TRANSFERABILITY ASSESSMENT OF OPEN-SOURCE DEEP LEARNING MODEL FOR BUILDING DETECTION ON SATELLITE DATA

A. Spasov[1] *, D. Petrova-Antonova[1,2]

[1] GATE Institute, Sofia University, Sofia, Bulgaria – (angel.spasov, dessislava.petrova)@gate-ai.eu
[2] Faculty of Mathematics and Informatics, Sofia University, Sofia, Bulgaria – d.petrova@fmi.uni-sofia.bg

KEY WORDS: Computer Vision, Remote Sensing, Convolutional Neural Networks, Image Segmentation, Transfer Learning.

ABSTRACT:

A great number of studies for identification and localization of buildings based on remote sensing data has been conducted over the past few decades. The majority of the more recent models make use of neural networks, which show high performance in semantic segmentation for the purpose of building detection even in complex regions like the city landscape. However, they could require a substantial amount of labelled training data depending on the diversity of objects targeted, which could be expensive and time consuming to acquire. Transfer Learning is a technique that could be used to reduce the amount of data and resources needed by applying knowledge obtained solving one problem to another one. In addition, if open-source data and models are used, this process is much more affordable. In this paper, the Transfer Learning challenges and issues are explored by utilizing an open-sourced pre-trained deep learning model on satellite data for building detection.

1. INTRODUCTION

Analysis of remote sensing data serves numerous applications ranging from environmental monitoring through land use assessment to urban planning and design. Within the city domain, a great number of studies have focused on the localization and segmentation of city objects such as roads and buildings. This information is used for different applications such as planning of urban development and resource allocation. Due to the great amount of data to be analyzed automatic methods have been extensively explored and continuously developed. In recent years deep learning methods have shown superior performance for this task (Zhu, 2017), (Kang Zhao) and (W. Zhao, 2020).

Deep learning methods could, however, need a vast amount of labeled data, the production of which could be time-consuming and expensive. The affordability of automatic analysis of remote sensing data could be greatly improved, if open-source data and models and widely accessible satellite images could be used for development of such solutions. The potential of this approach is recognized and explored in some works (Maggiori, Tarabalka, Charpiat, & Alliez, 2017) and (Li, et al., 2020).

This paper aims to explore the Transfer Learning challenges and issues through transferability assessment of deep learning models. The presented study analyses the performance of an open-source model, pre-trained on open-source data, applied on a type of satellite images which are globally accessible on a regular basis.

The rest of the paper is organized as follows. Section 2 presents the approach for transferability assessment. Section 3 discusses the obtained results, Finally, Section 4 concludes the paper and gives directions for future work.

2. APPROACH FOR TRANSFERABILITY ASSESSMENT

The process of transferability assessment can be divided into several steps. First, a tool for achieving the selected task (in this case – buildings localization) has to be developed or selected. Second, data on which it will be tested need to be specified and collected. Third, the acquired data have to be pre-processed to satisfy requirements of the tool. Forth, the tool needs to be applied on the test data and the results have to be compared with its performance on data for which it was developed.

2.1 Trained Model Selection

To assess the transferability potential of a modern approach for segmentation of urban objects, open-source solutions and datasets have been assessed based on 4 main criteria.

1. **Availability and readiness for use:** the model has to be publicly available and relatively easy to use.
2. **Possibility for refinement:** the model and the way it is provided had to allow for further improvements and adjustments such as further training and parameter tuning.
3. **Performance:** the model has to provide state-of-the art performance in terms of segmentation results. It furthermore, has to bear the potential to perform well on unseen urban areas.
4. **Data consuming:** the model has to perform on standard data that is relatively accessible and also suitable for the current task – building identification and localization.

