MONITORING THREATENED IRISH HABITATS USING MULTI-TEMPORAL MULTI-SPECTRAL AERIAL IMAGERY AND CONVOLUTIONAL NEURAL NETWORKS

Sara Perez-Carabaza, Oisin Boydell[*] *Jerome O'Connell*[†]

Ireland's Centre for Applied AI
University College Dublin, Ireland

ProvEye
Dublin, Ireland

ABSTRACT

The monitoring of threatened habitats is a key objective of European environmental policy. Due to the high cost of current field-based habitat mapping techniques there is a strong research interest in proposing solutions that reduce the cost of habitat monitoring through increasing their level of automation. Our work is motivated by the opportunities that recent advances in machine learning and Unmanned Aerial Vehicles (UAVs) offer to the habitat monitoring problem. In this paper, a deep learning based solution is proposed to classify four priority Irish habitats types present in the Maharees (Ireland) using UAV aerial imagery. The proposed method employs Convolutional Neural Networks (CNNs) to classify multi-temporal multi-spectral images of the study area corresponding to three different dates in 2020, obtaining an overall classification accuracy of 93%. A comparison of the proposed method with a multi-spectral 2D-CNN model demonstrates the advantage of including temporal information enabled by the proposed multi-temporal multi-spectral CNN model.

***Index Terms*—** Habitat mapping, Convolutional neural networks, Multi-temporal imagery, Aerial imagery

1. INTRODUCTION

The EU Habitat Directive [4], which addresses the conservation of natural habitats and of wild fauna and flora, directs EU member states to take measures in order to maintain the favourable conservation status of threatened habitats. Namely, the EU requires member states to periodically produce maps for change detection and conservation status assessments of priority habitat types listed in Annex I of the EU Habitats Directive [4]. Currently, Ireland reports the conservation status of its threatened habitats based on ecological field data. The current field-based mapping and assessment methodology requires significant time and financial resources. This work is motivated by the potential offered through the rapid advances in Unmanned Aerial Vehicles (UAVs) along with machine learning techniques.

Several remote sensing techniques have been recently proposed to deal with the need for cost-effective habitat mapping tools. Among them, UAV-based methods are emerging as a powerful tool for habitat mapping thanks to their low flight altitude relative to other remote sensing techniques like aircraft or satellites, as well as their possibility of flying over areas that are difficult to access. Additionally, the flexible scheduling of flights and their ability to provide visual imagery at a high resolution and biologically distinguishable level are also an advantage [5]. There is also an active research field in the application of machine learning techniques for processing the large amounts of data provided by aerial or satellite imagery. For instance, [5] proposes the use of the random forest algorithm for the classification of several Annex I habitats located in Northern Portugal using aerial imagery. In contrast, our work aims to benefit from advances that deep learning techniques have shown in image classification problems to classify aerial imagery from Annex I habitats in the Maharees, Ireland. Example images of the considered Annex I habitats are shown in Fig.1. Their spectral similarity makes their classification a highly challenging problem.

Over the last decade, deep learning has achieved breakthrough results across many varied applications, in particular in dealing with unstructured data such as images, audio or text. In computer vision tasks, Convolutional Neural Networks (CNNs) stand out due to their impressive results. Contrary to traditional machine models which are dependent on engineered features, CNNs have the advantage of being data-driven, which empowers their ability to automatically learn contextual features from raw input images, making them highly effective for large-scale image recognition and semantic segmentation tasks [6]. The outstanding results that CNNs have achieved in other domains have motivated the remote sensing community to apply them in image classification problems dealing with satellite or aerial imagery. However, the application of deep learning to remotely sensed data involves several challenges such as the high cost and expert knowledge required for obtaining la-

[*]This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No. 847402.

[†]The authors wish to thank the EPA funded iHabiMap project [3] for providing the data used in this work.

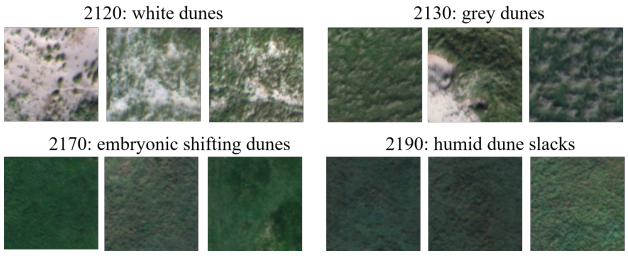

Fig. 1. Sample images of the four Annex I habitats.

