Development of high-resolution maps of vegetation cover to support land planning and grazing management in fire prone landscapes

Bianka Trenčanová
bianka.trencanova@tecnico.ulisboa.pt
Instituto Superior Técnico, Universidade de Lisboa, Portugal
January 2021

Abstract **– The focus of this thesis is to develop a classifier of shrub vegetation cover. Shrubs are a key vegetation type in dry Mediterranean climates, that is associated with an increased risk of fire. The classifier will be further used for sustainable land planning and grazing management for fire prevention. Two main objectives are 1.) to design a new dataset from an unmanned aerial vehicle (UAV) imagery using ordinary RGB channels and 2.) to develop a method to increase the accuracy of a convolutional neural network (CNN) with a U-Net architecture to detect shrubs in a complex heterogeneous forest environment within a study farm in Portugal. The tested methods and their feasibility for this particular task are data augmentation, tiling, rescaling, dataset balancing and hyperparameter tuning (namely the number of filters, dropout rate and batch size). The biggest improvements were recorded with data augmentation, tiling and rescaling practices. The developed classification model achieves an average F1 score of 0.72 on three separate test sets even though it is trained on a relatively small dataset with some degree of inaccurate labels. It takes around four hours to train the model. The major challenges identified in this work were precise manual image annotation, small sample size, time and memory limits of used tools, and high intra-class and low inter-class variance of the target vegetation class. The main contributions of this study are evaluating the performance of the state-of-the-art CNN for mapping fine-grained land cover patterns from RGB remote sensing data and proposing a method to improve the outputs.**

Keywords: **U-Net, convolutional neural network (CNN), shrub detection, heterogeneous land cover mapping, UAV imagery, Mediterranean forest**

1. Introduction

Food production is one of the major contributors to degradation of the environment. This sector accounts for approximately 26% (13.598 Gt CO2-eq/yr) of global greenhouse gas emissions (GHG), out of which one half (6.93 Gt CO2-eq/yr)[1] comes from crop production and land use, linked to turning natural ecosystems such as forests and grasslands, that act as carbon "sinks", into cropland and pastures, that release additional carbon dioxide (CO2)[1]. However, agricultural systems can become carbon sinks, if we change our food production systems and learn to harness ecosystem services. Proper sustainable management practices that are aligned with Earth system processes can lower our impact on the environemnt. One such example is cattle farming. Traditional livestock systems play a role in biodiversity conservation, climate adaptation, and socioecological resilience at regional and local scales [1]. Ecological processes, such as nutrient cycling, soil fertilization, maintenance of genetic diversity and regulation of vegetation growth, once supported by wild large herbivores [2], are now sustained by free-range livestock in areas where wild large herbivores are scarce or no longer present. However, due to strong socio-economic drivers resulting into rural-urban migration, an extensive abandonment of agricultural land is becoming problematic. The absence of large herbivores and the withdrawal from human activities increase fuel loads and promotes homogenization of vegetation in the affected areas, mainly the growth of shrubs, that are prone to fires and therefore especially dangerous in the dry climate of the Mediterranean Basin. Active re-introduction of herbivores into fire

[1] https://ourworldindata.org/food-ghg-emissions#:~:text=They%20are%20the%20direct%20emissions,for%2024%25%20of%20food%20emissions.

prone regions could serve as an environmentally sustainable and time and cost-effective method for wildfire prevention. Prescribed (or targeted) grazing is a silvopastoral practice that promotes heterogeneous landscapes, controls shrub encroachment and is officially considered as a wildfire prevention tool [3]. However, such interventions require thorough land planning, preventive management and regular monitoring for which a detailed land cover mapping is essential. Remote sensing is a primary source of data for vegetation mapping and thanks to continual developments in geo-information technologies this field is gradually becoming more universal. Limitations that satellite-based systems face, such as insufficient spatial, spectral and temporal resolutions, cloud cover or high cost of data acquisition, are resolved with the emergence of a new remote sensing aerial platform – unmanned aerial vehicles (UAVs). UAVs have very high spatial resolutions thanks to their low speed and flight altitude, they are cheaper, more flexible in obtaining data from target areas that are often difficult to reach, they minimize disturbances of inspected areas and provide real-time monitoring [4]. Acquired data is often used in combination with Artificial Neural Networks (ANNs), that have the capacity to speed up evaluation process of the input information even over large datasets. That is why these methods are becoming a fundamental tool in numerous fields from wildlife conservation and management and various agricultural applications to fire detection.

1.1 Case study: Quinta da França

This thesis uses a case study farm, Quinta da França, that integrates agricultural and forest land uses. The farm's management is guided by sustainability principles and focuses on promoting environmental services provided by agroforestry activities and sustainable forest management[2].

It maximizes synergies between forest production and agricultural production, that enhances multiple environmental services. In a big emission offset program in collaboration with EDP that ran from 2006 to 2012, the farm's forest area captured 7000 tons of CO_2/year, demonstrating that the provision of environmental services (in this case natural agroforestry carbon sink) can be a competitive agricultural market product.

The management of the forest, that experienced two big fire events in 80's and in 1996, is focused on the reduction of fire risk, increase of carbon sequestration, and biodiversity conservation. Vegetation cover and level of development is heterogenous. Trees are dominant and often accompanied by dense understory, which increases their vulnerability to fire spread and requires management measures to reduce that risk, namely the regular removal of shrub cover. The use of livestock for biomass regulation is now being implemented through targeted grazing and its impact is under investigation also within the forest site.

