ZAZNAVANJE STAVB Z UPORABO NEVRONSKIH MREŽ, UČENIH S PRENOSOM ZNANJA

BUILDING DETECTION WITH CONVOLUTIONAL NETWORKS TRAINED WITH TRANSFER LEARNING

Simon Šanca, Krištof Oštir, Alen Mangafić

UDK: 681.511.4:004.8
Klasifikacija prispevka po COBISS.SI: 1.01
Prispelo: 3. 8. 2021
Sprejeto: 3. 11. 2021

DOI: 10.15292/geodetski-vestnik.2021.04.559-593
SCIENTIFIC ARTICLE
Received: 3. 8. 2021
Accepted: 3. 11. 2021

IZVLEČEK

Rezultati klasifikacije stavb na ortofotu se uporabljajo kot vir za vzdrževanje katastra stavb. V zadnjih letih se za klasifikacijo stavb v svetu vse bolj uveljavljajo metode globokega učenja z uporabo konvolucijskih nevronskih mrež. V raziskavi predstavimo primer samodejne klasifikacije stavb z uporabo lastnih podatkovnih zbirk, izdelanih iz barvnih bližnje infrardečih ortofotov (BIR-R-G) in barvnih ortofotov (R-G-B). Preizkusili smo detekcijo stavb z uporabo predučenih uteži podatkovnih zbirk Microsoft Common Objects in Context (MS COCO) in ImageNet. Za detekcijo stavb smo uporabili Mask Region Convolutional Neural Network (Mask R-CNN). Namen raziskave je preizkusiti uporabniško vrednost globokega učenja za detekcijo stavb z uporabo predučenih uteži na podatkih drugega barvnega prostora s ciljem izgradnje klasifikacijskega modela brez ponovnega učenja.

ABSTRACT

Building footprint detection based on orthophotos can be used to update the building cadastre. In recent years deep learning methods using convolutional neural networks have been increasingly used around the world. We present an example of automatic building classification using our datasets made of colour near-infrared orthophotos (NIR-R-G) and colour orthophotos (R-G-B). Building detection using pretrained weights from two large scale datasets Microsoft Common Objects in Context (MS COCO) and ImageNet was performed and tested. We applied the Mask Region Convolutional Neural Network (Mask R-CNN) to detect the building footprints. The purpose of our research is to identify the applicability of pre-trained neural networks on the data of another colour space to build a classification model without re-learning.

KLJUČNE BESEDE

globoko učenje, konvolucijske nevronske mreže, klasifikacija stavb, Mask R-CNN, detekcija objektov, segmentacija objektov, samodejna klasifikacija

KEY WORDS

deep learning, convolutional neural networks, classification of buildings, Mask R-CNN, object detection, object segmentation, automatic classification

1 INTRODUCTION

Image classification aims to recognise and name geographical objects and phenomena on the Earth's surface (Veljanovski et al., 2011). Classification extracts the essential classes (objects) of interest from an image, e.g., roads, forests, crops, water areas, ships, aircraft or buildings. Automatic recognition and classification of buildings from aerial or high-resolution satellite imagery is an important research topic in photogrammetry and remote sensing (Zhu et al., 2017). Rapid advances in computer vision and deep learning using deep convolutional neural networks, and free access to remote sensing data are enabling the development of new methods for automatic building classification. There are many online tasks and competitions that use deep learning methods to classify buildings using satellite or aerial imagery automatically. Examples of such competitions are the DeepGlobe Buildings Extraction Challenge[1], the SpaceNet Building Extraction Challenge[2], the crowdAI Mapping Challenge[3], etc.

Deep learning relies on feedforward, multi-layer neural networks. A specific example of artificial neural networks for image classification and segmentation are convolutional neural networks. A simplified example of a convolutional neural network contains three basic layers that can be repeated. These are (1) convolutional layers, (2) pooling layers and (3) fully connected layers (Goodfellow et al., 2016). A deep neural network consists of many such layers, which make it capable of recognising detailed patterns and shapes in an image. Each layer of a neural network has a distinctive task. The convolutional layer is a combination of multiple filters moving across the image, combining higher-level information into a two-dimensional activation matrix. The convolutional layers progressively reduce the spatial resolution of the activation matrices to reduce the computational complexity of the neural network while also increasing the field of view of each filter. The fully connected layer acts as a classifier that returns a classification vector based on the extracted feature set values, which is used to determine the membership of a particular class (Šanca, 2020).