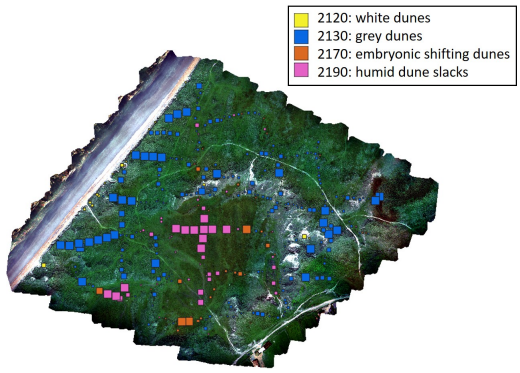

Fig. 2. A true color composite of the Maharees mosaic collected in July 2020 with habitat labels based on field data.

belled training data. Also, as most of the deep learning models are only designed to work with three channel RGB images, there is a need of adapting and proposing deep learning-based models that can utilize multispectral bands. Remote sensing images are georeferenced and may have high spectral dimensionality (multi-spectral or hyperspectral images) or multi-temporal information (which is especially relevant in remote sensing tasks dealing with crops or vegetation). Furthermore, while RGB images tend to have well-defined scene context, in several remote sensing problems including this work, the images present a potentially unlimited continuous space, where boundaries between classes are transitional as can be observed from the sample images presented in Fig. 1, making it a challenging image classification problem. Despite all these challenges, in recent years deep learning approaches applied to remotely sensed image classification have achieved significant breakthroughs, offering exciting opportunities for research [6].

In the literature we can find some recent deep learning classifiers that exploit the spatial, spectral and temporal information associated with remotely sensed imagery. In [7] a 3D-CNN model is proposed for learning three dimensional filters along the temporal and two spatial dimensions of a time series of satellite images in order to classify four types of crops. And [1] propose a model with two parallel branches in order to classify several land uses and crop types; where a recurrent neural network is used for learning the temporal features from a pixel-wise time series and a set of 2D convolutional layers for learning the spatial features of satellite imagery. In this work we propose a novel CNN-based architecture for leveraging the spatial and temporal information of aerial imagery of Annex I habitats in the Maharees, Ireland.

This work is organized as follows. Section 2 describes the data collected from the Maharees study site. Section 3 presents the proposed deep learning-based model to classify four habitats present in the Maharees. Section 4 evaluates the performance of the proposed method and compares it with a multi-spectral 2D-CNN model. Lastly, Section 5 summarises the main conclusions and describes the future research direction.

2. MAHAREES HABITAT DATA

The study area in the Maharees is a tombolo located in the Dingle Peninsula at south-west of Ireland. The sand dunes create a unique ecosystem with several threatened habitats listed in Annex I of the EU Habitats Directive [4]. The four habitats considered for the classification and their respective Annex I codes are: white dunes (2120), grey dunes (2130), embryonic shifting dunes (2170) and dune slacks (2190). Sample images of the habitats are displayed in Fig. 1.

This study considers three multi-spectral mosaics of the Maharees study site built from the aerial imagery collected by the multi-spectral sensor on the 26th of May, 28th of July and 8th of October 2020. The multi-spectral images in UTM/WGS84 projection cover a 0.7 km^2 area with a spatial resolution of 0.05 m per pixel and contain the reflectance values of five spectral bands: blue, green, red, red-edge and Near Infra Red (NIR). The multi-spectral mosaics were obtained by the combination of the image tiles collected by the multi-spectral sensor using ProvEye proprietary software [8]. In summary, this software uses back and forward projection in combination with automated algorithms for feature detection of control points during the mosaicking (fusion) process. The software uses upward-facing irradiance values to normalise for inflight illumination by creating a correction coefficient. Surface reflectance is derived using calibration panels of known albedo taken before and after each flight. This ensures that radiometric variation between mosaics is minimised while still maintaining the spectral integrity of the data. A true color representation of the multi-spectral mosaic corresponding to July 2020 is displayed in Fig. 2. This work also considers a a vector database in shapefile format with the information about the types of habitats provided by a team of ecologists that labelled multiple sample points across the study site with in-situ measurements [3]. Each on-the-spot habitat label has associated a square polygon that delimits the extent of the habitat label provided by the expert. Fig. 2 represents a true color representation of the multi-spectral

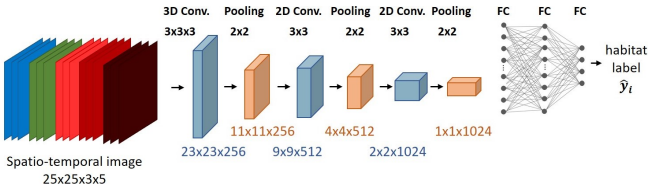

Fig. 3. The CNN architecture proposed for habitat classification from multi-temporal multi-spectral images.

mosaic in July 2020 with the labels represented in different colors according to the four habitat types. The heterogeneous distribution of the habitat types can be seen in Fig. 2, where habitats 2130 and 2190 cover 67% and 23% of the total labelled area, whereas the minority habitats of 2120 and 2170 only correspond to 2% and 8%, respectively.

