



DETECTION OF PLASTIC GREENHOUSES USING HIGH RESOLUTION RGB REMOTE SENSING DATA AND CONVOLUTIONAL NEURAL NETWORK

Balázs Jakab^{1*}, Boudewijn van Leeuwen¹, Zalán Tobak¹

¹Department of Geoinformatics, Physical and Environmental Geography, University of Szeged,
Egyetem u. 2-6, 6722 Szeged, Hungary

*Corresponding author, email: jakabbalazs501@gmail.com

Research article, received 5 February 2021, accepted 13 April 2021

Abstract

Agricultural production in greenhouses shows a rapid growth in many parts of the world. This form of intensive farming requires a large amount of water and fertilizers, and can have a severe impact on the environment. The number of greenhouses and their location is important for applications like spatial planning, environmental protection, agricultural statistics and taxation. Therefore, with this study we aim to develop a methodology to detect plastic greenhouses in remote sensing data using machine learning algorithms.

This research presents the results of the use of a convolutional neural network for automatic object detection of plastic greenhouses in high resolution remotely sensed data within a GIS environment with a graphical interface to advanced algorithms. The convolutional neural network is trained with manually digitized greenhouses and RGB images downloaded from Google Earth. The ArcGIS Pro geographic information system provides access to many of the most advanced python-based machine learning environments like Keras – TensorFlow, PyTorch, fastai and Scikit-learn. These libraries can be accessed via a graphical interface within the GIS environment. Our research evaluated the results of training and inference of three different convolutional neural networks. Experiments were executed with many settings for the backbone models and hyperparameters. The performance of the three models in terms of detection accuracy and time required for training was compared. The model based on the VGG_11 backbone model (with dropout) resulted in an average accuracy of 79.2% with a relatively short training time of 90 minutes, the much more complex DenseNet121 model was trained in 16.5 hours and showed a result of 79.1%, while the ResNet18 based model showed an average accuracy of 83.1% with a training time of 3.5 hours.

Keywords: plastic greenhouse, deep learning, convolutional neural network, satellite image, Google Earth

INTRODUCTION

In recent years, agricultural production in greenhouses showed a rapid growth (Agüera and Liu, 2009; Wu et al., 2016; Nemmaoui et al., 2019). In many arid and semi-arid countries, plastic greenhouses form a large share in the total number of greenhouses, since they are more affordable and can be used temporary as well. Plastic greenhouses are made of a partly transparent plastic cover to be able to control the environmental and growing conditions inside the greenhouse. It is important to monitor their spatial distribution, since this form of intensive farming requires large amounts of water and fertilizers and can have a severe impact on the environment. Estimation of the share of plastic greenhouses in the total agricultural activities can be performed by directly counting the number of greenhouses. This is slow, labor intensive and consequently expensive, therefore, it makes sense to apply remote sensing data-based algorithms to detect them. Apart from spatial planning and environmental protection, another reason to acquire knowledge of the number and location of greenhouses is that their registration is obligatory for taxation purposes.

Research on the classification of plastic or glass greenhouses using very high to medium resolution

(multi-spectral) remote sensing data and methodologies has been published earlier. Wu et al. (2016) applied random forest (RF) and support vector machine (SVM) on medium resolution multispectral data. Also, Yang et al. (2017) used medium resolution data, but they presented an index-based approach resulting in an overall accuracy of 91%. Koc-San (2013) reported high accuracies of classification of glass and plastic greenhouses using maximum likelihood (ML), RF and SVM methods based on Worldview-2 very high-resolution data. Agüera et al. (2008) received promising results when applying texture analysis combined with ML on very high-resolution satellite imagery. Agüera and Liu (2009) used ML classification to automatically delineate greenhouses. They report results with medium accuracy. Supervised classification based on a combination of orthophotos and Landsat data was proposed by González-Yebra et al. (2018). Novelli et al. (2016) used a combination of Landsat and Sentinel-2. They classified medium resolution data using Object Based Image Analysis (OBIA) and RF. Accuracies ranged between 89 and 93%. Very high-resolution satellite imagery was used by Nemmaoui et al. (2019) to derive surface and terrain models to extract plastic greenhouses. They report very high accuracies of up to 98%. Most recently, Yang et al. (2021) published a

manual approach to identify greenhouses to study urban fringe based on imagery downloaded from Google Earth.

The term artificial intelligence (AI) was first used in 1955 (McCarthy et al., 1955). Given that intelligence is difficult to define, this term is not easy to define either. It can be formulated as a process that mimics human abilities and behavior according to pre-programmed rules (Nilsson, 1980; Simon, 1995). Machine learning (ML) is a part of artificial intelligence that, based on collected data, can learn and develop itself in an iterative way using pre-programmed rules (Michie, 1968). Artificial Neural Networks (ANN) are a type of ML algorithms loosely based on the biological functioning of the brain. Artificial neurons process and transmit many input signals to a large number of neighboring neurons. The neurons are stored in layers, the final layer collects the signals and processes them to an output signal, with is the result of the network. The network learns from input and output data pairs and stores their combined relationship as weights (Müller et al., 1995). Deep learning is a group of ML algorithms that uses ANNs with many hidden layers. The more hidden layers, the deeper and more complex the neural network and the more complicated tasks it can potentially solve. In the present age, deep learning has become widespread and makes it possible to process large data sets that are otherwise often too big to manage. Examples of applications of deep learning include face recognition, image recognition, or self-driving vehicles (Goodfellow et al., 2016).