1.2 Objectives

The aim of this thesis is to develop a method for high resolution mapping of land cover in a forest area with heterogenous land cover composition, with a focus on fire prone shrub vegetation. A classifier of the target vegetation type (i.e. shrubs) in the areas susceptible to fire will be created based on exemplary data from UAV imagery, that can recognize the corresponding patterns in new images. Maps of vegetation cover will then serve as a foundation for better informed landscape planning and grazing management and for research of innovative ways of integrating livestock production, biodiversity conservation and fire prevention in fire prone landscapes in the Mediterranean regions.

The main objectives and contributions of this work are:

- Classification of fire-prone vegetation type (shrubs) from natural color UAV images – creating manually labeled dataset for training, validation and testing, using semantic segmentation;
- Using supervised learning approach to train a CNN (U-net) to automatically detect the key vegetation type in new images;
- Developing a method to increase the detection accuracy;
- Evaluating the feasibility and performance of the detection of an irregular shrub cover in a complex heterogeneous landscape.

2. Related work

Semantic segmentation is the most interesting type of image analysis for the land cover classification tasks, able to learn also spatial configuration of labels and class-specific structures [5]. The detection can be either of one specific class [6] or multiple classes at the same time [7]. Two big remaining challenges of the existing methods are intra-class inconsistency and inter-class indistinction [8].

[2] https://www.terraprima.pt/en/sobre-nos/

One of the main research topics nowadays is how to provide pixel-level high-resolution segmentation. Two approaches try to address this problem – 1.) using dilated (atrous) convolution and 2.) connecting pooling and un-pooling layers, e.g. DeconvNet, SegNet or U-Net [9]. Among the first networks focusing on semantic segmentation was a fully convolutional network (FCN) [10]. It uses traditional CNN as a feature extractor but replaces the fully connected layers with up-convolutions, producing spatial feature maps instead of classification scores, that are further up-sampled to a dense pixel-wise output. Improvement of the FCN is already mentioned SegNet [11], that consists of an encoder part, extracting spatial features, and a decoder part, up-sampling the feature maps. Similar to FCN and SegNet is a fully convolutional semantic segmentation network U-Net [12], that will be discussed further in the next section. SegNet and U-Net are able to densely label every pixel at the original resolution of the image thanks to their down-sample-up-sample architecture. High-level representations are learnt via convolutions and then up-sampled back to the original resolution via deconvolution. These nets are computationally efficient and able to learn spatial dependencies among classes. Their drawback is low geometric accuracy [13]. Other approaches are presented by [14] and their multi-scale FCN or [15] DeepLab with atrous convolutions for the semantic segmentation.

The research on how state-of-the-art classification tools perform in complex land cover mapping tasks is generally scarce [16]. Shrubs class is a very general and heterogeneous group of vegetation with individuals of variable shapes, sizes, and distribution patterns, forming irregular and complex clusters of individuals [17]. High intra-class and low inter-class variance is a challenge causing difficulties to distinguish them from their surroundings [18] or other vegetation classes. [16] used multispectral data, containing more complementary information, as a way to alleviate the problem of classification of spectrally similar vegetation types. They also found InceptionResNetV2 as the most efficient state-of-the-art convnet (compared to DenseNet121, InceptionV3, VGG16, VGG19, Xception and ResNet50) for classifying complex multispectral remote sensing wetlands scenes, when it reached an F1 score of 93%. In their pursuit of maximizing the distinction between the target vegetation type (weeds) and the surroundings, [18] proposed to consider phenological stage highlighting the differences in the vegetation appearance as the most promising approach, but also performing the survey at lower flight altitudes (below 100m [19]) or using higher resolution sensor to obtain more detail.

A study with similar objective to this work – shrubs detection is [17]. Objects of interest are Ziziphus lotus shrubs, however, it is surrounded by bare soil with sparse vegetation unlike shrubs in my case, that are located irregularly in a complex heterogeneous landscape. After combining GoogleLeNet with data augmentation, transfer learning (fine tuning) and pre-processing, F1 score of 97% was achieved. Pre-processing techniques improving the detection performance the most were background elimination and long-edge detection, and only random flipping, scaling, cropping and brightness were used for data augmentation.

3. Materials and methods

3.1 Study area

Quinta da França is a 500 ha property in a mild climate of Castelo Branco District. Summer is a critical season regarding the risk of forest fires, with the average temperature reaching 22.2°C and only 10 mm of rainfall in August[3].

Figure 1 shows the division of the farm into three main zones:

1. Quinta de Cima: Northwest area with beef production and grazing pastures.
2. Quinta de Baixo: South area with cattle and sheep production and pastures.
3. Serra: Northeast area with oak forest.

