (height); 20.8 in (width). Scale: 60 Rhynlandsche roeden (1:2600).
AMH-6702-NA	Title: Map of the Banda islands. Produced: between 1750 and 1796. Dimension: 76.3 in (height); 71.2 in (width). Scale: unidentified clearly, estimated 1000 Rhynlandsche roeden (1:50,000).

2.4. Georeferentiation

The utilisation of Geographic Information Systems (GIS) in reconstructing the heritage landscape has been significantly rising in recent decades (Gobbi, 2019). Integrating the cartographic images to GIS in particular for OBIA needs georeferencing as the prerequisite operation. The basis of this process is to find some markable points in the scanned images that have equivalents in referred geospatial data. As most object recognition tasks in computer vision engage the data labelling process, this recognition allows the operator to label points in the image with their position in the real world and calculate a mapping of the image into a coordinate system (Luft and Schiewe, 2021).

Several main steps of georeferencing (ESRI, 2022a,b): **(1) Aligning the raster with control points**, involves identifying a series of ground control points (GCPs) in x and y coordinates that link locations on the raster dataset with locations in the spatially referenced data. In this case, the historic maps of the Banda Islands refer to the polygons (in shapefile format) of the administrative area of the Banda Islands as a part of the Maluku Province, Indonesia. The administrative boundaries, as well as the coastline, are identified as linear features of arbitrary geometry that can be defined easily by the users (Papakosta et al., 2012). **(2) Transforming the raster**, after the creation of control points, the raster dataset can be transformed to the coordinates of the target data. **(3)**

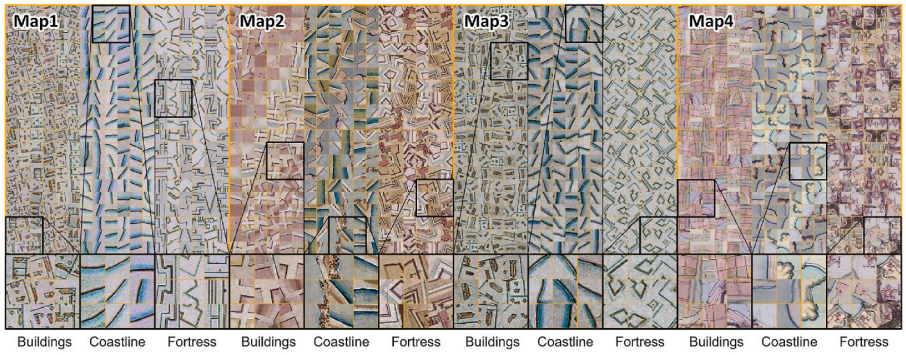

Fig. 3. Generated training samples from the augmentation process for each determined geographical feature (42 × 42 pixels) on each historic map: buildings, coastline, and fortress.

Interpret the root mean square error, the total error is computed by taking the root mean square (RMS) sum of all the residuals to compute the RMS error. The RMS indicates the error of ground control points (GCPs) caused by the gap between the scanned maps and the referred polygons. However, the GCP database is usually obtainable in the present aerial photograph, instead of scanned archives. Consequently, the georeferencing process of historic maps only shows the x and y coordinates of referred geospatial data and the scanned images due to the unavailability of the GCP database on the historic maps (Cléri et al., 2014). **(4) Persist the georeferencing information** and save the georeferenced images in.tiff format.

2.5. Experiments

Experiments on Convolutional Neural Networks (CNN) and Object-Based Image Analysis (OBIA) were conducted separately through different software. The CNN experiment utilised R with LeNet architecture, the simplest CNN architecture for object detection. LeNet (refers to LeNet-5) is the regular Convolutional Neural Networks (CNN) model, developed for grayscale image recognition from the MNIST dataset (Pearline & Kumar, 2019). The basis of convolutional networks combines three principles: local receptive fields, shared weights (or weight replication), and spatial or temporal subsampling. This is to certify the degree of shift, scale, and distortion invariance (Lecun, 1998). This architecture is constructed in seven layers as shown in Fig. 4.

The input layer constituted image samples resized 42 × 42, adapted from Uhl (2018) for the implementation of historic maps which need more detailed pixels to identify buildings and other spatial traces. The first convolution layer employed twenty 5 × 5 filters with stride, s = 1 and the subsampling layer consisted of the average pooling with 2 × 2 filters and stride, s = 2 following the first convolution layer. The calculation formula for one pixel in the next layer is expressed as follows (Albawi et al., 2017):

$$net(t,f) = (x.w)[t,f] = \sum_m \sum_n x[m,n]w[t-m,f-n] \quad (1)$$

where.

$net(t,f)$: output in the next layer

x: input image/layer

w: kernel/filter matrix

The second convolution had fifty 15 × 15 filters and was followed by an average pooling layer with 2 × 2 filters using a stride, s = 2. Rectified linear unit (ReLU) activation function is expressed:

$$ReLU(x) = max(0,x) \quad (2a)$$

$$\frac{d}{dx} ReLU(x) = \{1 \; tf \; x > 0; \; 0 \; otherwise\} \quad (2b)$$

The ReLU activation function effectively diminished the gradient vanishing problem, it trained the deep neural network the way of supervision, without relying on the unsupervised layer-by-layer pre-training, which significantly improved the performance of the deep neural network (Zhang, 2019). The outputs of the second max-pooling layer were converted from 2D vectors into 1D vectors. These 1D vectors were inputs to the first fully connected layer and the final layer utilised the softmax activation function. A test was conducted on Map1 to calculate all the parameters involved in this analysis. Table 2 indicates 328,573 parameters from the total layers of CNN using LeNet architecture.

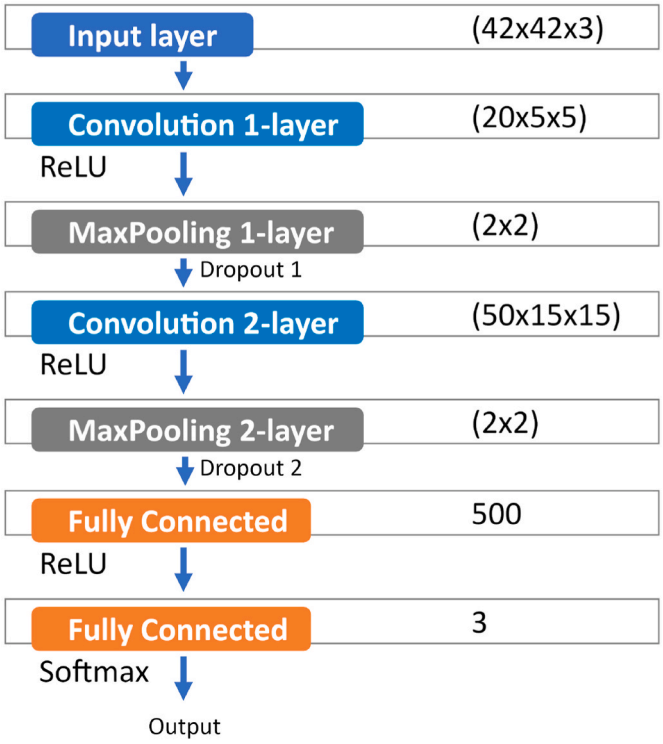

Fig. 4. LeNet architecture for CNN: 1-layer input, 2-layers convolution, 2-layers maxpooling, and 2-layers fully connected.

