# TRANSFER LEARNING IN THE CLASSIFICATION OF SATELLITE IMAGES SHOWING AMAZON RAINFOREST

K. Karwowska\*

\*Department of Imagery Intelligence, Faculty of Civil Engineering and Geodesy, Military University of Technology, 00-908 Warsaw, Poland - kinga.karwowska@wat.edu.pl

**KEY WORDS:** convolutional neural networks; scene classification; urban land cover classification; high-resolution satellite imagery; image analysis; rainforest.

### **ABSTRACT:**

In recent years, we have been dealing with the dynamic technological progress of the space sector, which allows for the observation of the Earth with better temporal, spatial and spectral resolution. The increasing availability of satellite data has contributed to the development of data processing algorithms. Thanks to the use of digital image processing methods and deep neural networks, it is possible to perform automatic image classification, segmentation or detection and recognition of objects on the images. This article presents the methodology that allows to accelerate the classification process of satellite images representing the Amazon rainforest based on the Transfer Learning method. Additionally, the influence of the choice of optimization, i.e. the network weight estimation strategy, on the classification of objects was checked. In order to verify the method, an additional raster image classifier was created on the basis of Lidar data. Research shows that the transfer learning method allows the preparation of an image classifier based on a small database (less than 100 images representing one class). The network training process can be shortened to a few minutes.

### 1. INTRODUCTION

Nowadays, we have observed significant technological development. It allowed increasing the possibilities of imaging the Earth. More and more sensors are placed in the orbits, allowing for the acquisition of images with high spatial resolution and in many spectral channels. Such possibilities are used in many areas, including agriculture (Efremova et al., 2021, p. 1; Metzger et al., 2021; Ru et al., 2021) and environmental protection (Sun et al., 2016; Zhong et al., 2015). Additionally, considering the computational capabilities of the working units, the processing of images acquired by satellites is much faster, which enabled the development of many algorithms. For several years, remote sensing development

solutions have been observed using classical image processing and deep learning methods. Deep learning algorithms, and more specifically convolutional networks, have been known to scientists for almost a quarter of a century. However, initially, they were not used due to limited computing power. Now, when we have a high-performance graphical processor at our disposal, and the possibility of using virtual machines, the development of deep learning algorithms in obtaining data from satellite imagery has significantly accelerated. Due to the developed solutions, it is possible to conduct automatic classification, object detection, segmentation, and create new images.

This work focuses on answering the following research question:



Figure 1. Network architecture example.

a) Is there a method to accelerate the training of neural networks for the classification of satellite images representing parts of the Amazon?

b) Does the developed methodology for training neural networks also work for other databases?

c) Does the application of Adam optimization allow to obtain the best classification results?

#### 2. METHODOLOGY

The proposed research methodology presents a method of training satellite image classifiers using deep neural sieves. This solution will speed up the process of classification of satellite images on the basis on the location of objects in the image and the correlation between them. This procedure is possible thanks to the use of weave layers. These layers contain kernels estimated during the training process, extracting the features of the objects in the image (Figure 1). The resulting images with object features are called feature maps. As a result of multilayer processing, a tensor is created that represents the characteristics of a given data set. In the proposed methodology, not only the existing network architectures were not used, but also the weights calculated on the basis of the ImageNet data set, responsible for the extraction of satellite image features (Figure 2)

Thanks to this approach to the problem, it is possible to prepare the classifier in a few minutes, using small databases.



Figure 2. Network architecture example.

Additionally, the influence of the network training method on the quality of classification was investigated. All calculations included in this paper were made on a PC equipped with an Nvidia Titan graphics card.