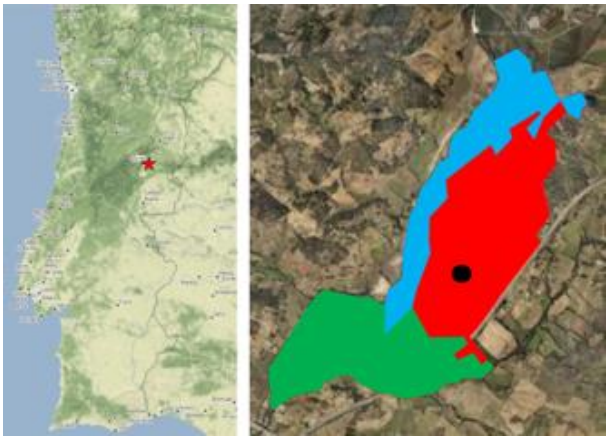

Figure 1 Left: Location of Quinta da França in Portugal (Source: QGIS). Right: Zones of the farm – Quinta de cima

[3] https://en.climate-data.org/europe/portugal/covilha/covilha-6944/

(blue), Quinta de Baixo (green) and Serra (red). The black point depicts the location of the test data used in this thesis (Source: Terraprima -Sociedade Agrícola Lda., 2012)

The forest in Serra, previously closed for animals, was divided into two parcels in January 2018: a southern grazing parcel, to test the effect of cattle presence on vegetation structure (grazing, trampling, etc.), and a northern parcel without grazing. In June 2018 the grazing parcel was opened for cattle. Mechanized removal of shrubs is maintained at both parcels.

3.2 Data description

The images were acquired by hexacopter with two cameras: VIS GITUP2 camera with RGB filter (370 – 680 nm) and 170° lens (fish-eye) and NIR Mapir Survey2 NDVI camera (Red: 660 nm, NIR: 850nm), with 90° lens. 16MP ((4608 x 3456) px) sensor Sony Exmor IMX206 (Bayer RGB) was used. The flight altitude relative to the take-off point was 120m, velocity 5m/s and photos were taken every 5s. The drone was assembled by Terraprima.

The set was composed of 21 (4608 x 3456) px original TIFF images in RGB, which were captured for the same test area during a single flight, that took place in August 2019. Some of the drawbacks of these images were fisheye and motion blur, which caused distortion and made the annotation more challenging, especially in peripheral areas of the images. This thesis uses ordinary RGB images. Limited number of spectral channels makes the presented method more convenient for use in combination with most aerial imaging systems, including off-the-shelf UAVs and wider range of data.

The images were converted into PNG format and sliced into smaller square-shaped tiles with dimensions (800 x 800) px, corresponding to approximately (50 x 50) m patches of land. The tile size was chosen based on the size of the objects of interest and the amount of context. The code can be found at https://github.com/aggiungi1procione/Thesis---Shrub-detection-with-U-Net. In total, 630 tiles were generated from the original images and 13 were selected for labelling. All 13 tiles came from the same image and were chosen as the best representation of all different land-cover configuration present in the image. Four land cover classes were identified: shrubs, trees, shadows and rocks. Labelbox[4] was used for labelling and managing the training data. Figure 2 is an example of the dense pixel-level semantic segmentation maps. The binary masks of labelled tiles were download for all the classes into separate class folders.

The central issue of labelling that could significantly impact the classification results is faulty labelling. Challenging was mainly visual interpretation and distinguishing between shrubs and other vegetation species, incoherent labelling of shadows that coexisted within other classes, interpretation of border parts of classes and border regions of skewed images.

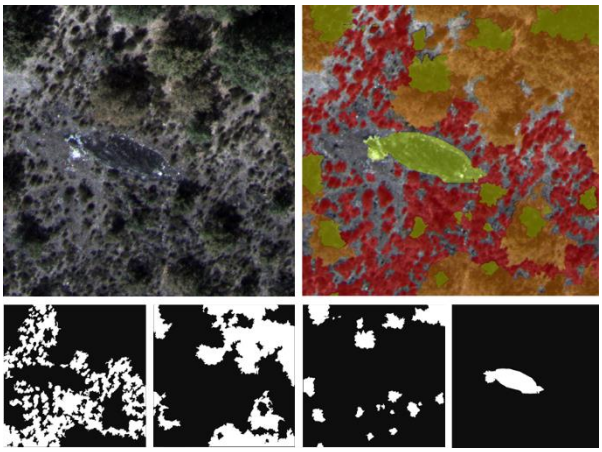

Figure 2 An example of a labelled tile and its binary masks. Up-left: original image tile, up-right: labeled image tile (red - shrubs, orange - trees, yellow - shadows, light yellow - rocks). Bottom (from left): binary mask of shrubs, trees, shadows and rocks

Subsequently, five sub-datasets with different sized patches were sliced from these tiles: A with 832 (100 x 100) px patches, B with 208 (200 x 200) px patches, C with 117 (300 x 300) px patches, D with 52 (400 x 400) px patches and E with 52 (500 x 500) px patches.

Data augmentation was then applied to each sub-dataset, generating three sets per each with the size of around 800, 1600 and 3800 samples. Techniques used for augmentation were random rotations, skews, flips, random brightness, elastic distortions and shears from the Augmentor[5] library. Figure 3 summarizes all the sets that were created.