The second experiment was conducted to examine the OBIA performance using ArcGIS Pro. As an alternative to pixel-based image classification, the developers of this method claim that object-based classification gives higher accuracy and obtrudes characteristic texture features which are neglected in conventional classifications (Blaschke and Strobl, 2001). This advantage brings OBIA as a reformer since its invention where most of the previous studies revealed that the performance of OBIA exceeds the results of pixel-based analysis such as

Table 2
Parameters calculation for the total layers.

Layer (type)	Output Shape	Param #
conv2d_1 (Conv2D)	(None, 38, 38, 20)	1520
max_pooling2d_1 (MaxPooling2D)	(None, 19, 19, 20)	0
dropout_2 (Dropout)	(None, 19, 19, 20)	0
conv2d (Conv2D)	(None, 5, 5, 50)	225050
max_pooling2d (MaxPooling2D)	(None, 2, 2, 50)	0
dropout_1 (Dropout)	(None, 2, 2, 50)	0
flatten (Flatten)	(None, 200)	0
dense_1 (Dense)	(None, 500)	100500
dropout (Dropout)	(None, 500)	0
dense (Dense)	(None, 3)	1503
Total params: 328,573		
Trainable params: 328,573		
Non-trainable params: 0		

supervised using Maximum Likelihood Classification (Uça-Avcl: 2011; Whiteside, 2011) or combination of supervised and unsupervised classification (Weih and Riggan, 2010). However, the results from Gao and Mas (2008) indicated the lower execution of OBIA against pixel-based both Nearest Neighbour (NN) and MLC utilising cubic-filtering and mean-filtering methods. The differences in the results from various previous studies are trying to be proven through this research.

In general, three major phases were carried out as shown in Fig. 5 for the OBIA scheme in this study: (1) raster segmentation; (2) classifier training; (3) classification and accuracy assessment. Some parameters should be fulfilled while segmenting an image such as spectral and spatial detail with valid values ranging from 1.0 to 20.0. Spectral detail is the level of importance given to the spectral differences of features in an image. Smaller values result in more smoothing and longer processing times. Spatial detail set the level of importance given to the proximity between features. A higher value is appropriate for a scene where the features of interest are small and clustered together. In contrast, smaller values create spatially smoother outputs. Minimum segment size parameters were set to determine the minimum mapping unit (in pixels) and best-fitting neighbour segment (ESRI, 2022a,b). In the next step, the creation of samples for a specific classifier was required as representative sites for all the classes were classified on the maps; otherwise, the default classification schema was from the 2011 National Land Cover Database (NLCD 2011).

As a standard parametric procedure, it is suggested that each class should be represented by a statistically significant and equivalent number of training samples with a normal distribution to produce reliable results (ESRI, 2022a,b). The total number of training samples was 360 polygons, divided into four classes (90 polygons for each class): buildings, coastline, fortress, and 'other'. This number was implemented for each map in the dataset. So, the total number of samples was 1440 polygons. The Maximum Likelihood algorithm used in this analysis classified pixels on multivariate probability density functions (pdf) for each established class. Statistical properties of training images from ground reference data were typically used to estimate the pdfs of the classes by assigning each unidentified pixel to the class with the highest probability at the pixel location with a decision rule (Lee and Warner, 2004), denoted as follows:

$$p(X|\omega_c)p(\omega_c) \geq p(X|\omega_i)p(\omega_i) \tag{3}$$

where.

$p(X|\omega_c)$: probability density functions (pdf) of X, given that X is a member of class c

$p(\omega_c)$: a priori probability of class c in the image

X: the spectral multivariate vector

i: class number among the m number of classes in the image

The final stage was performed by classifying the image based on the trained classifier and creating the confusion matrix to evaluate the model. Repetition of the training procedure is possible to perform if the previous accuracy level shows unsatisfying results.

2.6. Accuracy assessment

Once a classification model is established, an assessment is needed to evaluate the overall effectiveness of the learning models (Chefira and Rakrak, 2021). In image classification of historic maps, accuracy assessment is also important for exploring model generalisation and transferability to new data and/or geographic extents (Maxwell et al., 2021). Evaluation of the classification model's performance is divided into some criteria: Accuracy, Precision, Recall, and F1 score. The calculation is expressed as follows:

The overall model performance:

$$Accuracy = \left(\frac{TP + TN}{TP + FP + FN + TN}\right) \tag{4}$$

The accuracy of the positive predictions:

$$Precision = \left(\frac{TP}{TP + FP}\right) \tag{5}$$

The coverage of the positive sample:

$$Recall = \left(\frac{TP + TN}{TP + FN}\right) \tag{6}$$

And the harmonic recall-precision average:

$$F1score = \left(\frac{2 \times Recall \times Precision}{TP + FP + FN + TN}\right) \tag{7}$$

True Positive (TP), False Negative (FN), False Positive (FP) and True Negative (TN) measure variables are used to rate the performance of correct and incorrect predictions. Table 3 shows the standard confusion matrix for binary class predictions to describe the performance of CNN and OBIA models. Rows represent the instances in the reference (actual)

Table 3
Standard confusion matrix.

	Reference (0)	Reference (1)
Predicted (0)	True Positive (TP)	False Positive (FP)
Predicted (1)	False Negative (FN)	True Negative (TN)

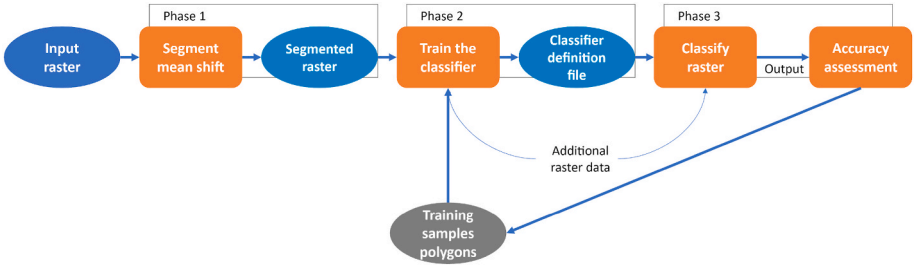

Fig. 5. Image processing for OBIA, consists of three main phases: segmentation, training, and classifying.

binary class, while each column represents the instances in the predicted binary class.

Another statistical measure used to compare the performance of CNN and OBIA is Cohen's kappa statistics. Many researchers claim this measure is very good to handle both multi-class and imbalanced class problems, particularly in the image classification model (Widmann, 2022). The kappa is expressed as follows:

$$K = \frac{P_0 - P_e}{1 - P_e} = 1 - \frac{1 - P_0}{1 - P_e} \tag{8}$$

where.

