Indicator plant species detection in grassland using EfficientDet object detector

Deepak Hanike Basavegowda[1,2], Paul Mosebach[3], Inga Schleip[3] and Cornelia Weltzien[1,2]

Abstract: Extensively used grasslands (meadows and pastures) are ecologically valuable areas in the agricultural landscape and part of the multifunctional agriculture. In Germany, the quality of these grasslands is assessed based on the occurrence of certain plant species known as indicator or character species, with indicators being defined at regional level. Therefore, the recognition of these indicators on a spatial level is a prerequisite for monitoring grassland biodiversity. The identification of indicator species for the status quo of grassland using traditional methods was found to be challenging and tedious. Deep learning-algorithms applied to high-resolution UAV imagery could be the key solution, where UAV with remote sensors can map a large area of grassland in comparison to manual or ground mapping methods and deep learning-algorithms can automate the detection process. In this research work, we use an EfficientDet based algorithm to train an object detection model capable of recognizing indicators on RGB data. The experimental results show that this approach is very promising in contrast to the difficult and time-consuming manual recognition methods. The model was trained with the momentum-SGD optimizer with a momentum value of 0.9 and a learning rate of 0.0001. The model was trained and tested on 1200 images and achieves 45.7 AP (and 85.7 AP_{50}) on test data set. The dataset includes images of four distinct indicator plant species: Armeria maritima, Campanula patula, Cirsium oleraceum, and Daucus carota.

Keywords: digital agriculture, indicators recognition, biodiversity in grassland, HNV farming, deep learning, object detection

1 Introduction

Grassland accounts for more than one third of the agricultural land in Germany, with a total area of around 5 million hectares [UB21], and is among the most species-rich habitats in Europe. It plays a very important role in biodiversity and ecosystem service provision across Europe, providing habitat for many endangered species and having a positive impact on plant diversity, soil stability, resilience to extreme weather events and carbon sequestration [DP09]. Therefore, these non-intensively used grassland regions are primary targets of nature conservation in the EU Common Agricultural Policy (CAP) [EC21] and are considered an important part of high nature value (HNV) farmland [DP09].

[1] Leibniz-Institut für Agrartechnik und Bioökonomie e.V., Technik im Pflanzenbau, Max-Eyth-Allee 100, 14469 Potsdam-Bornim, dbasavegowda@atb-potsdam.de; CWeltzien@atb-potsdam.de
[2] Technische Universität Berlin, FG Agromechatronik, Straße des 17. Juni 135, 10623 Berlin,
[3] Hochschule für Nachhaltige Entwicklung Eberswalde, Schickerstraße 5, 16225 Eberswalde, Paul.Mosebach@hnee.de; Inga.Schleip@hnee.de

In Germany, native grassland plant species are used as indicators to identify HNV grassland. Indicators are plants and animals that, by their presence, abundance, absence, or chemical composition, reveal a specific aspect of the character or quality of the site on which they thrive. These species are defined on a region level, resulting in seven different indicator species lists for Germany [BFH15]. To identify HNV grassland, every seemingly species-rich and homogenous area is examined for indicator species by using a standardized transect of 30 m length and 2 m width [BFH15; St17]. This transect-based approach is labour intensive as it has to be conducted regularly on about 5 million hectares of grassland and requires a good knowledge of grassland plant species at the regional level. Therefore, there is a need for new techniques to identify indicator plants in a given data sample.

Convolutional Neural Networks (CNN), a deep learning-oriented computer vision technique, are evolving as promising tools for plant species identification and biodiversity monitoring. The general structure of a typical CNN is composed of multiple convolutional layers interlaced with pooling layers, followed by some fully connected layers in the end. The convolutional layers automatically extract the hierarchical and high-level feature representation, while pooling layers reduce the image resolution and help achieve spatial invariance, and the fully-connected layers reduce the image dimension and output a 1-D distribution over classes. In general, the maximum output in the 1-D distribution corresponds to the predicted result, which classifies one image to one single label [LB95].

In this paper, we briefly explain the importance of indicator species identification for grassland biodiversity monitoring and explain our data collection effort. We train and evaluate an object detection-model based on a transfer learning-approach for identifying and locating grassland indicator species.

2 Materials and Methods

2.1 Indicator species selection

For this research work, indicator species were selected based on the fact that the species are used for monitoring HNV farmland areas and are included in the list of indicator species of the *NO* region (in total, the list includes 28 species and species groups), a region including the Federal States of Mecklenburg-Western Pomerania and Brandenburg in Germany [BF18]. The following indicator species were selected for our study (Fig. 1): *Armeria maritima*, *Campanula patula*, *Cirsium oleraceum* and *Daucus carota*.