[4] https://labelbox.com/

[5] Marcus D Bloice, Peter M Roth, Andreas Holzinger, Biomedical image augmentation using Augmentor, Bioinformatics, https://doi.org/10.1093/bioinformatics/btz259

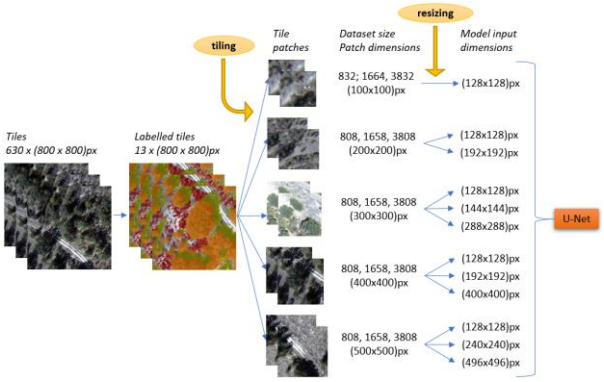

Figure 3 Flowchart of the development of the final sub-datasets for experiments

Test datasets were created in the same way from two additional sets of data that were obtained later: winter images from December 2019 and summer images from August 2020. Four test sets were created in total:

1. Using one (800 x 800) px tile from the same image from which training tiles were taken
2. Using two (800 x 800) px tiles from other images that were taken during the same flight, as the previous image
3. Using two (800 x 800) px tiles from one image from the new summer set (August 2020)
4. Using two (800 x 800) px tiles from one image from the new winter set (December 2019)

The reasoning of the selection was to see the performance on highly similar data (1 and 2), on seasonally similar data (3) and on highly distinct data, taken during different phenological stage (4).

3.3 Model

A U-Net model[6] (hereinafter the TGS U-Net) was used as a basis for the work. The model extracts features with convolutional layers in the encoding part and restores the original size of the image in the decoding part. The TGS U-Net uses the input image size (128 x 128 x 3) and gradually reduces its dimensions while increasing the depth (from (128 x 128 x 3) to (8 x 8 x 256)), and then gradually increases the dimensions and decreasing the depth (from (8 x 8 x 256) to (128 x 128 x 1)).

The main building block of the TGS U-Net consists of two consecutive 2D convolutional layers with batch normalization and ReLU. Batch normalization was stated by the author[7] to significantly improve the training. The number of filters starts at 16 and is doubled at every convolution step. There are four such blocks in the encoder side, each followed by max pooling layer, that halves the image dimensions, and a dropout layer. The fifth convolutional block forms a bottleneck with the maximum depth and minimum spatial dimensions after which comes the decoder side, with four symmetrical deconvolution layers concatenated with the feature maps from the encoder side. After comes a dropout layer and the convolutional block, which helps the model to assemble a more precise output. The number of filters is halved at each step, while the resolution is doubled. Ultimately, the output of a binary classification is sigmoid, which assigns each pixel a probability of belonging to the target class. The model is trained with Adam optimizer with a learning rate of 1e-5. Predictions are compared to labels with binary cross entropy loss function. Early stopping is implemented if the validation loss doesn't improve for 10 consecutive epochs to prevent overfitting. Learning rate is reduced when the validation loss doesn't improve for five consecutive epochs. For each pixel the probability of belonging to the target class is calculated, with the threshold of 0.5. The dataset is split into training and validation set with the ratio 9:1. The validation set is never used in the training process, it is only used to evaluate the model's performance. There are 50 epochs with batch size of 32.

A cloud service Google Colab was used for training and evaluating the model. The deep learning methods were implemented using Keras[8] with TensorFlow[9] backend. With the memory limit of 12GB and the time limit of 12h that comes with the free version of the service, this thesis also aims to explore the set ups with a reasonable tradeoff between working within these limits and yielding good results. This increases the usability and practicality for future students with limited access to advanced virtual machines that would like to build upon or further extend this thesis.

4. Results and discussion

This section presents methods used in order to improve the detection results. The main evaluation metric is F1 score (Equation (1)), is a class-specific measure of segmentation accuracy, suitable for unbalanced datasets like the ones used in this thesis.

[6] https://github.com/hlamba28/UNET-TGS

[7] https://towardsdatascience.com/understanding-semantic-segmentation-with-unet-6be4f42d4b47

[8] https://keras.io/

[9] https://www.tensorflow.org

$$F1\ score = 2\ x\ \frac{Precision\ x\ Recall}{Precision + Recall} \quad (1)$$

where precision:

$$Precision = \frac{TP}{TP + FP} \quad (2)$$

and recall:

$$Recall = \frac{TP}{TP + FN} \quad (3)$$

where TP means true positives, FP false positives and FN false negatives.

4.1 Sub-dataset A

The TGS U-Net was trained on the 832 sample datasets of all four classes separately and for the target class – shrubs also on the 1664 and 3832 sample datasets. The model input dimensions were (128 x 128) px.

The shrub class demonstrated that the performance rises with the growing size of dataset, with F1 = 0.31 for the smallest, and F1 = 0.68 for the biggest dataset. This is not a surprise – the model has more examples to learn from and the data augmentation aids in encoding more invariance, making the learning more robust. However, the tree class significantly outperformed shrubs even with the non-augmented 832 sample dataset (F1 = 0.83). The reason for that can be that trees were a much more balanced class without any artificial adjustments to the data, with 48.58% pixel representation across the dataset, while shrubs only account for 20.99%, but more importantly trees seem to be simply more distinct to other classes and suffer less from high intra- and low inter-class variance. In general, sub-dataset A performed poorly in comparison to other sub-datasets. Small patches probably failed to capture enough of spatial detail and fine-grained boundaries between the class and the background. There are assumably too many patches consisting of only a part of one object, not capturing enough of the context. Moreover, contrary to the other sub-datasets the patches here are up-scaled (from 100 to 128 px), which can bring more blur into them, making it even more difficult to see relevant patterns. On the top of that, scales larger than one don't incur much performance improvement because there is no additional information gained, and instead they occupy more space in GPU [20].