To train a classifier using neural networks, it is necessary to prepare: a training database and a model of the classifier network. Then, on their basis, it is possible to carry out network training, that is, to estimate the weights-masks between the layers of the model. The correctness of the classifier's operation is greatly influenced by the size and quality of the training database, because even when the best training strategy is selected, the best model will be designed, and the database will be too small (or will contain errors), the classifier will not work correctly (e.g. with due to network overfitting or incorrect inference). To eliminate the network overfitting error and to ensure the correct classification of various data obtained in different conditions, an extensive database should be created, containing images of different lighting, obtained at other times of the year or with different lighting. Moreover, one heuristic says that the training database should contain more images than the parameters trained during network training. In the case of creating satellite image classifiers, creating such large databases is very difficult due to the high cost of satellite data and difficult accessibility. Therefore, a solution was developed to adapt the estimated network weights to the classification of satellite images. Among the conducted research, it is popular to use pretrained networks and then train them on the basis of their databases (Risojević and Stojnić, 2021; Yuan et al., 2022). In this solution, all network parameters are trained. Unfortunately, few solutions use the transfer learning method to accelerate network training significantly. (Alem and Kumar, 2021; Pires de Lima and Marfurt, 2020). The article uses images of the Amazon rainforest. Based on this database, a number of solutions have been prepared, but the proposed methodologies for building classifiers are much more time-consuming, and the solutions obtained by the authors only slightly exceed the results obtained during the research (Chandak et al., n.d.; Kudli et al., n.d.).

### 2.1. Database

"Planet: Understanding the Amazon from Space database" (Planet, 2015) a part of the Planet provided by Kaggle was used to conduct the research. The database was created on the basis of images obtained by Planet Labs PBC. The data set used consisted of 4,849 randomly selected images composed of three channels (red, green, blue) with dimensions of 224x224 pixels, presenting fragments of scenes depicting Brazil, Peru, Uruguay, Colombia, Venezuela, Guyana, Bolivia, and Ecuador. Sample



Figure 3. Examples of training database images.

The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, Volume XLIII-B3-2022 XXIV ISPRS Congress (2022 edition), 6–11 June 2022, Nice, France



Figure 4. Network training diagram.

images are shown in Figure 3. Each of the images has been assigned at least one of the following classes: primary, clear, agriculture, road, water, partly cloudy, cultivation, habitation, haze, cloudy, bare ground, selective logging, artisanal mine, blooming, slash burn, conventional mine, blow down artisanal. The data set was divided into three data sets - training (70% of all images), validation (20%) and test (10%).

### 2.2. Evaluation metrics

The most popular classifier evaluation metrics used in the remote sensing and computer vision environment were used to assess the correctness of the classification of the trained models (Hossin and M.N, 2015). The first metric is Recall (True Positive Rate), which defines the ratio of true positive (TP) images to the sum of true positive and false negative (FN) images. Another metric is the average precision that determines the number of TP images to the sum of false and true positive images. The F1/F2 metric (also known as the F-beta score) takes the previous two results as their weighted average. For F1 score (beta=1.0), but for F2 score (beta=2.0). On the other hand, accuracy determines the percentage of well-classified image cases. Another assessment metric is logarithmic loss (also known as Cross-Entropy Loss, which is calculated as the product of the ground true class and the logarithm of the prediction class. One of the most reliable quality measures is AUC (area under the curve). It determines the area under the Receiver Operating Characteristics (ROC) curve, representing the relationship between True Positive Rate and True Negative Rate.

# 2.3. Solution

The solution to this problem is the use of the transfer learning method. The purpose of this solution is to use network weights estimated on the basis of large data sets such as ImageNET to prepare classifiers of satellite imagery or objects located on them. The use of network pre-learning, the so called Transfer learning on a freely selectable set of training data is possible because the last layers to be trained play a main role in object classification. This solution significantly reduces the number of trained parameters, and thus significantly shortens the network training (from a few hours to several minutes), and also significantly improves the ability to generalize the network. The most popular network models were used to train Amazon forest classifiers - VGG16 (Simonyan and Zisserman, 2015), VGG19 (Simonyan and Zisserman, 2015), Xception (Chollet, 2017), ResNet50 (He et al., 2015), MobileNet (Howard et al., 2017), MobileNetV2 (Howard et al., 2017). The studied networks differ in their network architecture, i.e. the arrangement of layers and the adopted hyperparameters. For the purposes of the research, pre-trained models were used based on the ImageNET database (Deng et al., 2009), allowing the classification of 1000 classes (e.g. alligator lizard, quail or crane). The images on which the networks were trained do not contain images from the air or space ceilings. Transfer learning was used to train the network, with the signals from the last

was used to train the network, with the signals from the last layers responsible for interpretation cut out. They were replaced with four layers: Flatten, Dense, Dropout and another Dense layer consisting of 17 neurons (the number of neurons is equal to the number of classes accepted) (Figure 4). Only the parameters located at the connections between these layers were calculated during network training. The layers in front of the flattened layer are responsible for the extraction of features in the image, on the basis of which the last layers (which are