P_0: the observed agreement

P_e: the expected agreement

This measure tells how much better the performance of created classifier over the performance of a classifier that simply guesses at random according to the frequency of each class. This is the advantage to overcome the imbalanced samples of images. Cohen's kappa is always less than or equal to 1. Values of 0 or less, indicate that the classifier is useless. There is no standardised way to interpret its values. According to Landis and Koch (1977), a value < 0 is indicating no agreement, 0–0.20 as slight, 0.21–0.40 as fair, 0.41–0.60 as moderate, 0.61–0.80 as substantial, 0.81–1 as almost perfect agreement (McHugh, 2012).

3. Results and discussion

3.1. Image segmentation

Image segmentation for CNN was created by assigning each label of determined classes from the image samples of geographical features. The comparison between the area traces of the output and the original images can be seen in Fig. 6. The lines created by the LeNet algorithm are distinguishable for features of buildings and fortress on Map1 and Map2, while the fortress is difficult to be observed on Map3 and Map4. The coastline is also slightly difficult to determine, especially in the upper left part of Map1, Map2, and some scattered parts on Map4. This is

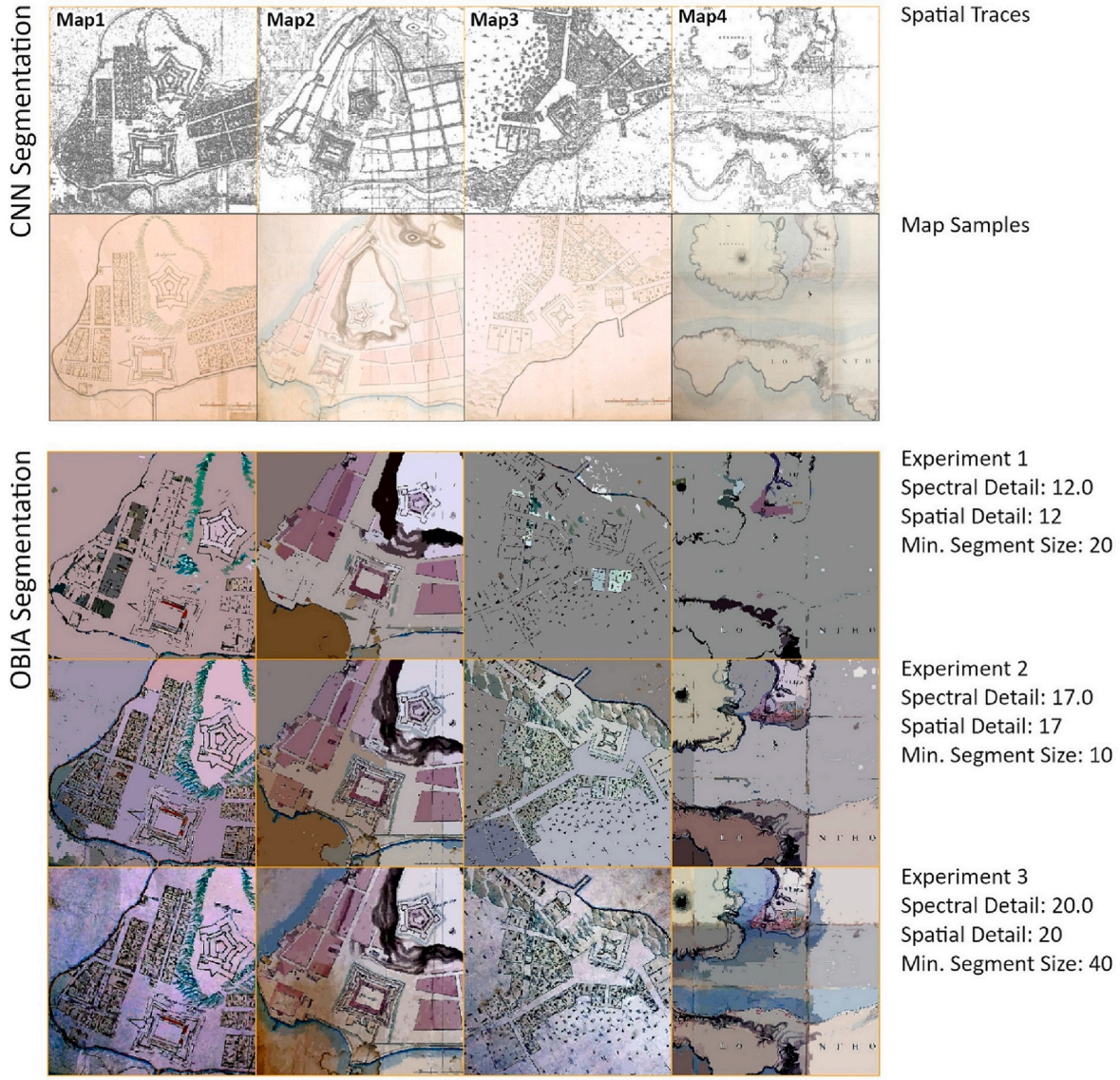

Fig. 6. Image segmentation for CNN shows some areas depict clear traces, while some others show very dense areas. This is affected by the texture and uneven colour of the maps. Image segmentation for OBIA shows that the second experiment gives the best result for image classification by eliminating some excessive spectra yet retaining the essential details of images. The first experiment loses many details of geographic features, while the third experiment complicates the classification process of images. The OBIA segmentation results are rotated due to the process of georeferencing and hence have a slightly different appearance than the map samples and CNN segmentation output. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

indicating that the CNN model learned to distinguish the areas from the classes based on the colour and texture (Uhl, 2018). The complexity of the visual components of the maps can also be observed in the smoothness of the boundaries of the building areas which show rugged patterns on Map1 and Map3, while the building areas on Map2 and Map4 reveal soft and smooth surfaces.

In OBIA classification, image segmentation plays a significant role. The main task of OBIA segmentation is to simplify a complex image into homogeneous areas followed by analysing and classifying the objects. In other words, the image segmentation process aims to eliminate the unnecessary elements associated with per-pixel classification to create objects representing spatial features classes (Grippa, 2017). Some parameters were required to construct an image segmentation: spectral detail, spatial detail, and minimum segment size which were set differently in three experiments. In the first experiment, the spectral detail and spatial detail are set at 12 and min. Segment size = 20. In the second test, spectral detail and spatial detail are set at 17 and min. Segment size = 10. In the third experiment, spectral detail and spatial detail are set at 20 and the minimum segment size = 40.

The results for the three experiments of creating OBIA segmentation as well as the function of parameters and image quality are shown in Fig. 7. Based on the experiments, the three segmentation parameters had different effects on the quality of segmentation output and were significantly visible on all maps. Quality in this context is defined as the proportion between missing and added pixel values from its original pixels and evaluated by the level of change of this proportion. In the results of the first experiment, a small spectral and spatial value gives an overgeneralisation of the image. Thus, many parts (pixels) are missing from objects. On the other hand, too high spectral and spatial values give very diverse segmentation (oversegmentation), producing almost the same output as the original image so that objects cannot be distinguished clearly. Meanwhile, a large minimum segment size value contributes to blurry image results, while a small value provides more detailed results.