4.2 Sub-datasets B-E

This was a set of 21 experiments exploring the impact of the patch size and rescaling of the model input on the performance. Data augmentation was also assessed simultaneously. Only the shrub class of sub-datasets B-E was used. Due to the time and computational constraints some experiments were omitted even though they exhibited the best performance (mainly the big datasets with 3808 samples). The total number of parameters was 1.18 million for all experiments. The assumptions were the following:

1) The patch size:
   a) Building on the studies of [21] and [22], the accuracy is expected to improve with increasing patch size, because bigger patch captures more spatial context. This is illustrated in Figure 4.
2) The model input size:
   a) Resizing images to smaller resolutions may lead to a loss of information [23]. Reina et al. (2020) indeed achieved a better performance with minimal down-scaling,
   b) whereas according to [24] and [25], down-scaling the input patch can benefit the results by better filtering the relevant spatial patterns. This can, therefore, depend on the content of the images and what is the target group. The goal was to find out which approach would work for the data used in this thesis.

Tested scales were 1:1 (patch size close to the original tile size) according to [22], and 1:2 according to [25]. Because the input has to be compatible with the 4 max-pooling layers contained in the architecture of the TGS U-Net, and therefore must be divisible by 2^4, the scales were not always exactly that.

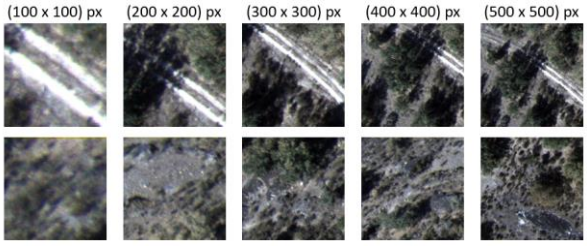

Figure 4 Examples of patches with different sizes (from left: sub-dataset A, sub-dataset B, sub-dataset C, sub-dataset D, sub-dataset E)

The results evidently support data augmentation as a means of improving the performance. The biggest differences among F1 scores of models trained on the same sub-dataset were between 808- and 1658-instance datasets, while it begins to plateau at 3808 samples (Figure 5). Apparently, there is not sufficient amount of information in the 808 sample datasets, whereas doubling it to 1658 seems to be already satisfactory and expanding it even further does not anymore yield such big differences in the F1 score.

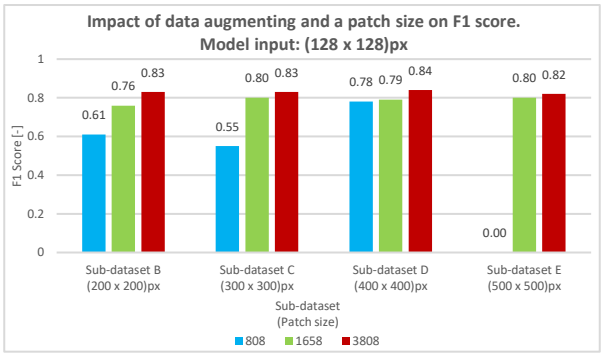

Figure 5 Impact of data augmenting and a patch size on F1 score. Model input: (128 x 128) px

Random data augmentation generates different data every time and could be treated as another hyperparameter, since changing the deformation types [22] or their argument values could yield different classification results. Increasing the patch size beyond (300 x 300) px proved not to be justified anymore, since it didn't improve the classification results, similarly as in case of [18], instead it increased time and computational requirements.

Degrading images into too small resolutions significantly hampered the ability to detect structures and textures. The closer was the size of the rescaled patch to the original size, the higher F1 score was achieved. This is especially important in cases where the size of the objects of interest is already small [14] or where downscaling would lead to a loss of relevant context information [22], [23]. However, it is an interesting technique for shortening the training time [14] and the scale 1:2 is a good tradeoff between the little drop in performance and a shorter training time [25]. The three best performing models (F1 = 0.90) that can be seen in Figure 6, took 50, 46 and 60 hours to train and qualitatively didn't bring much of a value.

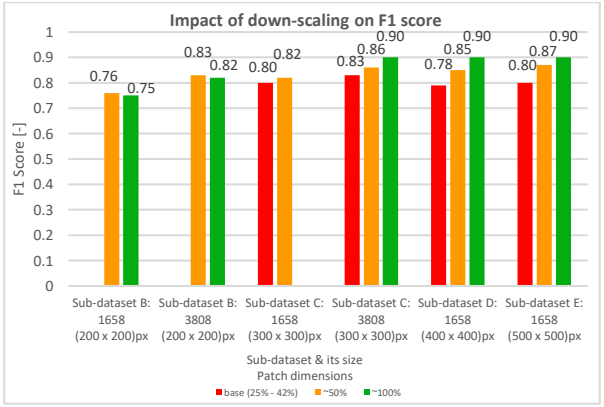

Figure 6 Impact of down-scaling on F1 score

The configuration of pre-processing techniques yielding the best results depends on the problem and on the object of interest [17]. Finding an optimal set of these methods for this particular problem would require further exhaustive research, but could bring a lot of benefit. In overall, inaccurate labelling certainly was one of the most important factors affecting the performance. High quality labels remain to be one of the central elements of image classification success.

The best tradeoff between training time and performance was achieved by model using sub-dataset C with 1664 (300 x 300) px samples with 50% reduction in spatial dimensions (C-1664_144x144). It achieved validation F1 score of 0.82 in about four hours of training.