Model	Optimization Algorithm	F-beta	Recall	Ave Precision	Accuracy	Log Loss	AUC	Error
VGG16	ADAM	0.85	0.90	0.31	0.36	15.64	0.78	14.52
Xception	SGD	0.88	0.91	0.36	0.44	13.63	0.71	11.62
ResNet50	RMSProp	0.66	0.98	0.18	0.01	8.44	0.54	33.50
VGG19	Adamax	0.84	0.88	0.30	0.37	18.19	0.66	15.63
MobileNet	Adagrad	0.88	0.91	0.34	0.43	14.91	0.70	12.45
MobileNetV2	SGD	0.87	0.90	0.37	0.46	15.76	0.70	12.79

Table 2. Operation correctness of the best classifiers after the first part of the research.

trained) determine the probability of the image belonging to one of the defined classes. The weights that were calculated on the basis of ImageNET databases (not containing satellite or aerial images) and the main network model (layers in front of the flatten layer) are responsible for the extraction of object features (edges, pixel groups with similar DN values, edge intersections) and generalized image features on the basis of which the trained part of the network predicts the probability of the image belonging to each of the defined classes (which is determined by the number of neurons in the last Dense layer).

Due to the application of the Transfer learning method, the number of trained parameters was significantly reduced (Table 1). Among the tested network models, the Xception network has the least trained parameters (only 14.30% of parameters are trained).

The same training procedure was adopted for each of the networks - only the layers that were added to the model were trained (the weights of the network trained on the ImageNET set were assumed for the remaining layers). In addition, these networks were trained only for 30 epochs, which allowed to determine the speed of learning the network and the correctness of operation after such a short training time. Additionally, the influence of the optimization network training method on the prediction results was examined by using various optimizers: ADAM (Kingma and Ba, 2017), SGD (Sutskever et al., 2013), RMSProp (Khan et al., 2017), Adagrad (Duchi et al., 2011), Adadelta (Zeiler, 2012), Adamax (Kingma and Ba, 2017), Nadam (Dozat, 2016). An additional method to accelerate the network learning process is the selection of an appropriate model parameter learning strategy. The optimization task is to

Model	Optimization Algorithm	F-beta	AUC	
Xception	SGD	0.90	0.86	
MobileNet	Adagrad	0.89	0.80	
MobileNetV2	SGD	0.89	0.80	

 Table 3. Operation correctness of the best classifiers after the second part of the research.

find the extreme of the global objective function. In the case of classifiers using convolutional networks, the optimization task is to match the weights (kernels) between the layers of the network in such a way that the image classification error is as small as possible. The correctness of the classification of the trained models was assessed on the based on the most popular metrics described above. Table 2 shows the best results of the trained models (taking into account the optimization algorithms used). The best result is marked in green, and the worst – is in red.

In the second part of the research, only one examined architecture model was trained. The ones that presented the best results in the first part of the study were selected. Each of the networks was trained for the next 50 epochs with a lower value of the learning speed parameter, which significantly improved the correctness of the classifiers. Table 3 shows the results of the F-beta and AUC metrics for the trained classifier. The classifier's training allowed for a significant improvement in the quality of its work, as evidenced by the increase in the value of the F-beta metric and a significant increase in the area under the ROC characteristic curve.

# 2.4. Method check for another database

In order to emphasize the potential of the presented methodology, an object classifier was built on rasters based on



Figure 5. Classes and number of images in new dataset.

LIDAR measurement data. For the purposes of the training, an own database was created, the classes and number of images of which are shown in Figure 5.

The Xception network architecture (from SGD Optimization Algorithm) was used for network training. As in the abovedescribed methodology, only the part of the model that is responsible for the classification was trained. In the preparation of this classifier, the training strategy was changed, the number of training epochs of the model was not defined. In this case, a parameter has been added to stop the model when F-beta stops growing for validation data. After 15 epochs the F-beta value stopped improving and the network training stopped. The correctness of the classifier operation is presented in Table 4. Additionally, the Confusion matrix was assessed (in the case of part of the Planet: Understanding the Amazon from Space

Classification model- Xception (SGD - optimization algorithm)					
F-beta	0.93				
Recall	0.93				
Ave Precision	0.75				
Accuracy	0.93				
Log Loss	3.51				
AUC	0.94				
Error	6.73%				

Table 4. Operation correctness of classifier Xception.