The graph on the left shows that the spectral detail reached its maximum quality on a scale of 17. The addition of the spectral value inversely decreased the image quality and objects were more difficult to distinguish. The graph in the middle illustrates the relationship between spatial detail and quality. The spatial detail gained its maximum quality on a scale of 17, relatively similar to spectral detail. However, there was a slight difference in that adding the spatial detail value after the peak gave maximum results for small-scale maps that require a very high level of detail to be able to identify objects. In contrast, increasing the value of spatial detail after the peak point on large-scale maps reduced image quality due to objects becoming more complicated to identify. Meanwhile, the chart on the right explains the effect of minimum segment size on output quality. A smaller segmentation value contributed to the higher quality of outputs. On the other hand, a large segment size value resulted in poor image quality by reducing the sharpness of the output.

3.2. Accuracy assessment and evaluation

The loss and accuracy functions are cardinal to be evaluated from the built CNN model. The loss describes how well or poorly the model performs after each iteration of optimisation, while the accuracy tells us about the performance of the algorithm to predict the given images. After the first iteration, the loss reduced (approximately to 0) and the accuracy increased to the highest point of approximately 100%. Several hyperparameters were set for optimisation: learning rate = 0.01, decay = 1e-6, momentum = 0.9, batch = 20, validation split = 0.2, and epochs = 100. The number of epochs was adapted from Pearline & Kumar (2019) who revealed that the model stabilised for 100 epochs, while the increase of epochs number (200, 300, 400, 500) did not contribute significantly where the loss keeps increasing and accuracy did not improve.

An illustration of the loss and accuracy curve for the validation set is shown in Fig. 8. The speeds of convergence were different for each map. Map1 and Map3 reached convergence around the 5th epoch. Map2 and Map4 showed the opposite performance where Map2 is slower (at the 15th epoch), while Map4 is faster (less than at the 5th epoch). This difference in behaviour was potentially due to the complexity level of the geographic features of each map.

A confusion matrix was generated to quantify the uncertainty of the image segmentation created by the CNN model as shown in Fig. 9 (left). The matrices were created by comparing the prediction results of the images to the labelled trained samples. The total number of samples was 360 images for each map with an equal division of buildings = 120, coastline = 120, and fortress = 120. Buildings and coastline gave a good result of 33.33% for all maps. This means that from the 120 images of buildings and 120 images of coastline, the algorithm correctly predicted all of the images to be of buildings and coastline. On the other hand, the fortress tended to be more difficult to predict as it shows a value of less than 33.33% on all maps. On Map1, Map2, and Map4, the model made the most errors by predicting fortresses as buildings, while on Map3, the error tended to predict the fortress as coastline.

A confusion matrix for the OBIA model (shown in Fig. 9 right) was generated based on the pixel-wise comparison between prediction results and manually recognised samples. The classes outside the three main geographic features (buildings, coastline, and fortress) must be defined as "other". This shows that the spatial approach in OBIA identifies all areas on the map into their respective classifications.

The total number of samples was 360 images for each map with an equal division of buildings = 90, coastline = 90, fortress = 90, and other = 90. Buildings and coastlines were classes that provide the highest

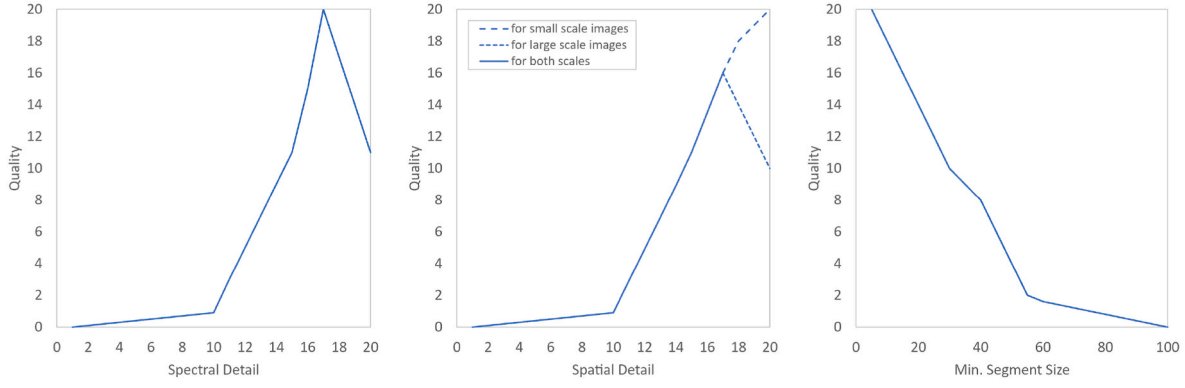

Fig. 7. Function of parameters for OBIA segmentation and quality of generated images: spectral detail, spatial detail, and minimum segment size. Spectral detail gives a maximum quality at a rate of 17. A higher rate of spatial detail gives better quality for small-scale maps. In contrast, the quality decreases after the maximum rate of 17 for large-scale maps. A lower minimum segment size contributes to a higher quality of images.