4.3 Balancing the dataset

The impact of creating training datasets with different target class representations on the model's performance is studied in this part. Non-augmented sub-dataset C (117 samples) was filtered out of samples containing less than 1% [26] and less than 45% [27] of shrub pixels. Under-sampling method was used.

The experiment showed dropping performance metrics with the increasing proportion of the shrub class representation in the dataset. The learning process seemed to be not robust enough, possibly because the used sub-dataset was created with heavy augmentation beacuse only 15 patches contained more than 45% shrub pixels, which significantly lowered down the representativeness of already small data sample. This small sub-dataset surely didn't effectively cover the different arrangements of land cover in such a diverse heterogeneous scene as was present in the study data, which means the model's recognizing abilities may not be sufficient with new data. The under-sampling had definitely led to the loss of important information.

5. Hyperparameter tuning

This section addresses the impact of different initial number of filters, dropout rate and batch size on the performance. The search was manual using the following values:

1. The initial number of filters: 16, 32 and 64 [26]. The total number of parameters in the models was 1.18, 4.71 and 18.82 million, respectively.
2. The dropout rate: 0.05, 0.2, 0.5 and 0.75 [28] [26].
3. The batch size: 15, 32 [29]–[31] and 50.

Similarly to the case of [26], adding more filters improved the performance only until certain point (32 filters) after which it started to drop (64 filters), disagreeing with the general notion that deeper networks achieve better accuracies [9]. Using more

filters made the network deeper and more complicated, which was probably not necessary for my kind of data or brought too many weights for the amount of available data that could cause overfitting. The F1 score of the best performing model with 32 filters was 0.84 but took 10 hours to train, while the model with 16 filters achieved 0.82 F1 score in half the time. The metrics generally worsened with the increasing dropout rate. With a difficult task that includes a landcover as complex as the one present by the data used in this thesis, the more neurons facilitate the learning process the better. Therefore, using high dropout rates might not be a reasonable choice in problems like this one. Batch size is a hyperparameter that, as many others, depends on many factors such as the type of a problem or data. Some [31] reported the best results when using batch size as small as 2 or 4, while others [32] favored batch sizes as big as 128. The batch size in didn't have much of an impact on the results in this study. Taking into account that further exploration of a batch size tuning would be reliant on computational resources and that the batch size of 32 is generally recommended as an optimum, further experimenting with this hyperparameter are not necessary in this particular case.

There are many other hyperparameters that could be further explored in order to improve the classification results, but the optimal model generally depends on the used data[10], so it is not only important to tune the hyperparameters, but also to choose them diligently, since some of them may have significant impact on the results while others can have almost none.

6. Test data

In this final test phase, all experiments were evaluated on independent tests sets described in 3.2

As expected, models performed worse on the new data. The reason is that the test data didn't come from the same dataset as training and validation data (excluding test set 1). The best average performance of all the experiments was achieved by test set 2 (F1 = 0.70), while test sets 1 and 3 performed equally. The biggest culprit behind the gap between validation and test results turned out to be an unlucky pick of testing patches that were somehow different from train and validation sets, e.g. had a different distribution of the vegetation. Patches in the test set 3 came from different images taken in a different year, that was most likely the single biggest factor worsening the performance. The best evaluations were on test set 2 because these images were taken on the same day as the training ones. The image from which training and validation patches were derived didn't supply various enough data and the patches were an unrepresentative sample of the shrub patterns in the area. A more robust model could have been obtained by training on a larger dataset of patches derived from different images, taken on different days and in different years, that would improve the representativeness of the data and could have increased the variety of features to learn. The winter images were too different to be extrapolated from the summer data. A separate model would be necessary.

Higher testing performances (0.76 to 0.77) were generally achieved by models using bigger patch sizes and model input dimensions, in accordance with the validation results from the 4.2 section. Data augmentation, patch size and model input dimensions (i.e. down-scaling) proved to be beneficial for the training and classification performance. The hyperparameter tuning didn't bring any significant improvements in the performance, neither for validation, nor for test sets. Generally, the gap between validation and test scores are relative to the data, selected metrics and models[11].

7. Conclusion

This thesis explored the potential of detecting the target vegetation type in a complex heterogeneous landscape with U-Net. Shrubs are wild plants with different shapes, sizes and distribution patterns. The difficulty of this task was increased further with the fact that the data contained more than species of shrubs scattered in the forest area. Shrubs are of a priority interest in terms of fire risk in dry Mediterranean regions and mapping them can serve as a basis for better informed land management and reduction of the forest fire hazard. This work consisted of two main parts: creating and manually labelling the datasets and developing a method to increase a detection accuracy using a U-Net neural network. The impact of data augmentation, tiling, rescaling, balancing the dataset and hyperparameter tuning (number of filters, dropout rate and batch size) was explored in this regard.

The beneficial methods were:

[10] https://jakevdp.github.io/PythonDataScienceHandbook/05.03-hyperparameters-and-model-validation.html

[11] https://machinelearningmastery.com/the-model-performance-mismatch-problem/

- Data augmentation: The biggest datasets containing 3808 samples yielded the highest F1 scores
- Patch size: (300 x 300) px was the optimum. Patches bigger than that didn't bring any significant improvement in performance, instead it increased the training time and computational demands
- Down-scaling: Degrading the image resolution leads to a loss of information, but the scale 1:2 significantly decreased the training time and didn't lead to a dramatic drop in performance.

Major identified limitations were a very little labelled data insufficient for learning of all the important features from scratch and incorrect labels. Using bigger datasets with patches derived from several images taken during multiple flights could have a significant positive effect on the results.