		CLASS						
		0	1	2	3	4	5	6
CLASS	0	16	0	0	0	0	0	0
	1	0	21	0	0	0	0	0
	2	0	0	20	0	0	1	0
	3	0	0	0	27	0	1	0
	4	0	0	0	0	24	0	0
	5	0	0	0	3	1	71	0
	6	0	1	0	0	0	0	3

 Table 5. Confusion matrix.

database, the implementation of the error matrix was difficult because one image could have more than one label) (Table 5). As shown in Tables 4 and 5, the model is very good at classifying objects after 15 epochs (3 minutes of training). The classifier sometimes occurs mistakes in the classification of electric poles (classifies them as wind farms).

# 3. DISCUSSION

The article presents the methodology of creating satellite image classifiers based on deep neural networks. In order to significantly accelerate the classifier model training process, the transfer learning method was used, which uses the trained model weights on the basis of ImageNET databases. Tables 2 and 3 present the results of the evaluation metrics for the tested network models, including the optimization method used. Comparing the obtained results with the results of the winners of the Kaggle competition, it can be noticed that the value of the F-beta metric presented by the competition participants (the result of the competition leader is F-beta = 0.93317) is slightly higher than the value of the F-beta error calculated for the Xception model (SGD Optimization Algorithm) after 80 epochs. Additionally, it should be noted that the model was trained using a small database (3,395 randomly selected images participated in the training, the remaining images were used to assess the classifier's work). In addition, in order to verify the methodology, an additional classifier was trained, allowing the classification of selected objects on the basis of raster images created on the basis of Lidar data. Table 4 presents the classifier performance evaluation metrics. What is more, Table 5 presents a matrix of errors that shows that the prepared classifier is only wrong six times out of 188 trials, which is only 3%.

# 4. CONCLUSION

The research focused on checking whether it was possible to use models that had been pre-trained from ImageNET images to classify satellite images. In addition, it was checked whether the use of the most popular Adam optimization training method is always the best solution. As shown by the results of the first part of the study, a fragment of which is presented in Table 1, after 30 epochs (whose training lasted less than 5 minutes), the classification accuracy in many cases exceeds 85%, and the values of the other qualitative metrics, including AUC, show very high potential of the method. Additionally, it can be seen that the use of the most popular optimization method - Adam allows for obtaining good results, but in most cases they are not the best. Among the tested classifiers, the best classification results were demonstrated by the Xception architecture [using the Stochastic Gradient Descent (SGD) optimizer], and the ResNET50 architecture demonstrated the worst.

The conducted research shows that the use of network weights trained on any data set can significantly shorten the preparation of the satellite image classifier and improve its generalization abilities. Moreover, considering the time needed to train such a classifier (which in the case of using a graphics processor, for this database it takes less than 10 minutes), this solution significantly exceeds the classic image classification methods. The results obtained during the research show that applying the transfer learning method to training the Xception model (using SGD optimization) allows to achieve slightly worse results than the winners of the Kaggle competition, where the value of the F-beta score differs by approx. 0.013. Further questions were raised due to the obtained results regarding the use of the transfer learning method of training algorithms for detection and classification of objects on satellite imagery. Future researches are planned to be devoted to this subject.

# REFERENCES

Alem, A., Kumar, S., 2021. Transfer Learning Models for LandCover and Land Use Classification in Remote Sensing Image.Appl.Artif.Intell.0,1–19.https://doi.org/10.1080/08839514.2021.2014192

Chandak, S., Chitters, V., Honnungar, S., n.d. Understanding the Amazon Rainforest from Space using CNNs 9.

Chollet, F., 2017. Xception: Deep Learning with Depthwise Separable Convolutions. ArXiv161002357 Cs.

Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K., Fei-Fei, L., 2009. ImageNet: A large-scale hierarchical image database, in: 2009 IEEE Conference on Computer Vision and Pattern Recognition. Presented at the 2009 IEEE Conference on Computer Vision and Pattern Recognition, pp. 248–255. https://doi.org/10.1109/CVPR.2009.5206848

Dozat, T., 2016. INCORPORATING NESTEROV MOMENTUM INTO ADAM 4.