Based on these criteria an open-source solution provided by Solaris (an open-source machine learning pipeline for geospatial imagery) has been selected (CosmiQ Works, 2021). The aim of the Solaris platform and its Python library is to *"accelerate research in the geospatial computer vision domain by providing efficient implementation of common utility functions"* and allow for easy use of existing geospatial computer vision models. As such the solution was relatively convenient to run only with few adjustments needed. For support the platform provides extensive documentation and tutorials. Furthermore, the models could be further trained with new data, whereby configuration files support the adjustments of the models. The model applied in the current study is an adjusted version of a winning solution from the *SpaceNet 2: Building Detection v2* challenge (SpaceNet LLC, n.d.). It has a U-Net architecture with a VGG16 encoder part and a decoder part consisting of five upsampling blocks with bilinear

---

* Corresponding author

interpolation, shown in Figure 1, (Ronneberger, Fischer, and Brox, 2015). The loss used is a 4:1 combination of BCE (binary cross entropy) and Jaccard Loss with an Adam optimizer. The model comes with parameter weights pretrained on a diverse dataset consisting of high-resolution satellite imagery from WorldView 3 satellite with around 300.000 buildings across four cities. These images are suitable for building footprint localization and are widely available for different geographical locations on a regular basis.
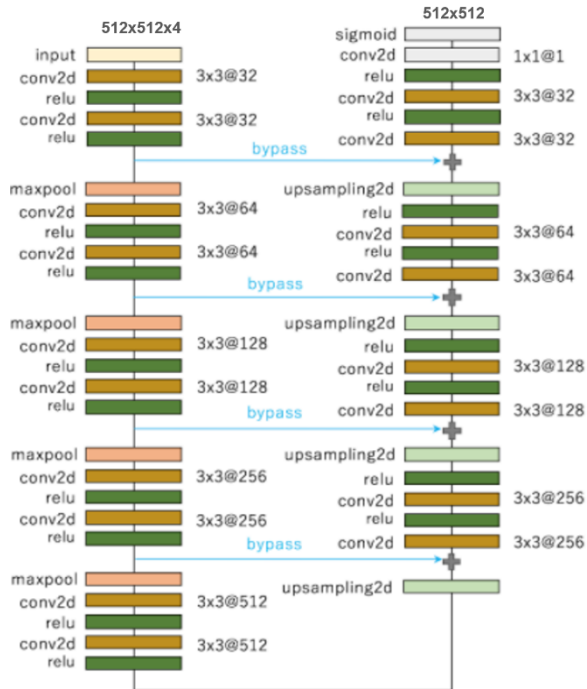

Figure 1. Model Architecture (SpaceNetChallenge, 2017).

2.2 Data Collection and Exploration

Once a pre-trained deep learning approach is selected, the next step is to analyze the data on which it was trained on. The model consumes 16-bit images with RGB + NIR channel. They have been pan-sharpened to a ground sampling distance of 0.3m. For each city, the images are single-strapped with a slight of-nadir angle apparent with zero cloud coverage, however, with different sun elevation across the four cities. Suitable test data for transferability assessment are ones showing as similar characteristics as the training data. Figure 2 sumarizes the requirements to the test data.

Area of Interest: Sofia – District "Lozenets" of Sofia, Bulgaria
Type of Imagery: WorldView3 satellite imagery:
  8-Bands multispectral imagery, georeferenced, orthorectified and pan-sharpened
- Off Nadir: < 12 degrees
- Max GSD: < 0.33m
- Sun Elevation: > 50.0 degrees
- Image Clouds: 0.00%
- Timeliness: within the last 12 months
- No snow or other whether phenomena
- Single stripped
- 16bit

Figure 2. Requirements on the test imagery.

A suitable image of a district of Sofia has been selected from the DigitalGlobe repository (Maxar, n.d.) and provided by DigitalGlobe distributer European Space Imaging (European Space Imaging, 2020) and their represener in Bulgaria – Vekom (Vekom Geo, 2019). It consisted of 8 bands with a spatial resolution of 1.2m and an additional panchromatic image with 0.3m spatial resolution. Importantly, the imagery was delivered georeferenced and ortho-rectified in order to represent more accurately the true location on the Earth's surface and match with official cadaster data.

In order to assess the segmentation performance of the models, the resulted inferences have to be evaluated against ground truth labels. Ideally, the ground truth annotation has to follow the same labelling convention as in the training dataset, whereby in a perfect scenario they are conducted by the same team. A manual annotation, however, is a time consuming and costly task. Therefore, official cadastral information on buildings footprints has been used instead. This approach is sufficient to assess the general transferability of the model.