One of the most successful convolutional neural networks for object detection is Mask R-CNN (He et al., 2017) (Mask Regional Convolutional Neural Network), which can achieve high accuracy in building classification (Šanca, 2020). Mask R-CNN was developed by Facebook AI Research (FAIR) in 2017. It is a deep convolutional neural network used for object detection, semantic segmentation and instance segmentation. Object detection with Mask R-CNN consists of two parts. The first part generates a suggestion of the area where the object should be located in the input image. The second part detects the object and assigns it a probability of belonging to a class, displays its position, and assigns an object mask. More details on Mask R-CNN can be found in the papers by He et al., (2017) and Abdulla (2017).

A detailed overview of the theory and application of deep learning in remote sensing is presented by Zhu et al. (2017). Most of the previous research based on the use of deep learning for automatic building detection uses high spatial resolution satellite imagery as the data source. In a study of automatic building detection, Shetty and Mohan (2018) used WorldView-2 satellite imagery. They used panchromatic imagery with a spatial resolution of 0.46 m to learn and test the Faster R-CNN neural network model. They found that using convolutional neural networks allows the detection of buildings of different shapes

[1] *http://deepglobe.org*
[2] *https://spacenetchallenge.github.io/*
[3] *https://www.crowdai.org/challenges/mapping-challenge*

with an accuracy of 88%, compared to the support vector machine (SVM) method. Using a large dataset of Google Earth imagery, Wen et al. (2019) researched the building detection capabilities of Faster R-CNNs. They compared the backbone architectures of VGG and ResNet101 when using Mask R-CNN and Faster R-CNN. In the proposed solution, Mask R-CNN was enhanced with a rotation matrix for building detection. The worst accuracy was achieved with Faster R-CNN-VGG (70%), the best with the proposed solution Mask R-CNN-VGG (91%). The other two solutions Faster R-CNN-ResNet101 and Mask R-CNN-ResNet101 achieve identical accuracies of 91%. Further, they also compared the results of semantic building segmentation with Mask R-CNN. The proposed method achieved an accuracy of 91%, and the conventional method Mask R-CNN-Resnet101 achieved an accuracy of 91%. Based on the study, they concluded that promising results can be achieved by using a large and complex building dataset and applying Mask R-CNN. Using the DeepGlobe dataset, Zhao et al. (2018) proposed an improved Mask R-CNN solution enhanced with a boundary recognition algorithm. Their proposed solution achieves more accurate results in building recognition and in detecting the footprints of the identified buildings than the baseline Mask R-CNN solution.

Compared to satellite imagery, aerial imagery has higher spatial resolution but fewer spectral bands. Using aerial imagery more complex objects can be identified with higher accuracy. Examples of building recognition datasets built from aerial imagery are the Massachusetts Buildings Dataset[4] (Mnih, 2013), Inria Aerial Image Labeling Dataset[5] (Maggiori et al., 2017), and AIRS Automatic Mapping of Buildings Dataset[6] (Chen et al., 2019). Research to date using aerial imagery and Mask R-CNN achieves high building detection accuracy. Ji et al. (2019) researched building changes using the Wuhan Building Change Detection Dataset. The proposed building detection solution consists of two convolutional neural networks; (1) a building recognition network based on Mask R-CNN and a Multi-Scale Fully Convolutional Network (MS-FCN), and (2) a building change detection network. The solution was tested in the Christchurch study area in New Zealand, comparing the 2011 and 2016 study areas. Both proposed solutions achieved high classification accuracies of over 89% for object classification and over 93% for pixel-based building classification. MS-FCN performed slightly better in building edge detection compared to Mask R-CNN. Building detection from aerial images using Mask R-CNN has been addressed in (Zhou et al., 2019), investigating the ability to identify buildings by varying the value of the anchor box parameter and the problem of segmenting the exact footprint of building edges. They compared two proposed solutions: (1) Mask R-CNN-S with a smaller anchor box and (2) Mask R-CNN-L with a larger anchor box. They found that the building detection results are better using the solution with a smaller anchor box, as it identifies smaller buildings and buildings with more detailed contents. They conclude with an important observation: Mask R-CNN is suitable for building detection, as classical convolutional neural networks do not preserve detailed spectral information when detecting objects, leading to poorer results. Another important finding is the importance of the anchor box parameter, which significantly impacts the quality of recognition of both small and larger, spectrally diverse objects. We present the results of the studies mentioned above in Table 1.