In short, the Maharees dataset consists of three multi-spectral mosaics corresponding to the same area in Maharees (Ireland) obtained at three different dates during May, July and October of 2020, and a shapefile with information about the on-the-spot habitat information. The Maharees dataset was provided by the iHabiMap project [3].

3. DEEP LEARNING APPROACH FOR HABITAT CLASSIFICATION

This section describes the proposed habitat mapping model that classifies multi-temporal multi-spectral images by means of a novel convolutional neural network architecture.

First, in order to extract the multi-temporal multi-spectral images to train the deep learning model from the Maharees mosaics a patch-based extraction method was followed. Patches are sampled at random locations within the labelled polygons with the intention of generating more samples for training the model. From each polygon p, n_p patches of dimensions $w \times h$ are extracted by randomly sampling the same area from the three multi-spectral mosaics of the Maharees corresponding to May, July and October 2020. The number of patches extracted from each polygon p is equal to $c_j \cdot (w_p \cdot h_p)/(w \cdot h)$, where w_p and h_p are the width and height of the polygon, and c_j a constant value associated to habitat j. Extracting in this way, a higher number of patches n_p from larger polygons. The resulting sampled patches from the three multi-spectral mosaics are concatenated along a temporal axis, resulting in a multi-temporal multi-spectral image of size $w \times h \times 3 \times 5$.

The proposed CNN classifier takes as input the multi-temporal multi-spectral image and predicts the most probable habitat label $\hat{y}_i$ among the four considered habitat types. The proposed architecture, which is sketched in Fig. 3, is composed of one 3D convolutional layer, two 2D convolutional layers, three pooling layers and three fully connected layers. First, the multi-temporal multi-spectral image of size $w \times h \times 3 \times 5$ is fed to a 3D convolutional layer that applies $3 \times 3 \times 3$ kernels with stride of 1 and no padding, where the two first dimensions of the 3D kernel correspond to the two spatial dimensions and the third one to the temporal dimension. The output of the 3D layer is then fed to a pooling layer that reduces by half the spatial dimensions by applying a 2×2 average pooling strategy. Next, a pair of 2D convolutional layers with kernels of size 3×3 and pooling layers with 2×2 average pooling are applied, followed by three Fully Connected (FC) layers with $1024, 512$ and 4 neurons respectively. All the convolutional layers consider no padding and a stride of 1. And all the layers have associated a rectified linear unit (ReLU) with the exception of the last fully connected layer which considers the softmax activation function that returns a probability distribution of the four habitat types. The categorical cross entropy loss function is optimized using the Adagrad optimizer and considering a weighting of the loss function based on the number of samples of each class [2].

Table 1. Confusion matrix of the results obtained over the Maharees test dataset.

Reference habitats \ Predicted habitats	2120	2130	2170	2190	Total	Recall
2120	146	5	0	0	151	0.97
2130	12	3723	38	115	3918	0.95
2170	0	44	394	68	506	0.78
2190	0	37	77	1071	1185	0.90
Precision	0.92	0.98	0.73	0.85		

4. EXPERIMENTAL RESULTS AND DISCUSSION

For evaluation purposes, the Maharees dataset has been split at polygon level in training and test sets following a stratified random sampling, where 25% of the polygons are used testing and the remaining 75% for training and validation. From the training and test sets of polygons the multi-temporal multi-spectral images are extracted following the batch extraction strategy defined in Sec. 3, where the patch width w and height h are both set to 25. With the purpose of mitigating the imbalance of the training dataset, the constant c_j that controls the number of images extracted from a polygon p labelled as class j is set to $c_{1:4} = \{4, 1, 3, 2\}$, extracting higher number of images from the least frequent habitats. No oversampling/undersampling is applied to the test dataset, where $c_j = 2$ for the four classes.

The following metrics summarize the performance of the proposed CNN classifier over the test dataset; overall accuracy: 0.93 and macro-averaged F1-score: 0.89. Hence, 93% of the images from the test dataset were correctly identified by the proposed model. The confusion matrix obtained by the comparison of the reference ground truth labels and predicted labels for the test dataset is shown in Table 1. Each element $e_{r,c}$ of the confusion matrix corresponds to the number of images predicted as habitat type r known to be habitat

Table 2. Comparison of the CNNs models.