The current revolution in deep learning algorithms for computer vision also provides opportunities to improve analysis of remote sensing data. Numerous studies have been published on the classification of medium (e.g., Watanabe et al., 2018; Gallwey et al., 2020; Rai et al., 2020; Virnodkar et al. 2020) and high resolution (e.g. Flood et al., 2019; Schiefer et al., 2020; Zhang et al., 2020) satellite images using deep learning methods (Kattenborn et al., 2021). Detection of individual objects in the imagery is not as common as classification but has been published as well (Ding et al., 2018; Jiang et al., 2019; Guo et al., 2020; Pi et al., 2020). The difference between the two results is important; classification provides a label for every pixel, specifying the class it belongs to. This is the most common approach when converting remote sensing data to thematic maps. On the contrary, object detection provides an output layer on top of the original remote sensing image where the objects of interest are shown with a square bounding box around them indicating their precise location and accuracy estimation.

The aim of our research is to evaluate if the current innovations in machine learning based technologies can be applied to detect plastic greenhouses. The presented methodology is based on object detection using a convolutional neural network (CNN). A CNN is an ANN that is designed to learn the spatial features, e.g. edges, corners, textures, or more abstract shapes, that best describes the target class or quantity. Like other ANNs, CNNs are based on neurons that are organized in layers and are connected through weights and biases. The initial layer is the input layer, e.g. remote sensing data,

and the last layer is the predicted output (Kattenborn et al., 2021).

In recent years, accessibility to machine learning algorithms and deep learning models in particular has been improved by implementations in user-friendly environments under Python or R. A next step in the development towards easier access to the algorithms is the implementation of graphical user interfaces on top of the functionality. One implementation is the Deep learning toolset in ArcGIS Pro (ESRI, 2021) which implements third-party deep learning frameworks – such as Keras – TensorFlow (Abadi et al., 2015; Chollet, 2015), PyTorch (Paszke et al., 2019), fastai (Howard and Guggen, 2020) and Scikit-learn (Pedregosa et al., 2011).

In this study, we present a methodology based on freely available images and a convolutional network to detect plastic greenhouses in an area in the south of Hungary. The area is mainly agricultural with a large amount of tunnel shaped plastic greenhouses. The earlier mentioned studies applied deep learning techniques for classification of high resolution remote sensing data, but none of them used CNN for object detection based on data from Google Earth.

STUDY AREA

A 230 km² area in the south east of the Great Hungarian Plain, near the town of Szeged (Fig. 1) has been selected to test the CNN algorithm. The area is mainly agricultural and has a large number of greenhouses. Other main land use/land cover classes in the area are forest, urban/build up and some water bodies. Most of the area has chernozem and sandy soils, but in some areas arenosol can be found. The sandy soils absorb water quickly causing the soil to dry out and reduce its fertility. The area suffers from high air pollution and dust content. In the 1750s, locust tree (*Robinia pseudoacacia*) was introduced in the region as an ornamental plant. The invasive species spread quickly through the region, and helps to reduce wind erosion, but it also reduces the nutrients in the soil. With 400-450 mm, the annual rainfall is low compared to the mean precipitation (600-700 mm) of the country (Mezősi, 2011).

DATA AND METHODS

The imagery used as input data for the presented detection algorithm was extracted from Google Earth. The high-resolution data is a georeferenced red-green-blue (RGB) image collected by CNES/Airbus in August 2020. The image was downloaded using the Tile+ extension in QGIS.

The concept of convolutional neural networks was introduced in the 1980s by Yann LeCun (LeCun et al., 1990). CNNs differ from fully connected ANNs by having each neuron being connected to only a limited number of neurons in the previous layer. CNNs assume that the input is an image and look for features through a kernel. The detection is performed through convolution between the input and the kernel thus the term convolutional neural networks. The kernels form a

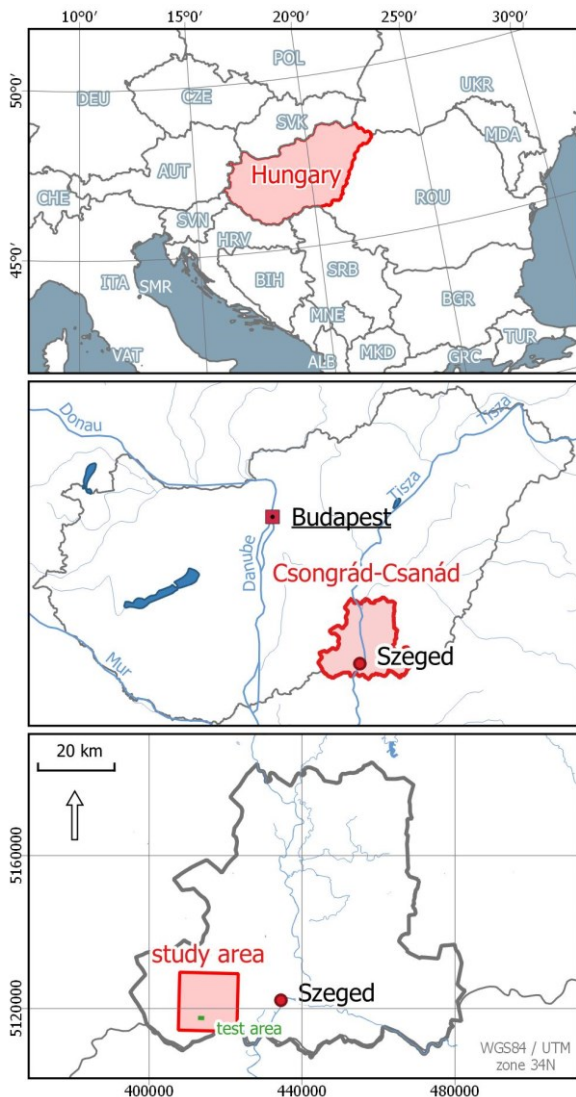


Fig. 1 Location of the study area

convolutional filter, and a set of stacked convolutional filters makes a convolutional layer (Fig. 2). Convolutional layers are followed by activation functions which introduce nonlinear behavior to the model. Each convolutional layer extracts features with increasing complexity from the input layer. After each convolutional layer, a pooling layer extracts the most prominent features and reduces the resolution of the previous input. A CNN thus contains a stack of convolutional layers followed by activation functions and pooling layers, and finally an addition of one or more Fully Connected (FC) layers. FC layers form an ANN head on top of the CNN that is used to classify the CNN output into a set of finite classes

(Davies, 2018). In case of object detection, objects are not only classified, but their locations are indicated with bounding boxes as well (Liu et al., 2016).