Based on the results achieved in this thesis I believe that further improvements in performance could be achieved by:

- Further enlargement of the datasets, primarily as a result of more labelled data from spatially independent samples, but also by employing more data augmentation
- Finding an optimal configuration of the augmentation techniques suitable for these data
- Finding an optimal configuration of pre-processing methods as well as hyperparameters for this particular data and task
- Employing transfer learning

This thesis demonstrated the capacity of U-Net for mapping the irregular shrub cover, presented methods improving the classification results and provided recommendations for a future research. The work has a potential to serve as an information tool for land planning and grazing management and could be also modified and repurposed to map other vegetation types, such as trees, or to be used as e.g. a forest inventory tool.

References

[1] V. Proença and C. M. G. L. Teixeira, "Beyond meat: Ecological functions of livestock," *Science*, vol. 366, no. 6468, pp. 962–962, Nov. 2019, doi: 10.1126/science.aaz7084.

[2] W. J. Ripple *et al.*, "Collapse of the world's largest herbivores," *Sci. Adv.*, vol. 1, no. 4, p. e1400103, May 2015, doi: 10.1126/sciadv.1400103.

[3] R. Lovreglio, O. Meddour-Sahar, and V. Leone, "Goat grazing as a wildfire prevention tool: a basic review," *iForest*, vol. 7, no. 4, pp. 260–268, Aug. 2014, doi: 10.3832/ifor1112-007.

[4] L. A. Pérez-Rodríguez, C. Quintano, E. Marcos, S. Suarez-Seoane, L. Calvo, and A. Fernández-Manso, "Evaluation of Prescribed Fires from Unmanned Aerial Vehicles (UAVs) Imagery and Machine Learning Algorithms," *Remote Sensing*, vol. 12, no. 8, p. 1295, Apr. 2020, doi: 10.3390/rs12081295.

[5] M. Volpi and D. Tuia, "Dense Semantic Labeling of Subdecimeter Resolution Images With Convolutional Neural Networks," *IEEE Trans. Geosci. Remote Sensing*, vol. 55, no. 2, pp. 881–893, Feb. 2017, doi: 10.1109/TGRS.2016.2616585.

[6] D. Wen, X. Huang, H. Liu, W. Liao, and L. Zhang, "Semantic classification of urban trees using very high resolution satellite imagery," *IEEE Journal of Selected Topics in Earth Observation and Remote Sensing*, vol. 10, no. 4, pp. 1413–1424, Jan. 2017, doi: 10.1109/JSTARS.2016.2645798.

[7] S. Paisitkriangkrai, J. Sherrah, P. Janney, and A. Van Den Hengel, "Semantic labeling of aerial and satellite imagery," 2016, doi: 10.1109/JSTARS.2016.2582921.

[8] C. Yu, J. Wang, C. Peng, C. Gao, G. Yu, and N. Sang, "Learning a Discriminative Feature Network for Semantic Segmentation," in *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Salt Lake City, UT, USA, Jun. 2018, pp. 1857–1866, doi: 10.1109/CVPR.2018.00199.

[9] R. Li *et al.*, "DeepUNet: A Deep Fully Convolutional Network for Pixel-Level Sea-Land Segmentation," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 11, no. 11, pp. 3954–3962, Nov. 2018, doi: 10.1109/JSTARS.2018.2833382.

[10] J. Long, E. Shelhamer, and T. Darrell, "Fully Convolutional Networks for Semantic Segmentation," *arXiv:1411.4038 [cs]*, Mar. 2015, Accessed: Nov. 02, 2020. [Online]. Available: http://arxiv.org/abs/1411.4038.

[11] V. Badrinarayanan, A. Kendall, and R. Cipolla, "SegNet: A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation," *arXiv:1511.00561 [cs]*, Oct. 2016, Accessed: Dec. 27, 2020. [Online]. Available: http://arxiv.org/abs/1511.00561.

[12] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional Networks for Biomedical Image Segmentation," *arXiv:1505.04597 [cs]*, May 2015, Accessed: Oct. 13, 2020. [Online]. Available: http://arxiv.org/abs/1505.04597.

[13] A. Stoian, V. Poulain, J. Inglada, V. Poughon, and D. Derksen, "Land Cover Maps Production with High Resolution Satellite Image Time Series and Convolutional Neural Networks: Adaptations and Limits for Operational Systems," *Remote Sensing*, vol. 11, no. 17, p. 1986, Aug. 2019, doi: 10.3390/rs11171986.

[14] N. Audebert, B. Le Saux, and S. Lefèvre, "Beyond RGB: Very high resolution urban remote sensing with multimodal deep networks," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 140, pp. 20–32, Jun. 2018, doi: 10.1016/j.isprsjprs.2017.11.011.

[15] L.-C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, and A. L. Yuille, "DeepLab: Semantic Image Segmentation with Deep Convolutional Nets, Atrous Convolution,

and Fully Connected CRFs," *arXiv:1606.00915 [cs]*, May 2017, Accessed: Dec. 27, 2020. [Online]. Available: http://arxiv.org/abs/1606.00915.

[16] M. Mahdianpari, B. Salehi, M. Rezaee, F. Mohammadimanesh, and Y. Zhang, "Very Deep Convolutional Neural Networks for Complex Land Cover Mapping Using Multispectral Remote Sensing Imagery," *Remote Sensing*, vol. 10, no. 7, p. 1119, Jul. 2018, doi: 10.3390/rs10071119.