[4] *https://www.cs.toronto.edu/~vmnih/data/*
[5] *https://project.inria.fr/aerialimagelabeling/*
[6] *https://www.airs-dataset.com/*

Table 1: Comparison of Faster R-CNN, Mask R-CNN and MS-FCN

Neural network	Method used	Data	Spatial resolution [m]	Accuracy [%]	Study
Faster R-CNN	SVM	satellite, panchromatic	0.46	88	Shetty and Mohan, (2018)
Faster R-CNN-VGG	Rotation matrix for building recognition	aerial R-G-B	0.26	70	Wen et al. (2019)
Mask R-CNN-VGG	Rotation matrix for building recognition	aerial R-G-B	0.26	91	Wen et al. (2019)
Faster R-CNN-Resnet101	Rotation matrix for building recognition	aerial R-G-B	0.26	91	Wen et al. (2019)
Mask R-CNN-Resnet101	Rotation matrix for building recognition	aerial R-G-B	0.26	91	Wen et al. (2019)
Mask R-CNN	boundary regularisation algorithm	satellite R-G-B	0.5	88	Zhao et al. (2018)
Mask R-CNN	SVM	aerial R-G-B	0.3	90	Ji et al. (2019)
MS-FCN	SVM	aerial R-G-B	0.3	84	Ji et al. (2019)
Mask R-CNN-S	small anchor box	aerial R-G-B	0.5	85	Zhou et al. (2019)
Mask R-CNN-L	large anchor box	aerial R-G-B	0.5	81	Zhou et al. (2019)

The Geodetic Institute of Slovenia carries out the automatic classification of buildings annually to update the spatial databases of the Surveying and Mapping Authority of the Republic of Slovenia and perform spatial monitoring for the Ministry of Environment and Spatial Planning. Currently, the classification of buildings is carried out using machine learning methods, object-based classification using support vector machines and random forest, where the digital surface model is also a key piece of information. This paper aims to test a new method for building detection using deep learning, entirely independent of the use of a digital surface model, on two new building datasets.

Two building datasets were produced as part of the research, using DOF050 colour orthophotos (R-G-B) and DOF050IR colour infrared orthophotos (NIR-R-G) from 2019 with a spatial resolution of 0.5 *m*. Colour infrared orthophotos reveal a different perspective of the terrain, as objects with high reflectance in the infrared spectrum (e.g. healthy vegetation) are shown in red, while objects with high reflectance in the red spectrum are shown in green and objects with high reflectance in the blue-green spectrum are shown in blue (Oštir, 2006). The main advantage of using colour infrared orthophotos is that it is easier to distinguish buildings from vegetation based on the spectral signature alone.

We prepared the building datasets in the MS COCO format (Lin et al., 2014), which represents the training instances of buildings in JavaScript Object Notation (JSON). This way of annotation is fast and transparent, so we used it to create the building dataset. We trained eight different building classification models using the pre-trained weights of the MS COCO and ImageNet datasets (Deng et al., 2009). We performed transfer learning because Mask R-CNN relies on it to train and generalize models on new custom datasets faster. We validated the performance of the trained models on a selected test sample of buildings in Slovenia and evaluated each model by computing evaluation metrics. We were particularly interested in how well Mask R-CNN identifies buildings, how good the identified building masks are, and whether the proposed method presents a potential for further application.

2 STUDY AREA AND DATA DESCRIPTION

A key element of a successful deep learning building classification model is a well-designed dataset with many labelled training features. The study area for the construction of the two building datasets is shown in Figure 1 — the area of 1,387 km^2 contains 98,425 registered buildings as of 28. 3. 2020. The building typology is primarily rural except for Murska Sobota, which has an urban building typology. Roofs are of different shapes and colours, with red, brown and dark grey or black roofing predominating. Many of the roofs used in the training examples also contain solar panels. The area was chosen because of personal knowledge and because it includes a wide variety of roofs that contribute to the detail of the building dataset.