The methodology to detect plastic greenhouses can be separated in 6 sequential steps. The first step is to download of the input image. The second is the generation of training and validation samples and the creation of image chips, and the third is the creation of the model architecture. The next step is the training of the model based on the training data. Then, the model parameters are fine-tuned based on the validation data set. The sixth and final step is the inference of the trained model with the total image.

The detection of greenhouses requires very high resolution data with at least 3 layers. This type of data can be provided by drones, aerial photographs or very high resolution satellite images. Since it is the aim of our research to apply the methodology on a large area, images collected by drones are not an option. Google Earth provides a source of high-resolution aerial photographs and satellite images that can be downloaded for free for non-commercial use. Therefore, an RGB image was downloaded from Google Earth with a resolution of 2000 dpi. This resulted in a 1 gigabyte three layer TIF file with an approximate spatial resolution of 50 cm. Selection criteria for the image were cloud cover percentage, spatial resolution and number of greenhouses. The extracted image covers an area of 230 km².

During the second step, samples of plastic greenhouses were identified in multiple subsets of the image. In each subset, all greenhouses were digitized manually to make sure that the model would not be trained with pixels that belong to greenhouses, but that were labeled as non-greenhouse pixels. In total, 2352 greenhouse samples were created. The higher the spatial resolution of the image, the easier is the identification of individual greenhouses; the downloaded image was of sufficient spatial resolution. The selection of the samples is of decisive important for the result of the detection of greenhouses. Using rotation, it is possible to perform data augmentation, with is the artificial creation of more training samples by capturing the sample created by the user at multiple angles.

The samples were used to generate training data in the PASCAL visual object classes meta data format (Everingham et al., 2010) and serve as input for the process of sub-setting the total image into individual chips. Only chips with (a part of) at least one sample in it were stored. With each chip an .xml file is produced that stores the location of the sample within the chip.

The training data was used as input for the training phase, where the algorithm aims to detect the

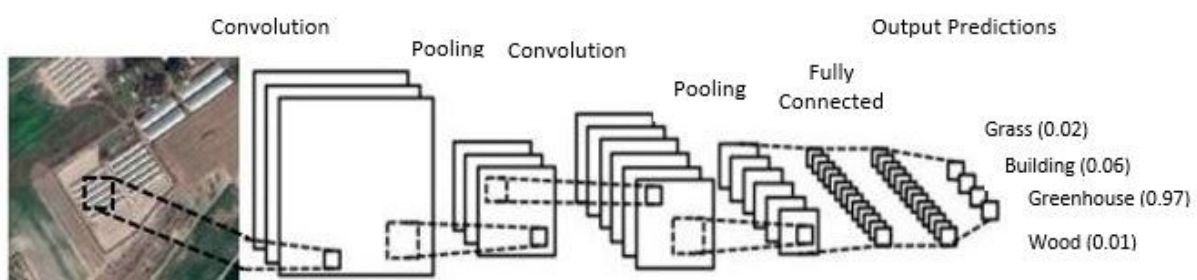


Fig. 2 Classification using a convolution neural network

greenhouses. During object detection, the algorithm uses bounding boxes to delimit the objects' location. The purpose of the training is to minimize the difference between the real and the modeled bounding boxes. In case there are several different objects in the image, all possible positions and box sizes need to be evaluated, which is calculation intensive. For this reason, originally, the R-CNN method was developed, which provides region suggestions. Single-shot detectors are an improved detection algorithm and are designed to skip making region suggestions and solve classification and regression tasks in one step, making them more efficient and faster. The two best known algorithms are YOLO (You Look Only Once) and SSD (Single Shot Detector) (Liu et al., 2016). The latter is used in this research, because it is the most accurate and fastest (Poirson et al., 2016). Training performance can be increased considerably by applying transfer training based on a model with many parameters that was pre-trained for a different task (Howard and Gugger, 2020). A large number of pre-trained models can be downloaded via the internet. Each model has its specific architecture, among others they differ in number of layers and filter sizes. They can also vary in the type of data that was used to train them. The selection of the backbone model determines the architecture of the model used for the training. Many experiments were carried out to determine the best backbone model. In the presented research, we evaluated the ResNet18, DensNet121 and VVG_11 models for training and inference. The learning rate is an important parameter during this phase. It determines the size of the adaptation of the weights during one pass of the training data through the network. If the learning rate is too low, the optimal solution for the model may not be found, if it is too high the model may take too long to converge, and training will never end. The best learning rate can be found manually, but in ArcGIS Pro, it is possible to use fast.ai's learning rate finder, which suggests an optimal learning rate. The maximum number of epochs is used to specify the maximum number of times the training data is used to adapt the weights of the network and therefore limits the training time. The batch size is a hyperparameter that defines the number of samples to work through before updating the model parameters. Other parameters are the number of grids cell in which the image is divided, and the size and ratio parameters for the detection boxes.

Once the model and hyperparameters were determined, they were stored and used for inference on the image of the complete area. During the inference, parameters for the confidence threshold and non maximum suppression (nms) need to be determined. The confidence threshold is the minimum confidence that is required for an object to be stored. For example, in our research a setting of 0.5 was used, which means that the algorithm must be at least 50% confidence that it has found a plastic greenhouse. Objects with a lower confidence are ignored for further processing. The output of the inference gives many overlapping boxes with different confidences. The nms parameters is used to remove overlapping boxes of the same objects and to determine how much overlap is allowed between adjacent

boxes. The plastic greenhouses in the study area are located close to each other, therefore a 40% overlap setting was used in this research.