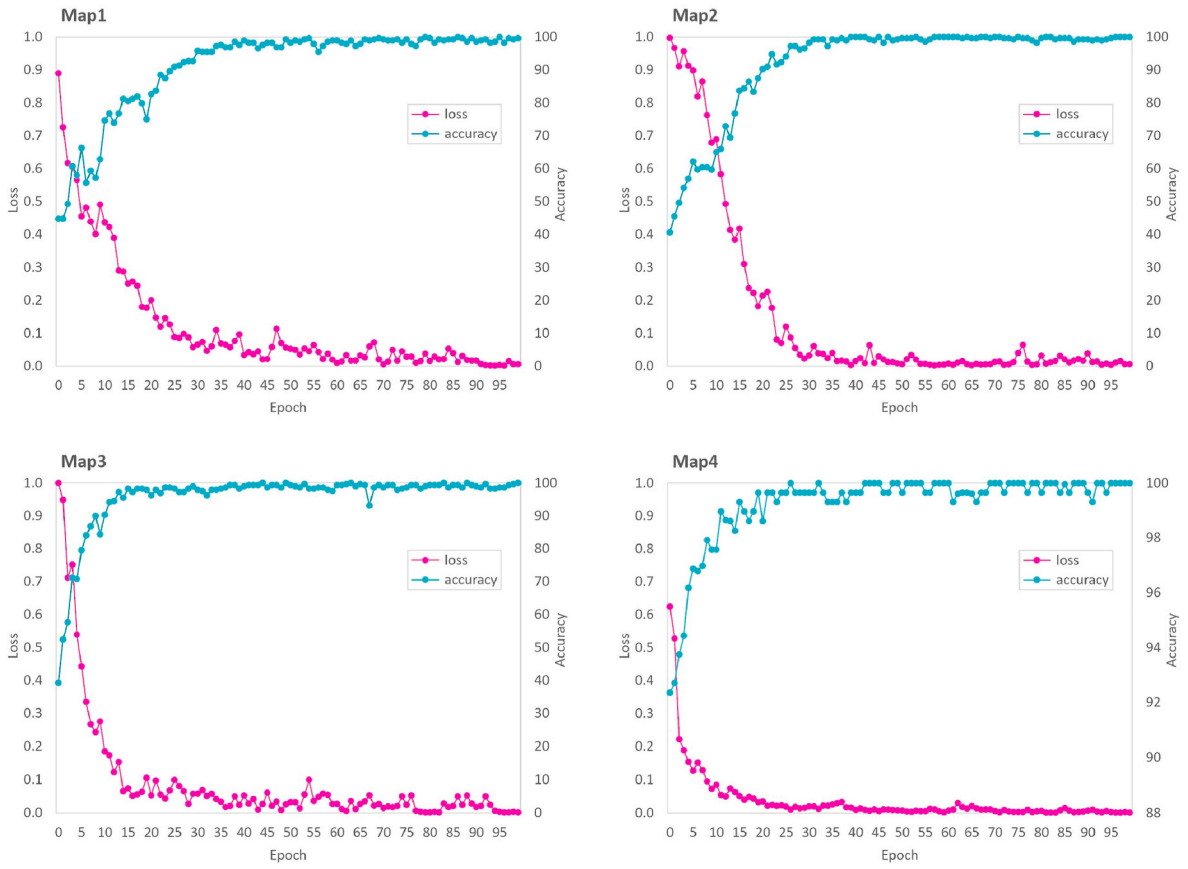

Fig. 8. The value of loss and accuracy as a function of epochs for the CNN model test.

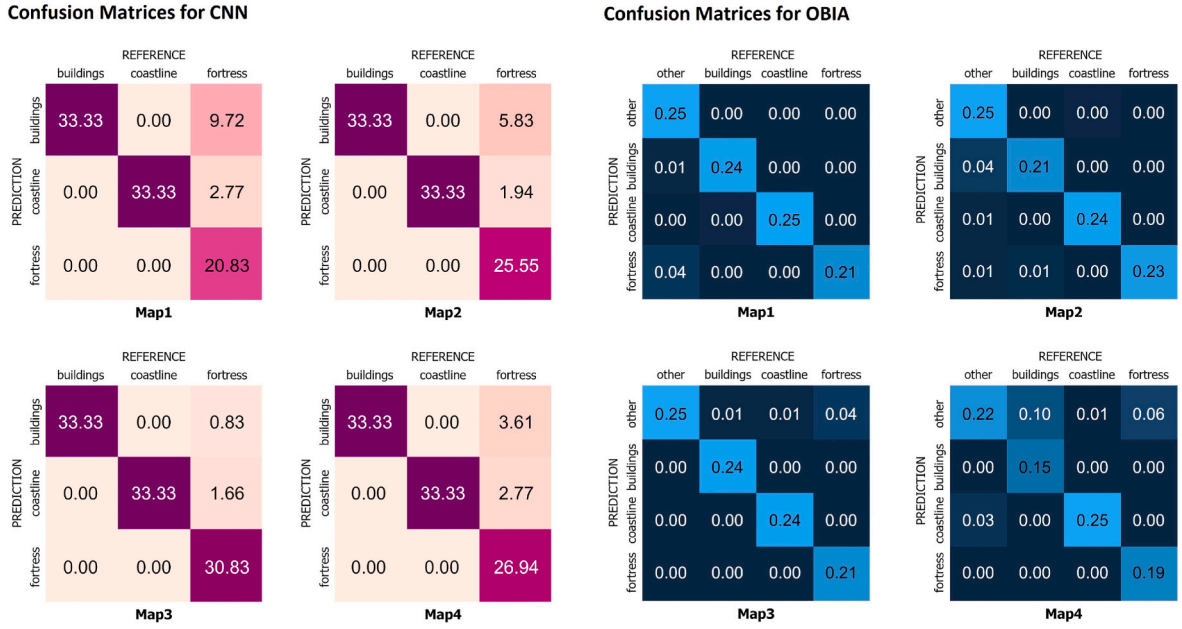

Fig. 9. Confusion matrices of the four maps for CNN (left) and OBIA (right), displayed in per cent of the total trained images.

accuracy on Map1, Map2, and Map3 with a percentage of approximately 25% of total samples, while on Map4, the percentage of all classes that were correctly predicted tends to be lower. This indicates that the errors resulted from the prediction bias of these classes with the "other" class. The difference in the level of prediction accuracy of the model was high potentially related to the scale at which large-scale maps (Map1, Map2, and Map3) had higher accuracy than small-scale maps (Map4).

The class-specific accuracy measures were evaluated by precision, recall, and F1 score. The precision measures the percentage of correct images, the recall evaluates the model, and the F1 score combines the precision and recall (Zeya, 2022). Precision is calculated by dividing the True Positives and all the Positives. For this case, that would be the

measure of one class of geographic features that are correctly identified as its true class divided by all the correctly predicted classes. Meanwhile, recall measures the model correctly identifying True Positives (Huilgol, 2022). Thus, for all images that were identified as actually their true classes, recall tells how many times the model correctly identified the images as true classes of geographic features.

Table 4 shows the precision, recall, and F1 score for each class of geographic features on each map for CNN and OBIA. The evaluation for CNN indicates that buildings and coastline on all maps had high precision and recall (the recall value of these two classes was 1). The high precision value relates to a low false positive rate, and the high recall relates to a low false negative rate. In contrast, the fortress assigned a high value for precision (1) yet a lower recall value in particular for Map1. This shows that the classifier was returning accurate results but failed to return some of the positive results. For the case of Map1, the recall value of 0.6250 given by the fortress indicates that the model was going to miss around 40% of the identified fortress images as True Positive. Instead, they could be identified as buildings or coastline. Furthermore, the F1 score as the harmonic mean of recall and precision indicates a good model with a value of approximately 1 for each class of geographic features on each map.

For the OBIA evaluation, the model gave a good performance. The precision value on Map3 produced a maximum rate of 1 for all classes and the highest recall value was on Map1 with a maximum rate generated by coastline and fortress. The F1 score result reveals that the model works well with an overall value of approximately 1.

3.3. Comparison of methods

Analysis of images from CNN and OBIA gives different results. Various factors such as spatial information contained in the map affect the level of difficulty of the machine in reading and redrawing according to the determined classification. The physical condition of the historical map when it is scanned is also an important factor, particularly for image segmentation. Excellent segmentation is only possible to be produced by the traditional method of computer vision (such as CNN) for good-quality images with large and homogeneous corpora (Muhs, 2016). In other words, inflexibility occurs for any image analysis, particularly for historic maps and the users should rely on some consideration of what is the best method to be implemented on their dataset. Hence, the researchers need precise knowledge of the nature and the visual characteristics of the maps: size, shape, texture, colour, etc (Graeff and Carosio, 2002).