Model	Convolution	MT	MS	Samples	Accuracy
This work	3D & 2D	✓	✓	May-July-Oct.	0.93
2D-CNN	2D		✓	May	0.87
				July	0.86
				October	0.84

type c. Moreover, the recall (user's accuracy) and precision (producer's accuracy) values per class are placed respectively at the last column and row of Table 1. All the habitats were correctly classified with percentage equal or higher than 78%.

Furthermore, with the purpose of analysing the benefits of including temporal information for the habitat classification problem, the proposed CNN model is compared with a 2D-CNN model that considers the multi-spectral reflectance values but without any temporal information. Table 2 summaries the main characteristics of the two models: the model names, the type of convolutions considered by the models ($Convolution$), whether the models consider multi-temporal (MT) and multi-spectral (MS) information, the multi-spectral mosaics used for the training and testing samples ($Samples$) and the overall accuracy obtained over the test dataset ($Accuracy$). In order to implement the 2D-CNN model the following changes are applied to the architecture of the proposed CNN model (sketched in Fig. 3): the first layer of the model is fed with multi-spectral images and applies 2D kernels of size 3×3 (instead of the 3D filters used by the proposed model). For training the 2D-CNN model the multi-temporal multi-spectral images are adapted from 4D tensor of size $w \times h \times t \times s$ to 3D tensors of size $w \times h \times s$. This is done by splitting each MT MS image into three MS images that correspond to the spectral values of the same area and habitat type at different dates of 2020. Hence, both models are trained considering images corresponding to the same areas and habitat types, but while the images used by the proposed model correspond to the spectral values of the terrain at three different dates, the images used by the 2D-CNN model correspond to the spectral values at a certain date (either May, July or October). Table 2 shows the overall accuracy values obtained by the proposed CNN model ($acc. = 0.93$) and the 2D-CNN model when considering the MS images obtained from the May mosaic ($acc. = 0.87$), the July mosaic ($acc. = 0.86$) and the October mosaic ($acc. = 0.84$). The better results obtained by the proposed CNN model in comparison with the ones obtained by the 2D-CNN model show the benefits of incorporating temporal information through the proposed CNN-based model.

5. CONCLUSIONS AND FUTURE WORK

This work proposes a deep learning approach using convolutional neural networks to classify Annex I habitat types from multi-temporal multi-spectral aerial imagery. The proposed model achieves an accuracy of 93%, obtaining better predictions than a multi-spectral 2D-CNN model that shows the advantage of including multi-temporal information.

As future work, we plan to incorporate digital elevation information and expect to increase the classification performance of the models leveraging the dependence of the Maharees' habitats on the terrain information. Moreover, using the new aerial imagery from Maharees and different Irish habitats data that is currently being gathered by the EPA-funded iHabiMap project [3], we plan to i) test the proposed model with imagery from Maharees in 2021 to assess the model's performance not only over unseen areas but also with data corresponding to a different year than the one used for training and ii) test similar deep learning models on more types of Annex I habitats located in different regions of Ireland.

6. REFERENCES

[1] Benedetti, P., Ienco, D., Gaetano, R., Ose, K., Pensa, R., & Dupuy, S. (2018). M3fusion: A deep learning architecture for multi-Scale/Modal/Temporal satellite data fusion. arXiv preprint arXiv:1803.01945.

[2] Cui, Y., Jia, M., Lin, T. Y., Song, Y., & Belongie, S. (2019). Class-balanced loss based on effective number of samples. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (pp. 9268-9277).

[3] Cruz, C., O'Connell, J., McGuinness, K., Martin, J., Perrin, P., & Connolly, J. (2019). iHabiMap: habitat mapping, monitoring and assessment using high-resolution imagery. IEOS, NUIG, Ireland.

[4] Directive, H. (1992). Council Directive 92/43/EEC of 21 May 1992 on the conservation of natural habitats and of wild fauna and flora. Official Journal of the European Union, 206, 7-50

[5] Gonçalves, J., Henriques, R., Alves, P., ... & Honrado, J. (2016). Evaluating an unmanned aerial vehicle-based approach for assessing habitat extent and condition in fine-scale early successional mountain mosaics. Applied Vegetation Science, 19(1), 132-146.

[6] Li, Y., Zhang, H., Xue, X., Jiang, Y., & Shen, Q. (2018). Deep learning for remote sensing image classification: A survey. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 8(6), e1264

[7] Ji, S., Zhang, C., Xu, A., Shi, Y., & Duan, Y. (2018). 3D convolutional neural networks for crop classification with multi-temporal remote sensing images. Remote Sensing, 10(1), 75.

[8] ProvEye 2021. ProvUAV v1.3, ProvEye Limited, Ireland. www.proveye.ie