To our knowledge, there is no open source image dataset available for the indicator plant species. Although a few images are available under the Creative Commons (CC) license in various search engines, they are biased for a specific growth stage and are available in very small numbers. To collect images of the indicators in their vegetative growth phase, the selected indicators were grown in a greenhouse.

Fig. 1: Selected indicator plant species

(i) Armeria maritima, (ii) Campanula patula, (iii) Cirsium oleraceum, and (iv) Daucus carota (left to right) (photo: Hanike Basavegowda, Deepak)

2.2 Data Collection

A four-wheeled hand-driven equipment carrier 1.4 m wide, 1.5 m long and 2.7 m high above the ground was designed and fabricated to mount an image sensor. Images were acquired with the DSLM camera ILCE-6000 (Sony, Japan) with an APS-C type sensor chip (23.5 × 15.6 mm) and a 50 mm lens attached. The camera was mounted on an equipment carrier in the nadir position to capture ortho-images. The images used for this work were taken during the 4th week of June 2020 by randomly placing indicator plants on a 15 m long and 2 m wide meadow outside of the green house. The distance between the camera and the plant surface was set to capture images with a ground surface distance (GSD) of 0.15 mm/pixel on a projected area of 0.9 m × 0.6 m on the ground. The camera was set in autonomous mode to capture images with a frequency of one image per second as the camera carrier moved across the meadow. Six hundred indicator plant samplings, one hundred fifty from each class, were used for data collection work. We had collected approximately 1200 images during this work.

2.3 Object Detection Model

EfficientDet, a scalable and efficient object detection model, was proposed by the Google Brain Team [TPL20]. The family of EfficientDet models are a great choice for object detection tasks considering their higher efficiency and better efficiency across a wide spectrum of resource constraints. The architecture of EfficientDet consists of three basic components: (i) EfficientNet as a backbone network to extract features from the given image; (ii) BiFPN or Weighted Bi-directional Feature Pyramid Network as a feature network that takes multiple levels of features from the backbone as input and outputs a list of fused features representing salient features of the image; and (iii) the final class/box network that uses the fused features to predict the class and location of each object [TPL20].

3 Results and Discussion

The CNN-based object detector was trained using the transfer learning-approach since training the model from scratch requires a large dataset. The pre-trained EfficientDet object detection model was downloaded from the TensorFlow 2 Detection Model Zoo [Hu17]. A pre-trained model is a stored network that has been previously trained on a large dataset. During training, through backpropagation, errors from the new data were used to update the weights of the convolutional layers to fit the model into a new task.

The model was trained and evaluated on an Ubuntu workstation with images of size 768 × 768 pixels, using Python 3.9 as the programming language and TensorFlow 2 (TF 2) version 2.6 as the deep learning-framework. Four CUDA 11.4-compatible GeForce RTX 2080 Ti graphics cards (NVIDIA, California, USA) were used for model training. The images used for training were randomly augmented during training to increase the robustness and transferability of the model. The augmentation techniques used are random horizontal and vertical flipping and random cropping with a pad to ensure the uniform input size.

The model was trained for 500 epochs with a batch size of 12. The final class values were calculated using the sigmoid function. The Momentum SGD optimizer with a momentum value of 0.9 and a learning rate of 0.0001 was used for optimization. Figure 2 shows the loss convergence of the detection model during training. The model weights of an epoch were only considered if the accuracy exceeded the accuracy obtained in the previous epoch. We evaluated the accuracy of the final model using the test data, i.e., the images that were not used for training the CNN model. The model achieved 45.7 AP (and 85.7 AP_{50}) on the test data set.

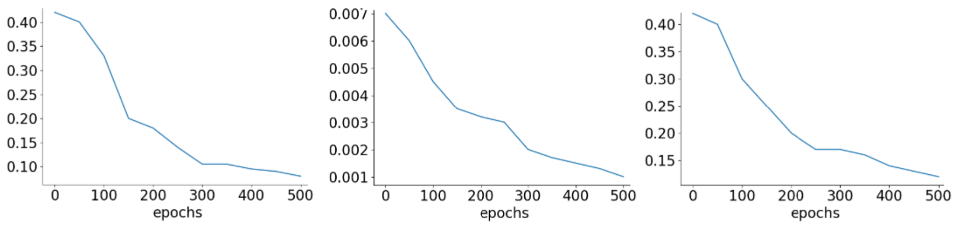

Fig. 2: Training losses of the object detection model

(i) Classification Loss, (ii) Localization Loss, and (iii) Total Loss (left to right)