[17] E. Guirado, S. Tabik, D. Alcaraz-Segura, J. Cabello, and F. Herrera, "Deep-Learning Convolutional Neural Networks for scattered shrub detection with Google Earth Imagery," *arXiv:1706.00917 [cs]*, Jun. 2017, Accessed: Dec. 23, 2020. [Online]. Available: http://arxiv.org/abs/1706.00917.

[18] C. Hung, Z. Xu, and S. Sukkarieh, "Feature Learning Based Approach for Weed Classification Using High Resolution Aerial Images from a Digital Camera Mounted on a UAV," *Remote Sensing*, vol. 6, no. 12, Art. no. 12, Dec. 2014, doi: 10.3390/rs61212037.

[19] A. Ashapure, J. Jung, A. Chang, S. Oh, M. Maeda, and J. Landivar, "A Comparative Study of RGB and Multispectral Sensor-Based Cotton Canopy Cover Modelling Using Multi-Temporal UAS Data," *Remote Sensing*, vol. 11, no. 23, p. 2757, Nov. 2019, doi: 10.3390/rs11232757.

[20] L. Zheng, Y. Zhao, S. Wang, J. Wang, and Q. Tian, "Good Practice in CNN Feature Transfer," *arXiv:1604.00133 [cs]*, Apr. 2016, Accessed: Dec. 02, 2020. [Online]. Available: http://arxiv.org/abs/1604.00133.

[21] T. Kattenborn, J. Eichel, S. Wiser, L. Burrows, F. E. Fassnacht, and S. Schmidtlein, "Convolutional Neural Networks accurately predict cover fractions of plant species and communities in Unmanned Aerial Vehicle imagery," *Remote Sens Ecol Conserv*, p. rse2.146, Feb. 2020, doi: 10.1002/rse2.146.

[22] G. A. Reina, R. Panchumarthy, S. P. Thakur, A. Bastidas, and S. Bakas, "Systematic Evaluation of Image Tiling Adverse Effects on Deep Learning Semantic Segmentation," *Front. Neurosci.*, vol. 14, p. 65, Feb. 2020, doi: 10.3389/fnins.2020.00065.

[23] W. Zhang, P. Tang, and L. Zhao, "Remote Sensing Image Scene Classification Using CNN-CapsNet," *Remote Sensing*, vol. 11, no. 5, p. 494, Feb. 2019, doi: 10.3390/rs11050494.

[24] J. Müllerová, J. Brůna, T. Bartaloš, P. Dvořák, M. Vítková, and P. Pyšek, "Timing Is Important: Unmanned Aircraft vs. Satellite Imagery in Plant Invasion Monitoring," *Front. Plant Sci.*, vol. 8, p. 887, May 2017, doi: 10.3389/fpls.2017.00887.

[25] A. Rakhlin, A. Davydow, and S. Nikolenko, "Land Cover Classification from Satellite Imagery with U-Net and Lovász-Softmax Loss," in *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, Salt Lake City, UT, USA, Jun. 2018, pp. 257–2574, doi: 10.1109/CVPRW.2018.00048.

[26] P. Zhang, Y. Ke, Z. Zhang, M. Wang, P. Li, and S. Zhang, "Urban Land Use and Land Cover Classification Using Novel Deep Learning Models Based on High Spatial Resolution Satellite Imagery," *Sensors*, vol. 18, no. 11, p. 3717, Nov. 2018, doi: 10.3390/s18113717.

[27] Q. Wei and R. L. D. Jr, "The Role of Balanced Training and Testing Data Sets for Binary Classifiers in Bioinformatics," *PLOS ONE*, vol. 8, no. 7, p. e67863, Jul. 2013, doi: 10.1371/journal.pone.0067863.

[28] F. Zhang, B. Du, and L. Zhang, "Saliency-Guided Unsupervised Feature Learning for Scene Classification," *IEEE Trans. Geosci. Remote Sensing*, vol. 53, no. 4, pp. 2175–2184, Apr. 2015, doi: 10.1109/TGRS.2014.2357078.

[29] Y. Bengio, "Practical recommendations for gradient-based training of deep architectures," *arXiv:1206.5533 [cs]*, Sep. 2012, Accessed: Dec. 14, 2020. [Online]. Available: http://arxiv.org/abs/1206.5533.

[30] N. S. Keskar, D. Mudigere, J. Nocedal, M. Smelyanskiy, and P. T. P. Tang, "On Large-Batch Training for Deep Learning: Generalization Gap and Sharp Minima," *arXiv:1609.04836 [cs, math]*, Feb. 2017, Accessed: Dec. 14, 2020. [Online]. Available: http://arxiv.org/abs/1609.04836.

[31] D. Masters and C. Luschi, "Revisiting Small Batch Training for Deep Neural Networks," *arXiv:1804.07612 [cs, stat]*, Apr. 2018, Accessed: Nov. 09, 2020. [Online]. Available: http://arxiv.org/abs/1804.07612.

[32] V. Iglovikov, S. Mushinskiy, and V. Osin, "Satellite Imagery Feature Detection using Deep Convolutional Neural Network: A Kaggle Competition," *arXiv:1706.06169 [cs]*, Jun. 2017, Accessed: Dec. 02, 2020. [Online]. Available: http://arxiv.org/abs/1706.06169.