The cadastral footprints were provided georeferenced in the same projection as the satellite image and time mismatch between both was reasonably close. The received cadastral footprints showed systematic, however, locally different misalignment to the satellite image as it is shown in Figure 3. Furthermore, footprints' shapes and sizes have not always matched too and occasionally there were missing footprints or such without correspondence in the satellite imagery.

Figure 3. Shift misalignment of cadastral footprint and image.

2.3 Data Pre-processing

The first step in the pre-processing was to pan-sharpen all bands to the same 0.3m resolution as per training data. To do this a free open-source tool *PanFusion* has been used (Vaiopoulos, 2021). The software provides easy to use implementation of a number of pan-sharpening algorithms. In this study several algorithms such as BDSD (band-dependent spatial detail), GSA (adaptive Gram-Schmidt) and HCS (hyperspherical color sharpening) have been applied and compared. The BDSD has proofed to be suitable for very high-resolution multispectral images in other studies (Garzelli, Nencini and Capobianco, 2007), and showed the best visual results. Next, the single stripe imagery has been tiled into images with pixel dimensions of 512 by 512 as expected by the model. Importantly, the geo-spatial information for each tile had to be retained, since it was used to tile the label data to the same spatial bounding boxes. Finally, the channels of the pan-sharpened tiles were separated, reordered, and selected to match the training input. The whole process is depicted in Figure 4.
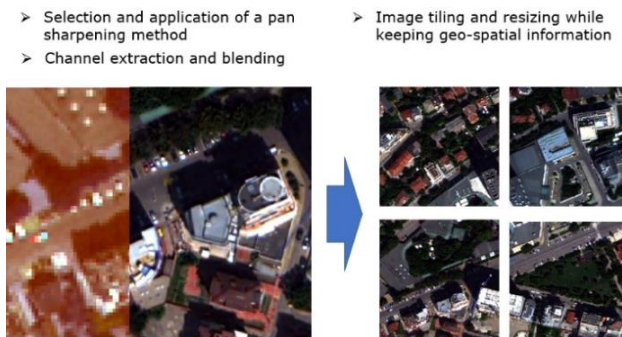
Figure 4. Image pre-processing.

Manual alignment of the cadastral footprints to better match with the buildings on the imagery was performed using QGIS (QGIS, 2021). Further label inaccuracies such as incorrect footprint size or shape have not been corrected since these mismatches with the actual image were not significant for an initial transferability assessment. Next, the cadastral information has to be tiled into parts corresponding to the satellite image tiles, whereby footprints spread across several tiles have been split. Finally, the polygons have been converted into binary masks as expected by the model. The whole process is depicted in Figure 5.
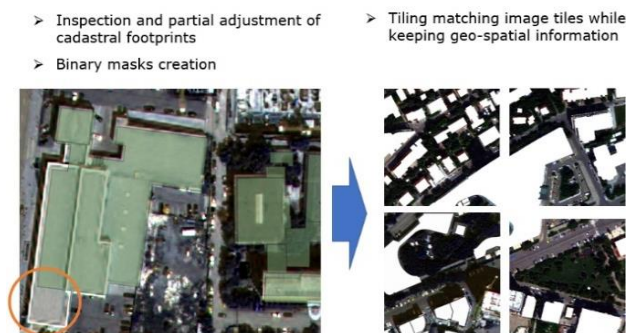

Figure 5. Ground Truth pre-processing.

2.4 Inference and Evaluation Methodology

The application of the model itself – running inferences – is done with few commands through the Solaris API once the input data is formatted in the right way. Notably, the last layer in the architecture of the provided model does not have an activation layer and returns raw values. Therefore, an additional *sigmoid* or *softmax* function should be applied in order to receive results for each pixel that could be interpreted as probabilities.

There are different metrics that could be considered for assessing the segmentation performance of a model. Simple *Mean Accuracy* gives a proportion of right versus false guesses for all pixels. This metric, however, is not suitable in cases where there is high class imbalance. In the current case there is strong predominance of the non-building class and so another metric has to be considered.