Each step of the workflow to detect plastic greenhouses was performed using the Deep learning toolset of ArcGIS Pro 2.7. The separate tools for the creation of samples, export of image chips, and the training and inference of the model provide a user-friendly interface to the complex algorithms that are used to detect the objects (ESRI, 2021). To be able to use the toolset, an ArcGIS Pro license is required, and an open-source deep learning environment based on python implementations of well-known machine learning libraries like Keras – TensorFlow, PyTorch, fastai and Scikit-learn needs to be installed. Although, it is possible to train the models using a CPU, it is highly recommended to use a GPU.

RESULTS AND DISCUSSION

For the training in total 2352 plastic greenhouses were digitized with an average size of 200 m². These were used to create 6228 partly overlapping image chips. Each image chip had a size of 256 x 256 pixels and the average number of greenhouses per chip was 5. Examples of image chips are shown in Figure 3.

Many settings of the hyperparameters for the training were tested, and the optimum combination was reached with 50 iterations, a batch size of 8, grid values of 4, 2, and 1, zoom values of 0.7, 1.0, 1.3 and [1,1], [1, 0.5], [0.5, 1] for the ratio values. The learning rate was set to automatic, and 20% of the data was used for model validation. During the training phase, numerous experiments have been executed to determine the architecture of the model and the values of the hyperparameters. All training and inference tests were executed on a PC with an Intel Core I5, 8th generation processor, 8 GB RAM and a Geforce GTX 1050 graphics card. First, backbone models ResNet18, ResNet34, ResNet50 and DenseNet121 were tested as architectures for the training.

A subset of the result of the training with the ResNet18 backbone model is shown in Figure 4.

Obviously, the larger the model, the longer the time required for training. In our case, it took 16h 31m to train the large DenseNet121 model, while the ResNet18 model with the same parameters took only 3h 24m.



Fig. 3 Image chips with multiple samples of plastic greenhouses



Fig. 4 Training results of model with Resnet18 backbone, at the left the input sample is shown, while the right image shows the detected result

The main problem with the larger backbone models was overfitting, therefore more samples were added to the training set by data augmentation, where the original training samples were rotated with 45° and 180° angles. DenseNet121, with 121 layers was finally trained with 6228 chips but did not provide better detection. This model gave the highest training accuracy of 79.1%. Also, the ResNet18 model was trained with the same sample set, acquired by data augmentation. The ResNet18 training with 50 iterations and a batch size of 8 and resulted in an average accuracy of 80.1%. Figure 5 shows the training and validation loss and clearly proves that the model is converging to an optimal solution.

In the first test, the VGG_11 backbone model was evaluated for the training. This model has only 11 layers and is therefore much faster to train. The training and validation loss plot is shown in Figure 6. The training took 90 minutes and the maximum number of epochs was set to 25, the number of samples was 5168.

To prevent overfitting of the VGG_11 model, different settings for dropout were tested. Dropout is a regularization method where randomly a part of the output of a layer is ignored and not read into the next layer. A value of 0.3 (30% of the data is ignored) gave the best result (Fig. 7).

The results of inference shown here are all executed on the same smaller test area. The main consideration for the selection of the area was that there are many plastic greenhouses that were not included in the training data set, so that it is possible to assess the quality of the inference. Additionally, it was important to select as many different types of greenhouses as possible (different in size, color and damage) to be able to evaluate the capabilities of the models to detect all plastic greenhouses. Figure 8 shows the inference results of the VGG_11 model.

The model successfully found the large majority of objects of interest. The red symbols indicate the 222 greenhouses that the model detected. Some objects were detected by this model, that are not a plastic greenhouse. These are indicated in black in Figure 8. For example, the model also recognized large tents that are very similar in shape to greenhouses. The average accuracy of the bounding boxes detected during the inference using VGG_11 in the test area is 79.2%.

The next backbone model used for inference was DenseNet121 (Fig. 9). The model showed a slight over fitting in the initial trials, but adaptation of the hyper

parameters and enlarging the training set showed that the model can detect greenhouses successfully. As a result of the inference using the DenseNet121 model, 230 plastic greenhouses were found in the test area. Comparing with the results of the VGG_11 model, it can be observed that the inference made with the DenseNet121 model is more accurate. DenseNet121 did not erroneously detected the large tents as plastic greenhouse unlike VGG_11 model (yellow circles in Fig. 9). The DenseNet121 model yielded an average of 79.1% for the accuracy value.