The method of CNN has a major weakness in reading shape, especially if the object being analysed has a complex or diverse shape, in this study such as the corners of the fortress. Some previous research also reported some similar troubles in detecting morphological features such as lines (Miao, 2013) or closed polygons (Miyoshi, 2004). Reconstruction algorithms are required to solve the problem of incomplete lines, due to the degradation of the document or graphical choices (e.g., dashed lines) (Mello, Costa, & J., 2012). Furthermore, the information overlay as a common feature in cartographic images inhibits the extraction process of the geometries.

Table 4
Evaluation matrix for CNN and OBIA.

	CNN			OBIA		
	Precision	Recall	F1 Score	Precision	Recall	F1 Score
Map1						
Buildings	0.7742	1.0000	0.8727	0.9556	0.9885	0.9718
Coastline	0.9231	1.0000	0.9600	0.9889	1.0000	0.9944
Fortress	1.0000	0.6250	0.7692	0.8333	1.0000	0.9091
Map2						
Buildings	0.8511	1.0000	0.9196	0.8222	0.9367	0.8757
Coastline	0.9449	1.0000	0.9717	0.9667	0.9886	0.9775
Fortress	1.0000	0.7667	0.8679	0.9222	1.0000	0.9595
Map3						
Buildings	0.9756	1.0000	0.9876	1.0000	0.9778	0.9888
Coastline	0.9524	1.0000	0.9756	1.0000	0.9778	0.9888
Fortress	1.0000	0.9250	0.9610	1.0000	0.8222	0.9024
Map4						
Buildings	0.9023	1.0000	0.9486	1.0000	0.6111	0.7586
Coastline	0.9231	1.0000	0.9600	0.8824	1.0000	0.9375
Fortress	1.0000	0.8083	0.8940	1.0000	0.7444	0.8535

OBIA seemingly appears to patch these certain issues in recognising spatial images that could not be solved by CNN. Image segmentation gives much improvement by suppressing isolated pixels and small clusters. Thus, classification error resulting from high within-object variance was efficiently controlled by this method (Grippa, 2017). The differences between image analysis using segmentation and without segmentation are shown in Fig. 10. However, this method comes with a prominent disadvantage in that the segmentation process consumes a large amount of device memory.

The final result of geographic features recognition using OBIA (see Fig. 10) shows that the spatial patterns were well drawn on large-scale maps (Map1, Map2, and Map3). On Map1 and Map2, clean polygons depicted the coastline, fort Belgica in the north and fort Nassouw in the south with buildings in the surrounding area. On Map3, Fort Hollandia was a landmark with buildings and coastline around. However, the small-scale map (Map4) required a more complex refinement process to remove inappropriate polygons. The image output was also less clear, especially in very small parts such as fortresses and buildings, while the coastline was identified well.

During the operation, the image segmentation by OBIA may be experiencing issues of oversegmentation or undersegmentation. Carleer et al. (2005) argued that oversegmentation is preferable to undersegmentation, as the former can be corrected during classification, contrary to the latter. Furthermore, Csillik (2017) highlighted that oversegmentation, as long as it remains at an admissible level, could be a minor issue in regard to the final classification result. The refinement process answers the final classification tasks that enable the users to manually reclassify the conflicting classes caused by errors during the automatic classification process by deleting or moving the inappropriate feature classes based on their class number attribute. For example, if there are some buildings polygons which should be coastline, this error can be fixed by changing the classes in the polygons' attributes (e. g., 1 for buildings and 2 for coastline) then the inappropriate buildings polygons' classes are changed from 1 to 2 and the classification will be changed as well.

Table 5 shows the comparison of CNN and OBIA based on the experiments. In terms of the statistical measure of performance, CNN is surpassing OBIA both in accuracy (0.9270 vs 0.9049) and kappa statistics (0.8906 vs 0.8731). This indicates that CNN's classification model is better at predicting images than the OBIA model although both methods tend to acquire high value. On the other hand, the OBIA method offers a more convenient way to extract geographic information from historic maps. These two contrary results should be taken into account in determining the tools based on user types and knowledge requirements of computer and data science. For this case study, OBIA is the preferable method that can be implemented by related experts such as archaeologists and historians for heritage landscape mapping of the Banda Islands. Other advantages of the OBIA method is training sample creation. The training sample plotting on the maps simplifies the process of training data generation whereas, in CNN analysis, this step should be done via data augmentation.

4. Conclusions

This study discusses CNN and OBIA as alternatives to geographic features recognition from historic maps for the heritage landscape of the Banda Islands, Indonesia. The importance of heritage landscape exploration from cartographic images rising along with the advancement of digital technology. In particular, the historic maps of the Banda Islands

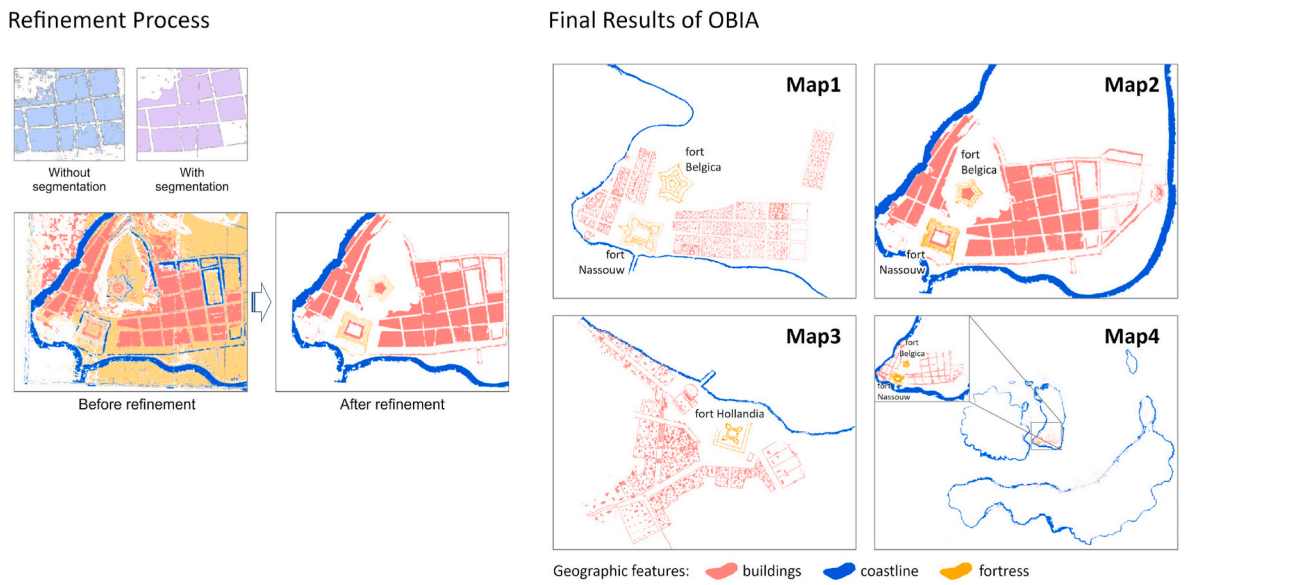

Fig. 10. Image segmentation in OBIA (left) helps to reduce the scope of image classification to produce a clearer output and the refinement in the post-processing allows users to manually reclassify the conflicting geographic features. Geographic features output from OBIA after image segmentation and refinement process (right). Different scales affected the results particularly Map4 less clear than the three other image outputs.