The SpaceNet 2 Challenge rates its competitors base on a F1 score inspired by ImageNet Large Scale Visual Recognition Challenge (Russakovsky, et al., 2015).

$$F1 = 2 \times \frac{Precision \times Recall}{Precision + Recall} \qquad (1)$$

Where:

$$Precision = \frac{True\ Positive}{True\ Positive + False\ Positicve} \qquad (2)$$

$$Recall = \frac{True\ Positive}{True\ Positive + False\ Negative} \qquad (3)$$

The score is a harmonic mean of Precision and Recall. A footprint prediction is considered a True Positive when there is an Intersection over Union (IoU) between a proposed polygon and the ground truth polygon larger than a threshold of 0.5, otherwise it is a False Positive. If there is no proposal for a certain ground truth, the footprint with IoU exceeding the threshold a False Negative is counted.

The SpaceNet 2 evaluation metric considers each building in a scene separately. However, this requires to have also a separate inference for each separate footprint. The model provided by Solaris fails at this separation and predicts one "bubble" for close buildings without further improvement. Furthermore, attached buildings very often cannot be recognized as different buildings. Therefore, their footprints polygons have to be merged into one if the SpaceNet 2 metric had to be applied.

Due to the above issues, this study evaluates the results slightly different. Here, pixel-level Precision, Recall, F1 and IoU for the building class have been compared. Consequently, each building is not assessed separately, but a result for a tile based on all buildings is calculated. Therefore, the results are not directly comparable with the ones reported on the SpaceNet challenges but are sufficient for a transferability assessment.

3. RESULTS

The model's transferability is assessed by analyzing its performance on a data from Paris and Sofia. It is important to note that the model has been trained on data from the former and has seen the test images in an initial training stage and therefore a higher performance in this case is expected. Table 1 shows the results averaged over the tiles (tiles with no buildings were filtered out).

Metric	Paris	Sofia
IOU	0.55	0.48
Precision	0.76	0.71
Recall	0.69	0.60
F1	0.69	0.63

Table 1. Comparison on performance on test data.

The average results are slightly higher on the training set. The analysis of error shows that this difference can be attributed to incorrect labels and rare object appearance to a large extend, as it is shown in Figure 6. Furthermore, rooftops with very bright near whit colors were not correctly detected.
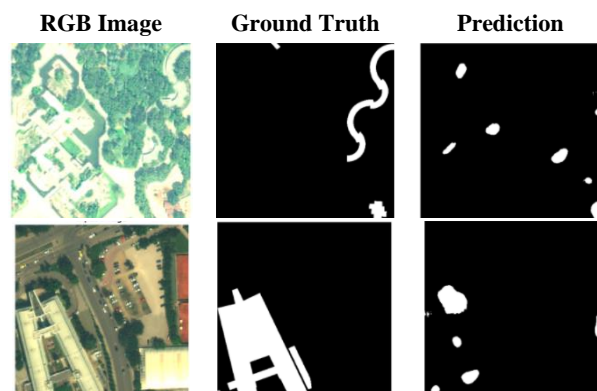

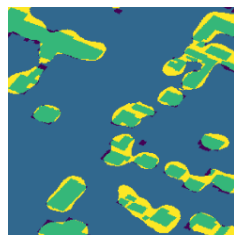**Figure 6.** Examples of tiles with low performance.

Single test images with no or small cadastral mismatch show comparable results with the training images as it is shown in Figure 7.

Paris results
Precision: 0.635, Recall: 0.931, F1: 0.755, IOU: 0.607
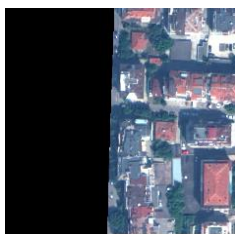

a) Image tile of Paris   b) Mask and prediction of Paris

Sofia results
Precision: 0.709, Recall: 0.93, F1: 0.805, IOU: 0.673

c) Image tile of Sofia   d) Mask and prediction of Sofia

Legend: ■ False Negative ■ False Positive ■ True Positive ■ True Negative

Figure 7. Comparison of performance on a single image.