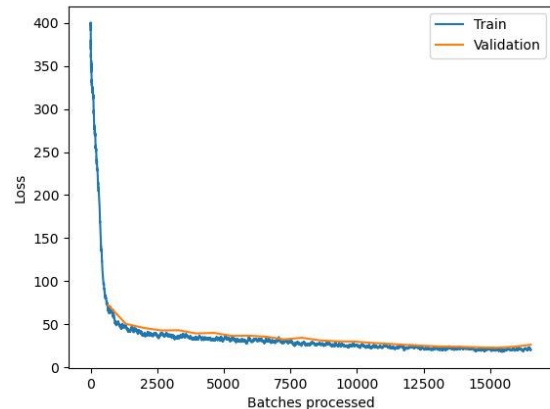


Fig. 5 Training and validation loss using the ResNet18 backbone and 2352 samples

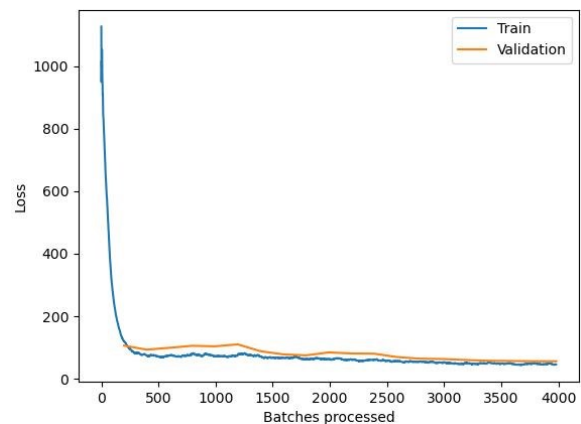


Fig. 6 Training and validation loss using the VGG_11 backbone

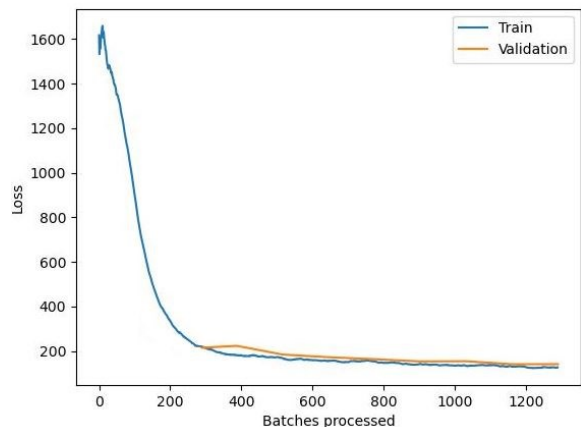


Fig. 7 Training and validation loss using the VGG_11 with dropout 0.3

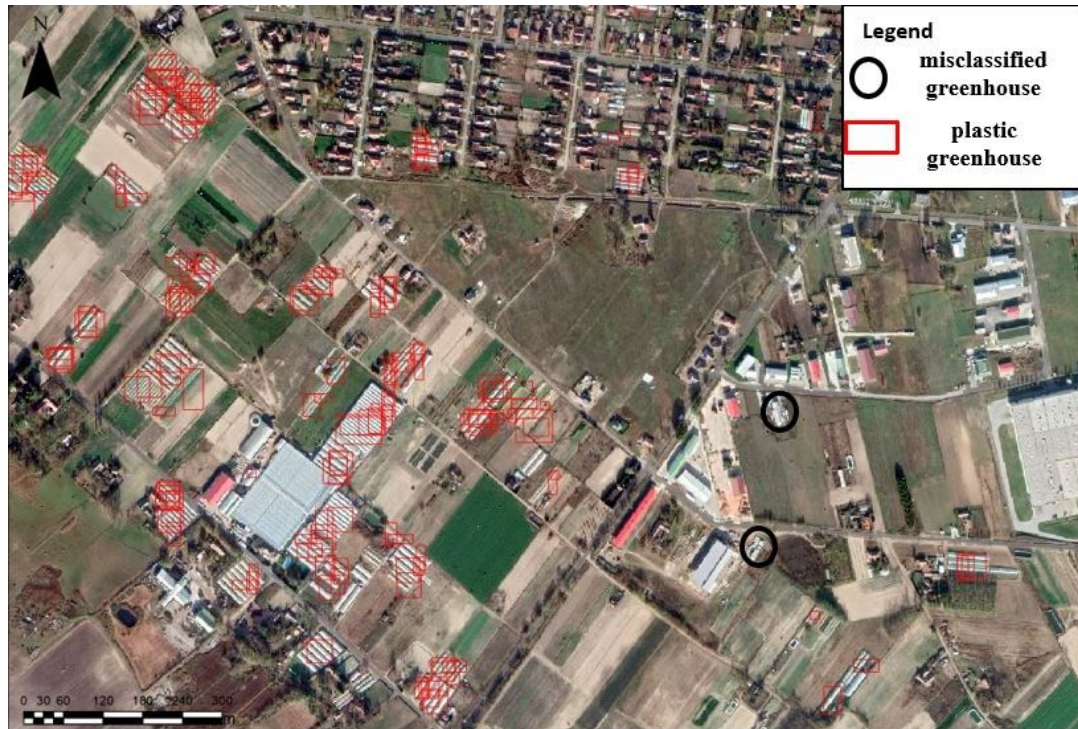


Fig. 8 Detection of plastic greenhouses on the test area using VGG_11

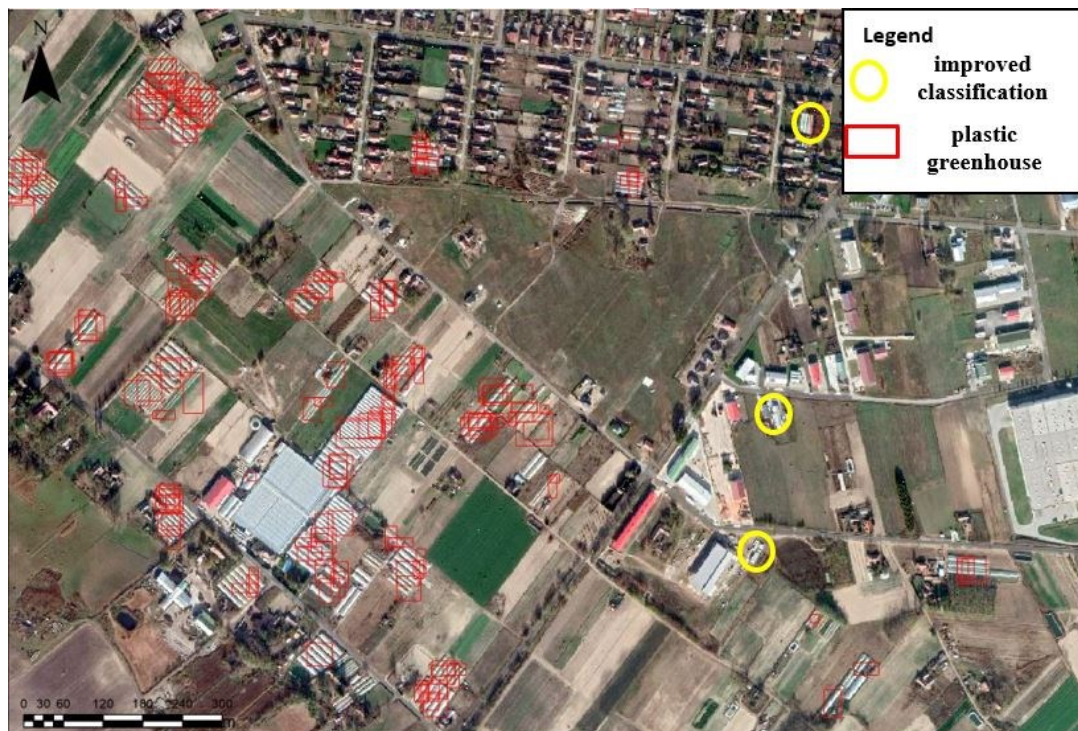


Fig. 9 Detection of plastic greenhouses on the test area using DenseNet121

Finally, the results of the ResNet18 model are presented in Figure 10. During the first training, a higher learning rate of 0.03 was used, but this did not give satisfying results. Then, the training was performed with the learning rate finder, suggesting far lower rates that provided a much better result. For the presented inference, the ResNet18 model trained with the suggested learning rates was used. The manual learning

rate setting resulted in a worse result, giving an average accuracy value of 78.3% and the model placed 225 bounding boxes during detection. The most accurate result was achieved by the ResNet18, with the optimal learning rate. It gave an average accuracy of 83.1% for the detected 232 bounding boxes. Improvement over the DenseNet121 result is indicated in yellow.

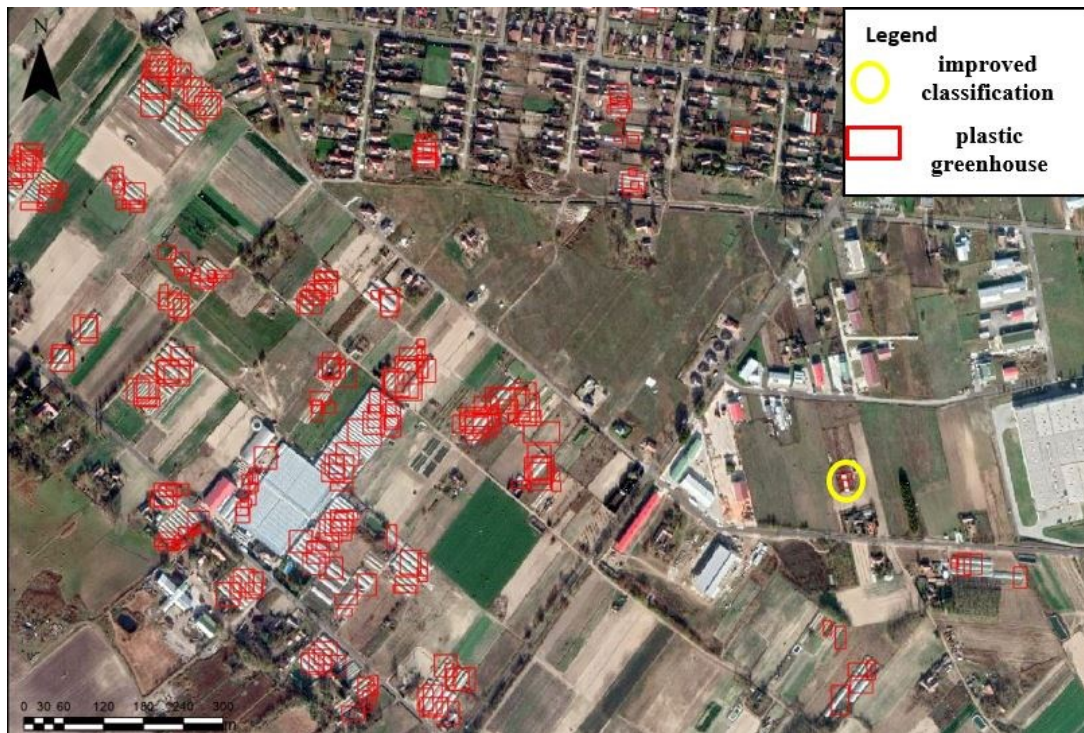


Fig. 10 Detection of plastic greenhouses on the test area using ResNet18