Duchi, J., Hazan, E., Singer, Y., 2011. Adaptive Subgradient Methods for Online Learning and Stochastic Optimization. J. Mach. Learn. Res. 12, 2121–2159.

Efremova, N., Seddik, M.E.A., Erten, E., 2021. Soil Moisture Estimation using Sentinel-1/-2 Imagery Coupled with cycleGAN for Time-series Gap Filing. IEEE Trans. Geosci. Remote Sens. 1–1. https://doi.org/10.1109/TGRS.2021.3134127

He, K., Zhang, X., Ren, S., Sun, J., 2015. Deep Residual Learning for Image Recognition. ArXiv151203385 Cs.

Hossin, M., M.N, S., 2015. A Review on Evaluation Metrics forData Classification Evaluations. Int. J. Data Min. Knowl.Manag.ProcessProcess5,01–11.https://doi.org/10.5121/ijdkp.2015.5201

Howard, A.G., Zhu, M., Chen, B., Kalenichenko, D., Wang, W., Weyand, T., Andreetto, M., Adam, H., 2017. MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications. ArXiv170404861 Cs.

Khan, M.E., Liu, Z., Tangkaratt, V., Gal, Y., 2017. Vprop: Variational Inference using RMSprop. ArXiv171201038 Cs Stat.

Kingma, D.P., Ba, J., 2017. Adam: A Method for Stochastic Optimization. ArXiv14126980 Cs.

Kudli, S., Qian, S., Pastel, B., n.d. Kaggle Competition: Understanding the Amazon from Space 10.

Metzger, N., Turkoglu, M.O., D'Aronco, S., Wegner, J.D., Schindler, K., 2021. Crop Classification Under Varying Cloud Cover With Neural Ordinary Differential Equations. IEEE Trans. Geosci. Remote Sens. 1–12. https://doi.org/10.1109/TGRS.2021.3101965

Pires de Lima, R., Marfurt, K., 2020. Convolutional Neural Network for Remote-Sensing Scene Classification: Transfer Learning Analysis. Remote Sens. 12, 86. https://doi.org/10.3390/rs12010086 Planet, 2015. Planet: Understanding the Amazon from Space [WWW Document]. URL https://kaggle.com/c/planetunderstanding-the-amazon-from-space (accessed 1.10.22).

Risojević, V., Stojnić, V., 2021. The Role of Pre-Training in High-Resolution Remote Sensing Scene Classification. ArXiv211103690 Cs.

Ru, C., Duan, S.-B., Jiang, X.-G., Li, Z.-L., Jiang, Y., Ren, H., Leng, P., Gao, M., 2021. Land Surface Temperature Retrieval From Landsat 8 Thermal Infrared Data Over Urban Areas Considering Geometry Effect: Method and Application. IEEE Trans. Geosci. Remote Sens. 1–16. https://doi.org/10.1109/TGRS.2021.3088482

Simonyan, K., Zisserman, A., 2015. Very Deep Convolutional Networks for Large-Scale Image Recognition. ArXiv14091556 Cs.

Sun, G., Chen, X., Jia, X., Yao, Y., Wang, Z., 2016. Combinational Build-Up Index (CBI) for Effective Impervious Surface Mapping in Urban Areas. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 9, 2081–2092. https://doi.org/10.1109/JSTARS.2015.2478914

Sutskever, I., Martens, J., Dahl, G., Hinton, G., 2013. On the importance of initialization and momentum in deep learning, in: Proceedings of the 30th International Conference on Machine Learning. Presented at the International Conference on Machine Learning, PMLR, pp. 1139–1147.

Yuan, Y., Lin, L., Liu, Q., Hang, R., Zhou, Z.-G., 2022. SITS-Former: A pre-trained spatio-spectral-temporal representation model for Sentinel-2 time series classification. Int. J. Appl. Earth Obs. Geoinformation 106, 102651. https://doi.org/10.1016/j.jag.2021.102651

Zeiler, M.D., 2012. ADADELTA: An Adaptive Learning Rate Method. ArXiv12125701 Cs.

Zhong, B., Yang, A., Nie, A., Yao, Y., Zhang, H., Wu, S., Liu, Q., 2015. Finer Resolution Land-Cover Mapping Using Multiple Classifiers and Multisource Remotely Sensed Data in the Heihe River Basin. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 8, 4973–4992. https://doi.org/10.1109/JSTARS.2015.2461453