Table 5
Comparison of CNN and OBIA.

Aspect	Sub-aspect	CNN	OBIA
Average performance statistics	Accuracy	0.9270	0.9049
	P-Value [Acc > NIR]	< 2.2e-16	< 2.2e-16
	Kappa	0.8906	0.8731
Technical procedure	Pre-processing	Requires data augmentation for training data	Training samples created directly on the map
	Post-processing/ finalisation	Geographic features filtering is performed in more complicated steps	Easier refinement process
Additional qualitative aspect		Requires a knowledge base in the field of computer and data science, in particular, the coding	More convenient for various non-IT expert users (archaeologists, anthropologists and other researchers)

as the colonised region have very different characteristics and conditions compared to ancient maps of Europe and this becomes a critical aspect which must be considered in the selection of relevant methods for spatial information extraction.

The results indicate that in terms of accuracy, CNN gives a better value than OBIA. Buildings and coastline draw excellent results for CNN analysis, while fortress seemed to be more difficult to be interpreted by the model. On the other hand, OBIA generates a very satisfying result for detailed accuracy of all feature classes (buildings, coastline, and fortress). However, OBIA results depend also on the resolution of the digital images because segments with a large number of pixels are easier to classify. This is very sensitive to the image segmentation process where the suitable segmentation result will give better output quality.

Although CNN's accuracy and kappa statistics are generally better than OBIA's, OBIA's image classification method has other advantages that CNN does not provide, such as georeferentiation, training sample plotting, and image segmentation which can be set up easily through a wizard. This makes OBIA easier to operate by various groups of related researchers. This gives consideration to each method to be selected according to the user's requirements and the historic maps dataset.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

I greatly appreciate the support from Dr Josep Grau-Bové as my supervisor, all lecturers, and involved parties of the Institute for Sustainable Heritage, University College London who provided the learning materials.

References

Albawi, S., Mohammed, T.A., Al-Zawi, S., 2017. Understanding of a convolutional neural network, 2017. In: International Conference on Engineering and Technology (ICET), pp. 1–6.
Andrade, H., Fernandes, B., 2020. Synthesis of satellite-like urban images from historical maps using conditional gan. Geosci. Rem. Sens. Lett. IEEE 1545–598X 1, 1–4.
Archana, J.N., 2016. A review on the image sharpening algorithms using unsharp masking. Int. J. Electr. Syst. Control (6), 8729–8733.
Baily, B.e., 2011. Extracting digital data from the first land utilisation survey of great Britain – methods, issues and potential. Appl. Geogr. 31 (3), 959–968.
Bianchetti, S., 2020. Traces of scientific geography in pliny's Naturalis historia. Shagi/ Steps 6, 10–25.
Blaschke, T., Strobl, J., 2001. What's wrong with pixels? Some recent development interfacing remote sensing and GIS. GeoBIT/GIS 6, 12–17.
Carleer, A., Debeir, O., Wolff, E., 2005. Assessment of very high spatial resolution satellite image segmentations. Photogramm. Eng. Rem. Sens. 71, 1285–1294.
Carman, C.C., Evans, J., 2015. The two earths of Eratosthenes. Isis 106 (1), 1–16.
Chefira, R., Rakrak, S., 2021. Accuracy assessment of applied supervised machine learning models on usual data probability distributions. J. Phys. Conf. 1743.
Christophe, S.e., 2022. Neural map style transfer exploration with GANs. International Journal of Cartography 8 (1), 18–36.
Cléri, I., Deseilligny, M., Vallet, B., 2014. Automatic georeferencing of a heritage of old analog aerial photographs. Aug ISPRS Annals of Photogrammetry, Remote Sensing and Spatial Information Sciences II (3), 33–40.
Csillik, O., 2017. Fast segmentation and classification of very high resolution remote sensing data using SLIC superpixels. Rem. Sens. 9 (243), 1–19.
Ekim, B., Sertel, E., Kabadayl, M.E., 2021. Automatic road extraction from historical maps using deep learning techniques: a regional case study of Turkey in a German world war II map. ISPRS Int. J. Geo-Inf. 492 (10), 1–15.
ESRI. (2022, 07 8). Overview of georeferencing. Retrieved from ArcGIS Pro: https://pro.arcgis.com/en/proapp/2.8/help/data/imagery/overview-of-georeferencing.htm.
ESRI. (2022, 8 07). Segmenting an image. Retrieved from ArcGIS Pro: https://pro.arcgis.com/en/proapp/2.8/help/analysis/image-analyst/segmentation.htm.