The result of this work, the object detection model for identifying indicator species in grassland, can be readily applied to grassland images to search for indicator species. Species recognition can then be linked to the HNV farmland (extensive grassland) type addressed and its quality classes. For example, based on the number of indicator species detection, the grassland is classified into one of three classes [An03; St17]:

- HNV 1: exceptionally high nature value farmland (8 or more HNV character species)
- HNV 2: very high nature value farmland (6 or 7 HNV character species)
- HNV 3: moderately high nature value farmland (4 or 5 HNV character species)

Previous studies focused on HNV farmland mostly addressed other HNV farmland types [An03], and limited works focused on HNV grassland detection relied on spectral variability of grassland vegetation to distinguish HNV grassland from others [St17] and were unable to differentiate grassland quality levels. Our method can be used to identify both HNV grasslands and their quality levels. Furthermore, the method can be extended to include more indicator species and can be easily applied to all regions, as most indicator species overlap in the regional species list.

Fig. 3: Indicator species identification using the object detection model

4 Conclusion

In this work, RGB images of indicator species were collected by growing them in a greenhouse, and then a CNN-based object detection-model was proposed for species recognition and localization. Since the image dataset was not sufficient to train a model from scratch, a transfer learning approach was used to train the model. Our results were in line with those of the original model [TPL20] and show that indicator species detection using deep learning is a promising way to monitor grasslands. It should also be noted that indicator species with less conspicuous leaves or small characteristic species in tall vegetation might be more difficult to identify using this method. In our future work, the training data will include more images to cover the complete growth stages of the indicators, and the method will be applied to high resolution UAV-imagery.

Acknowledgment

This research was conducted as part of the *Digital Agricultural Knowledge and Information System* (DAKIS) project and funded by the Federal Ministry of Education and Research (BMBF), Germany. DAKIS Homepage: https://adz-dakis.com/

References

[An03] Andersen, E.; Baldock, D.; Bennett, H.; Beaufoy, G.; Bignal, E.; Brouwer, F.; Elbersen, B.; Eiden, G.; Godeschalk, F.; Jones, G.; McCracken, D.I.; Nieuwenhuizen, W.; vanEupen, M.; Hennekens, S.; Zervas, G: Developing a High Nature Value Indicator. Report for the European Environment Agency (EEA), 2003.

[Ar12] Armsworth, P. R.; Acs, S.; Dallimer, M.; Gaston, K. J.; Hanley, N.; Wilson, P.: The cost of policy simplification in conservation incentive programs. Ecology letters, 15(5), S. 406-414, 2012.

[BFH15] Benzler, A., Fuchs, D., Huenig, C.: Methodik und erste Ergebnisse des Monitorings der Landwirtschaftsflächen mit hohem Naturwert in Deutschland. Beleg für aktuelle Biodiversitätsverluste in der Agrarlandschaft. Natur und Landschaft 90 (7), S. 309-316, 2015.

[BF18] Benzler, A.; Fuchs D.: Biodiversität in der Agrarlandschaft: erstmals ein Stopp des Rückgangs? Natur und Landschaft 93(9/10), S. 470-471, 2018.

[DP09] Dierschke, H.; Peppler-Lisbach, C.: Erhaltung und Wiederherstellung der Struktur und floristischen Biodiversität von Bergwiesen – 15 Jahre wissenschaftliche Begleitung von Pflegemaßnahmen im Harz. Tuexenia, 29, 145-179, 2009.[EC21] European Commission, https://ec.europa.eu/info/sites/default/files/food-farming-fisheries/key_policies/documents/technical-handbook-monitoring-evaluation-framework_june17_en.pdf, 20.11.2021.

[Hu17] Huang, J.; Rathod, V.; Sun, C.; Zhu, M.; Korattikara, A.; Fathi, A.; Murphy, K.: Speed/accuracy trade-offs for modern convolutional object detectors. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 7310-7311), 2017.

[LB95] LeCun, Y.; Bengio, Y.: Convolutional networks for images, speech, and time series. The handbook of brain theory and neural networks, 3361(10), 1995.

[St17] Stenzel, S.; Fassnacht, F. E.; Mack, B.; Schmidtlein, S.: Identification of high nature value grassland with remote sensing and minimal field data. Ecological indicators, 74, S. 28-38, 2017.

[TPL20] Tan, M.; Pang, R.; Le, Q. V.: Efficientdet: Scalable and efficient object detection. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (pp. 10781-10790), 2020.

[UB21] Umwelt Bundesamt, https://www.umweltbundesamt.de/daten/land-forstwirtschaft/gruenlandumbruch#gefahrdung-des-grunlands, 20.11.2021.