ArcGIS Pro provides a user-friendly environment to sophisticated deep learning algorithms. The dialogs hide the complexity of working with machine learning algorithms from the user, but to use the functionality optimally, it is required to have detailed knowledge of the hidden algorithms. This also helps to make efficient decisions on the many options that need to be specified during the creation of the training data, the training and the inference. Erroneous settings of the hyperparameters during training can easily result in models that never reach an optimal solution.

The selection of the backbone model is of decisive importance for the CNN processing. The complexity of the model, combined with setting for the learning rate, batch size and maximum number of epochs determine the accuracy of the results and the time required to train the model. Often, it might be more efficient to allow a slightly lower accuracy over much better performance.

The size and quality of the training data is another important condition for successful use of the deep learning functionality. Data augmentation is a powerful procedure to generate more training data without digitizing more examples. It also reduces the chance of overfitting, since more and different types of examples that may occur in other areas, are shown to the model.

Detection of objects in remotely sensed images provides a different result than the classification results that are common in remote sensing studies. For this reason, it is difficult to compare the results of traditional classifications and object detection as presented in our research. Traditional classifications provide one class label for each pixel in the image, while object detection aims to detect all objects of interest and their locations in the image. The metrics used for the estimation of the accuracy are also

different since the location of an object is not an output of classification algorithms.