Faxon, H.O., 2022. Territorializing spatial data: controlling land through One Map projects in Indonesia and Myanmar. Polit. Geogr. 98, 789–810, 102651.
Flachot, A., Gegenfurtner, K.R., 2021. Color for object recognition: hue and chroma sensitivity in the deep features of convolutional neural networks. Vis. Res. 182, 89–100.
Gao, Y., Mas, J.F., 2008. A comparison of the performance of pixel based and object based classifications over images with various spatial resolutions. Online J. Earth Sci. 2, 27–35.
Gobbi, S., 2019. New tools for the classification and filtering of historical maps. ISPRS Int. J. Geo-Inf. 8 (455), 1–24.
Godfrey, B., Eveleth, H., 2015. An adaptable approach for generating vector features from scanned historical thematic maps using image enhancement and remote sensing techniques in a geographic information system. J. Map Geogr. Libr. 11, 18–36.
Graeff, B., Carosio, A., 2002. Automatic Interpretation of Raster-Based Topographic Maps by Means of Queries. *TS3.10 Spatial Information in Mapping and Cadastral Systems*.
Grippa, T.e., 2017. An open-source semi-automated processing chain for urban object-based classification. Rem. Sens. 9 (4), 1–20.
Guirado, E. e, 2021. Mask R-CNN and OBIA fusion improves the segmentation of scattered vegetation in very high-resolution optical sensors. Sensors 21 (1), 1–17.
Huilgol, P., 2022. Precision vs. Recall – an intuitive guide for every machine learning person. July 19, Retrieved from analyticsvidhya.com. https://www.analyticsvidhya.com/blog/2020/09/precision-recallmachine-learning/.
Iosifescu, I., Hugentobler, M., Hurni, L., 2010. A solution to cartographic challenges of environmental management. Environ. Model. Software 25, 988–999.
Jiang, L.e., 2017. The heritage and cultural values of ancient Chinese maps. J. Geogr. Sci. 27 (12), 1521–1540.
Kaufman, L., 2012. Content-aware automatic photo enhancement. Comput. Graph. Forum 31 (8), 2528–2540.
Kersapati, M.I., 2021. Banda Neira: bandar rempah di Timur nusantara. December 6, Retrieved from Jalur Rempah: http://jalurrempah.kemdikbud.go.id/publikasi/banda-neira-bandar-rempah-ditimur-nusantara.
Knudsen, B.T., 2021. Decolonizing Colonial Heritage: New Agendas, Actors and Practices in and beyond Europe. Routledge, London.
Landis, J.R., Koch, G.G., 1977. The measurement of observer agreement for categorical data. Biometrics 33.1, 159–174.
Lape, P.V., 2002. Historic maps and archaeology as a mean of understanding late precolonial settlement in the Banda islands, Indonesia. Asian Perspect. 41 (1), 43–70.
Lashgari, E., Liang, D., Maoz, U., 2020. Data augmentation for deep-learning-based electroencephalography. J. Neurosci. Methods 346, 108885.
Lecun, Y.e., 1998. Gradient-based learning applied to document recognition. Proc. IEEE 86 (11), 227–2324.
Lee, J., Warner, T.A., 2004. Image classification with a region based approach in high spatial resolution imagery. Int. Arch. Photogram. Rem. Sens. Spatial Inf. Sci. 181–187.
Luft, J., Schiewe, J., 2021. Automatic content-based georeferencing of historical topographic maps. Aug Trans. GIS 1–19.
Mahdianpari, M.e., 2018. Wetland classification using deep convolutional neural network. IEEE Int. Geosci. Remote Sensing 9249–9252.
Maxwell, A., Warner, T., Guillén, L.A., 2021. Accuracy assessment in convolutional neural network-based deep learning remote sensing studies—Part 1: literature review. Rem. Sens. 13, 2450.
McHugh, M.L., 2012. Interrater reliability: the kappa statistic. Biochem. Med. 22 (3), 276–282.
Mello, C.A., Costa, D.C., J, S.T., 2012. Automatic image segmentation of old topographic maps and floor plans. IEEE Int. Conf. Syst. Man Cybern. 132–137.
Miao, Q.e., 2013. Linear feature separation from topographic maps using energy density and the shear transform. IEEE Trans. Image Process. 22 (4), 1548–1558.
Microsoft, 2022. Edit Photos and Videos in Windows. December 2. Retrieved from Microsoft Support: https://support.microsoft.com/en-us/windows/edit-photos-and-videos-in-windows-a3a6e711-1b70-250a-93fa-ef99048a2c86.
Miyoshi, T.e., 2004. Automatic extraction of buildings utilizing geometric features of a scanned topographic map. Proceedings of the 17th International Conference on Pattern Recognition 626–629.
Muhs, S.e., 2016. Automatic delineation of built-up area at urban block level from topographic maps. Comput. Environ. Urban Syst. 58, 71–84.
Papakosta, C., Ioannidis, C., Vassilaki, D., 2012. The use of linear features as ground control information for the georeferencing of old aerial photos. In: Proceedings of FIG Working Week 2012, pp. 1–12.
Parkin, A., 2018. Computing Colour Image Processing. Springer, New York.
Pchelov, E.V., 2019. Historical cartography of Muscovy: symbols and emblems. IOP Conf. Ser. Earth Environ. Sci. 350 (1), 012014.
Pearline, S.A., Kumar, V.S., 2019. Plant species recognition using modified LeNet-5 CNN architecture. Recent Trends in Instrumentation and Control 103–113.
Petitpierre, R., Kaplan, F., Lenardo, I., 2021. Generic semantic segmentation of historical maps. Computational Humanities Research Conference 2989, 228–248.
Rees, R., 1980. Historical links between cartography and art. Geogr. Rev. 70 (1), 61–78.
Skaloš, J., Engstová, B., 2010. Methodology for mapping non-forest wood elements using historic cadastral maps and aerial photographs as a basis for management. J. Environ. Manag. 91 (4), 831–843.
Sutton, A., Yingling, C., 2020. Projections of desire and design in early modern caribbean maps. Hist. J. 63 (4), 789–810.
Tortora, A., Statuto, D., Picuno, P., 2015. Rural landscape planning through spatial modelling and image processing of historical maps. Land Use Pol. 42, 71–82.
Turner, S., 2018. Historic Landscape Characterisation: an Archaeological Approach to Landscape Heritage. Taylor Francis Group, pp. 37–50.
Uça-Avcl, Z.e., 2011. A comparison of pixel-based and object-based classification methods, A case study: istanbul, Turkey. ISPRS Proceedings 1–4.
Uhl, J.H., 2018. Spatialising uncertainty in image segmentation using weakly supervised convolutional neural networks: a case study from historical map processing. The Institution of Engineering and Technology Journal 12 (11), 2084–2091.
UNESCO, 2015. The historic and marine landscape of the Banda islands. January 30, Retrieved from UNESCO World Heritage Convention: https://whc.unesco.org/en/tentativelists/6065/.
Weih, R., Riggan, N., 2010. Object-based classification vs. Pixel-based classification: comparitive importance of multi-resolution imagery. Int. Arch. Photogram. Rem. Sens. Spatial Inf. Sci. XXXVIII-4/C7, 1–6.
Whiteside, T.G., Boggs, G.S., Maier, S.W., 2011. Comparing object-based and pixel-based classifications for mapping savannas. Int. J. Appl. Earth Obs. Geoinf. 13 (6), 884–893.
Widmann, W., 2022. Cohen's kappa: what it is, when to use it, and how to avoid its pitfalls. August 2, Retrieved from The Newstack: https://thenewstack.io/cohens-kappa-what-it-is-when-to-use-it-and-how-to-avoid-its-pitfalls/.
Wong, S.C., 2016. Understanding Data Augmentation for Classification: when to Warp? *2016 International Conference on Digital Image Computing: Techniques and Applications*. DICTA), pp. 1–6.
Yan, Z., 2016. Automatic photo adjustment using deep neural networks. ACM Trans. Graph. 35, 1–15. https://doi.org/10.1145/2790296.
Yu, e. a., 2017. CNN-GRNN for Image Sharpness Assessment. *Conference*. Asian Conference on Computer Vision, pp. 50–61.
Zatelli, P.e., 2019. Object-based image analysis for historic maps classification. Int. Arch. Photogram. Rem. Sens. Spatial Inf. Sci. XLII-4/W14, 247–254.
Zeya, L.T., 2022. Essential things you need to know about F1-score. July 19, Retrieved from towardsdatascience.com. https://towardsdatascience.com/essential-things-you-need-to-know-about-f1score-dbd973bf1a3#:~:text=1.,Introduction,competing%20metrics%20%E2%80%94%20precision%20and%20recall.
Zhang, C.e., 2019. Pedestrian detection based on improved LeNet-5 convolutional neural network. J. Algorithm Comput. Technol. 13, 174830261987360.