The results show that the model's knowledge for certain geographical locations can be transferred to another one. They prove that additional training with data from the new location e.g., containing building shapes or colors more distinctive for the new location, would serve the model's performance.

4. CONCLUSION

This paper presents results from a study on transferability assessment of an open-source deep learning model for localization of buildings on satellite imagery. The model was pre-trained on open-source data from WorldView3 satellite and applied to images of another location made by the same device. The outcome of this study indicates that deep learning knowledge obtained from satellite images could be transferred to other cities and countries. Importantly, this suggests that accessible satellite images and open-sourced data in combination with deep learning methods can reduce the costs and effort for development and application of automatic analysis of large areas as long as the image quality and information is sufficient.

The model's performance could be improved, if more labelled data from the new location are included in the training stage along with the open-source data. Different strategies such as Deep Active Learning methods (Pengzhen Ren, 2020) and specific data augmentations or use other techniques or models could be applied to make the process more affordable. The application of such strategies is a subject of further work.

ACKNOWLEDGEMENTS

REFERENCES

CosmiQ Works, 2021. An open source machine learning pipeline for geospatial imagery. solaris.readthedocs.io (20 April 2021).

European Space Imaging, 2020. euspaceimaging.com (20 April 2021).

Li, Q., Shi, Y., Auer, S., Roschlaub, R., Möst, K., Schmitt, M., Zhu, X. X., 2020: Detection of Undocumented Buildings Using Convolutional Neural Network and Official Geodata. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences, 5*(2), 517-524.

Maggiori, E., Tarabalka, Y., Charpiat, G., Alliez, P., 2017: Can semantic labeling methods generalize to any city? the inria aerial image labeling benchmark. *IEEE Int. Geoscience and Remote Sensing Symposium (IGARSS)*, 3226-3229.

Maxar. (n.d.). *Digital Globe Repository*. discover.digitalglobe.com (20 April 2021).

QGIS, 2021 QGIS User Guide, docs.qgis.org/3.16/en/docs/user_manual (15 April 2021).

Ronneberger, O., Fischer, P., Brox, T., 2015: U-net: Convolutional net-works for biomedical image segmentation. *Int. Conf. on Medical Image Computing and Computer-Assisted Intervention*, 234-241, doi.org/10.1007/978-3-319-24574-4_28.

Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., Huang, Z, Karpathy, A., Khosla, A., Berstein, M., Berg, C. A., Fei-Fei, L., 2015: ImageNet Large Scale Visual Recognition Challenge. *Int. Journal Computer Vision,* 115, 211–252, doi.org/10.1007/s11263-015-0816-y.

SpaceNet, 2021. Accelerating Geospatial Machine Learning. spacenet.ai (10 April 2021)

SpaceNetChallenge, 2017. Marathon Match - Solution Description. github.com/SpaceNetChallenge/BuildingDetectors _Round2/tree/master/1-XD_XD (31 March 2021)

Vaiopoulos, D., 2021. PanFusion: Image Fusion – Pansharp App. www.pansharp.com/applications/panfusion (15 April 2021)

Vekom Geo, 2019. www.vekom.bg (10 April 2021).

Zhao, W., Ivanov I., Persello, C., 2020: Building outline delineation: From very high resolution remote sensing imagery to polygons with an improved end-to-end learning framework. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XLIII-B2-2020, 731-735. doi.org/10.5194/isprs-archives-XLIII-B2-2020-731-2020.

Zhao, K., Kamran, M., Sohn, G., 2020: Boundary Regularized Building Footprint Extraction From Satellite Images Using Deep Neural Network. *ArXiv, abs/2006.13176.*

Zhu, X. X., Tuia, D., Mou, L., Xia, G., Zhang, L., Xu, F., Fraundorfer, F., 2017: Deep Learning in Remote Sensing: A Comprehensive Review and List of Resources. *IEEE Geoscience and Remote Sensing Magazine*, 5, 8-36. doi.org/10.1109/MGRS.2017.2762307.