In the presented research, RGB channel images downloaded from Google Earth are used as input data. The limited number of channels is a disadvantage compared to other data sets, when the data would be used for traditional classification. The models used as backbone for the convolutional neural network model are trained with three channel data though and are therefore particularly suitable as input data.

CONCLUSION

The presented research explores the possibilities for detection of plastic greenhouses in an agricultural area in the south east of Hungary using freely available high resolution satellite imagery and a convolutional neural network. The aim was to use state of the art deep learning techniques without the need to go into the depths of writing code, therefore we used the recent ArcGIS Pro deep learning implementation. This user-friendly environment allows to experiment with many setting for creation of the training data, the backbone model, the training, and the inference. It also provides feedback to the user on the success of the training. The connection between the deep learning algorithms and the GIS functionality of the software makes it easy to perform all steps in the detection of greenhouses in a spatial environment and display the results as maps and images.

The results of the inference show that - with careful selection of the network architecture and hyper parameters - it is possible to achieve high accuracy output maps. The calculation intensive experimentation

requires a high-performance computer. The use of pre-trained backbone models is essential.

We successfully tested a possibility to determine the number of plastic greenhouses and their locations based on freely available, high resolution data and we are optimistic that the technology will be used in the future for applications like statistics on agriculture, environmental impact studies and taxation purposes.

References

- Abadi, M., Agarwal, A., Barham, P., Brevdo, E., Chen, Z., Citro, C., Corrado, G.S., Davis, A., Dean, J., Devin, M., et al. 2015. TensorFlow: Large-scale machine learning on heterogeneous systems. Online available at: <https://arxiv.org/pdf/1603.04467.pdf>
- Agüera, F., Aguilar, M.A., Aguilar, F.J. 2008. Using texture analysis to improve per-pixel classification of very high resolution images for mapping plastic greenhouses. *ISPRS Journal of Photogrammetry and Remote Sensing* 63 (6), 635–646. DOI: 10.1016/j.isprsjprs.2008.03.003
- Agüera, F., Liu, G. G. 2009. Automatic greenhouse delineation from QuickBird and Ikonos satellite images. *Computers and Electronics in Agriculture* 66, 191–200. DOI: 10.1016/j.compag.2009.02.001
- Chollet, F. 2015. Keras. Online available at: <https://github.com/fchollet/keras>
- Davies, E.R. 2018. Computer Vision: Principles, Algorithms, Applications, Learning. Academic Press, 5th edition, 866 p. DOI: 10.1016/C2015-0-05563-0
- Ding, P., Zheng, Y., Deng, J-W., Jia, P., Kuijper, A. 2018. A light and faster regional convolutional neural network for object detection in optical remote sensing images. *ISPRS Journal of Photogrammetry and Remote Sensing* 141, 208–218. DOI: 10.1016/j.isprsjprs.2018.05.005
- ESRI 2021, ArcGIS Pro online help. Online available at: <https://pro.arcgis.com/en/pro-app/latest/tool-reference/image-analyst/an-overview-of-the-deep-learning-toolset-in-image-analyst.htm>
- Everingham, M., Gool, V., L., Williams, I., K., C., Winn, J., Zisserman, A. 2010. The PASCAL Visual Object Classes (VOC) Challenge. *International Journal of Computer Vision* 88, 303–338. DOI: 10.1007/s11263-009-0275-4
- Flood, N., Watson, F., Collett, L. 2019. Using a U-net convolutional neural network to map woody vegetation extent from high resolution satellite imagery across Queensland, Australia. *International Journal of Applied Earth Observation and Geoinformation* 82, 101897. DOI: 10.1016/j.jag.2019.101897
- Gallwey, J., Robiati, C., Coggan, J., Vogt, D., Eyre, M. 2020. A Sentinel-2 based multispectral convolutional neural network for detecting artisanal small-scale mining in Ghana: Applying deep learning to shallow mining. *Remote Sensing of Environment* 248: 111970. DOI: 10.1016/j.rse.2020.111970
- Goodfellow, I., Bengio, Y., Courville, A. 2016. Deep Learning. MIT Press, Online available at: <http://www.deeplearningbook.org>
- González-Yebra, Ó., Aguilar, A. M., Nemmaoui, A., Aguilar, J., F. 2018. Methodological proposal to assess plastic greenhouses land cover change from the combination of archival aerial orthoimages and Landsat data. *Biosystems Engineering* 175, 36–51. DOI: 10.1016/j.biosystemseng.2018.08.009
- Guo, Y., Xu, Y., Li, S. 2020. Dense construction vehicle detection based on orientation-aware feature fusion convolutional neural network. *Automation in Construction* 112. 103124. DOI: 10.1016/j.autcon.2020.103124
- Howard, J., Gugger, S. 2020. Fastai: A layered API for Deep Learning. *Information* 11 (2), 108. DOI: 10.3390/info11020108
- Jiang, B., Ma, X., Lu, Y., Li, Y., Feng, L., Shi, Z. 2019. Ship detection in spaceborne infrared images based on Convolutional Neural Networks and synthetic targets. *Infrared Physics & Technology* 97, 229–234. DOI: 10.1016/j.infrared.2018.12.040
- Kattenborn, T., Leitloff, J., Schiefer, F., Hinz, S. 2021. Review on Convolutional Neural Networks (CNN) in vegetation remote sensing. *ISPRS Journal of Photogrammetry and Remote Sensing* 173, 24–49. DOI: 10.1016/j.isprsjprs.2020.12.010
- Koc-San D. 2013. Evaluation of different classification techniques for the detection of glass and plastic greenhouses from WorldView-2 satellite imagery. *Journal of Applied Remote Sensing* 7 (1): 073553. DOI: 10.1117/1.JRS.7.073553
- LeCun y., Boser, B., Denker, S. J., Henderson, D., Howard, E. R., Hubbard, W., Jackel, D. L. 1990. Handwritten Digit Recognition with a Back-Propagation Network. pp. 396–403.
- Liu, W., Anguelov, D., Erhan, D., Szegedy, C., Reed, S., Fu, Y. C., Berg, C. A. 2016. SSD: Singe Shot Multibox Detector. *European Conference on Computer Vision* 2016, 21–37. DOI: 10.1007/978-3-319-46448-0_2
- McCarthy, J., Minsky, I. M., Rochester, N., Shannon, E., C. 1955. A proposal for the Dartmouth summer research project on artificial intelligence. *AI Magazine*, 27 (4), pp. 12–14. DOI: 10.1609/aimag.v27i4.1904
- Mezősi, G. 2011. Magyarország természetföldrajza, (Physical geography of Hungary) Academic Press, Budapest, pp. 393.
- Michie, D. 1968. „Memo” Functions and Machine Learning. *Nature* 218 (5136), 19–22. DOI: 10.1038/218019a0
- Müller, B., Reinhardt, J., Strickland, M. T. 1995. Neural Networks: An Introduction. Springer, Berlin, pp. 307.
- Nemmaoui, A., Aguilar, J. F., Aguilar, A. M., Qin, R. 2019. DSM and DTM generation from VHR satellite stereo imagery over plastic covered greenhouse areas. *Computer and Electronics in Agriculture* 164, 104903. DOI: 10.1016/j.compag.2019.104903
- Nilsson, N., J. 1980. Principles of artificial intelligence. Morgan Kaufmann, California, pp. 475.
- Novelli, A., Aguilar, A.M., Nemmaoui, A., Aguilar, J. F., Tarantino, E. 2016. Performance evaluation of object based greenhouse detection from Sentinel-2 MSI and LANDSAT 8 OLI data: A case study from Almería (Spain). *International Journal of Applied Earth Observation and Geoinformation* 52, 403–411. DOI: 10.1016/j.jag.2016.07.011
- Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., Killeen, T., Lin, Z., Gimelshein, N., Antiga, L., Desmaison, A., Köpf, A., Yang, E., DeVito, Z., Raison, M., Tejani, A., Chilamkurthy, S., Steiner, B., Fang, L., Bai, J., Chintala, S. 2019. PyTorch: An Imperative Style, High-Performance Deep Learning Library. Cornell University. Online available at: <https://arxiv.org/pdf/1912.01703v1.pdf>
- Pedregosa, F., Varoquaux, G., Gramfort, A., Michael, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., Duchesnay, É. 2011. Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research* 12, 2825–2830. Online available at: <https://arxiv.org/pdf/1201.0490.pdf>
- Pi, Y., Nath, D. N., Behzadan, H. A. 2020. Convolutional neural networks for object detection in aerial imagery for disaster response and recovery. *Advanced Engineering Informatics* 43, 101009. DOI: 10.1016/j.aei.2019.101009
- Poirson, P., Ammirato, P., Fu, C. Y., Liu, W., Košecák, J., Berg, C. A. 2016. Fast single shot detection and pose estimation. *Fourth International Conference on 3D Vision (3DV)*, Stanford, CA, USA, 2016, pp. 676–684, DOI: 10.1109/3DV.2016.78
- Rai, K. A., Mandal, N., Singh, A., Singh, K. K. 2020. Landsat 8 OLI Satellite Image Classification using Convolutional Neural Network. *Procedia Computer Science* 167, 987–993. DOI: 10.1016/j.procs.2020.03.398
- Schiefer, F., Kattenborn, T., Frick, A., Frey, J., Schall, P., Koch, B., Schmidlein, S. 2020. Mapping forest tree species in high resolution UAV-based RGB-imagery by means of convolutional neural networks. *ISPRS Journal of Photogrammetry and Remote Sensing* 170, 205–215. DOI: 10.1016/j.isprsjprs.2020.10.015
- Simon, A., H. 1995. Artificial intelligence: an empirical science. *Artificial Intelligence* 77 (1), 95–127. DOI: 10.1016/0004-3702(95)00039-H
- Virnodkar, S.S., Pachghare, C.V., Jha, K.S. 2020. CaneSat dataset to leverage convolutional neural networks for sugarcane classification from Sentinel-2. *Journal of King Saud University – Computer and Information Sciences*. DOI: 10.1016/j.jksuci.2020.09.005 (in press)
- Watanabe, S., Sumi, K., Ise, T. 2018. Using deep learning for bamboo forest detection from Google Earth images. *bioRxiv* 351643, DOI: 10.1101/351643
- Wu, C., Deng, J. S., Wang, K., Ma, L. G., Tahmassebi, A. R. S. 2016. Object-based classification approach for greenhouse mapping

- using Landsat-8 imagery. *International Journal of Agricultural and Biological Engineering* 9, 79–88. DOI: 10.3965/j.ijabe.20160901.1414
- Yang, D., Chen, J., Zhou, Y., Chen, X., Chen, X., Cao, X. 2017. Mapping plastic greenhouse with medium spatial resolution satellite data: Development of a new spectral index. *ISPRS Journal of Photogrammetry and Remote Sensing* 128, 47–60. DOI: 10.1016/j.isprsjprs.2017.03.002
- Yang, G., Xu, R., Chen, Yi., Wu, Z., Du, Y., Liu, S., Qu, Z., Guo, K., Peng, C., Chang, J., Ge., Y. 2021. Identifying the greenhouse by Google Earth Engine to promote the reuse of fragmented land in urban fringe. *Sustainable Cities and Society* 67, 102743 DOI: 10.1016/j.scs.2021.102743
- Zhang, D., Pan, Y., Zhang, J., Hu, T., Z, J., Li, N., Chen, Q. 2020. A generalized approach based on convolutional neural networks for large area cropland mapping at very high resolution. *Remote Sensing of Environment* 247, 111912. DOI: /10.1016/j.rse.2020.111912