Figure 1: Map of the study area for the creation of the building dataset.

3 METHODS

In this chapter, we present the methodology for the creation of the building dataset. First, we define the building class, then create the training samples and their overview for the study area shown in Figure 1.

3.1 Definition of building class

The aim of the building cadastre is to register all buildings in the Republic of Slovenia. The concept of a building and part of a building is defined in ZEN (ZEN - Official Gazette of the Republic of Slovenia No. 47/06). The data on buildings and parts of buildings is described in the National Topographic Model (DTM) (GURS, 2020), which is defined by the law (Article 11 ZDGRS - Official Gazette of the Republic of Slovenia No. 25/14 and No. 61/17). A building is defined as an object permanently located in one place (Boguszewski et al., 2020). An example of a correct and considered building footprint is presented in Figure 2 on the left. Tall buildings pose a problem as the building footprint obtained from the building cadastre do not spatially coincide with the buildings. Such examples have been excluded from the dataset.

Figure 2: Examples of correct training samples (left) and incorrect training samples (right).

3.2 Methodology for the creation of the building dataset

The building dataset creation flowchart is presented on Figure 3. Each step is further explained below.

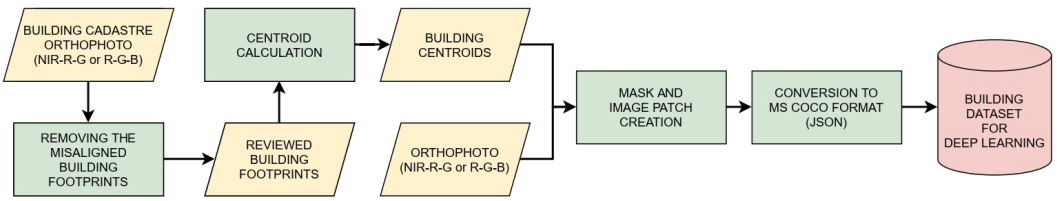


Figure 3: Deep learning dataset creation flowchart.

The building cadastre used to create the training samples was cleaned before the actual production of the dataset. We removed the building footprints which are, according to the generalised land use (MOP, 2020) located in: (1) manufacturing areas, (2) transport infrastructure areas and networks, (3) communication infrastructure areas and (4) energy infrastructure areas. Before the filtering, there were 98,425 footprints in the study area. First, 1,271 footprints were eliminated, which mainly included large industrial buildings. These were excluded because they are too large and cannot be adequately represented with 128 x 128 px patches. We also excluded buildings whose spectral properties are equivalent to transport infrastructure, mainly roads and transformers, as the outlines of these are too small for the algorithm to detect the transformer mask and compute the image coordinates of the created masks. Buildings located in areas of energy, communication and transport infrastructure are atypical. We wanted to create a dataset that is as balanced as possible and has as few outliers as possible. After the data cleaning process, we calculated the new centroids from the inspected and cleaned building footprints, which were then used to prepare the training patches.

3.3 Overview of sample images and building masks

By manually inspecting the training features, we improve the quality of the dataset, which leads to better building detection results. The most typical errors are building footprints that cover meadows, fields and shadows. This is due to the non-updated nature of the building cadastre. For the study area with 98,425 building footprints, we detected 8,226 cases of such errors, representing 8% of all footprints. We did not consider the erroneous building footprints in the production of the sample images and binary masks. After reviewing the produced sample images and binary masks, we obtained 58,000 training examples to create the building dataset. Some examples of the most typical errors are shown in Figure 4.

Figure 4: Examples of incorrect building footprints. The most typical errors are footprints that cover meadows, fields and shadows. Examples of partially overlapping footprints and building are un useful.

The dataset was split 80/20, with 80% of the training samples used for learning and 20% for validation. We created building masks and sample building patches of 128 x 128 px from the newly computed building centroids. Each building is represented in the dataset with a mask and a corresponding sample image. After that, we converted the dataset to MS COCO format, which, as mentioned, uses JSON annotation. In the first step, we labelled the locations of all the sample patches and the corresponding binary masks according to the building class. In the second step, we converted the sample and the building masks patches into the JSON annotation used by the MS COCO dataset. The detailed procedure is further described in Šanca (2020).

3.4 Transfer learning from pre-trained neural networks

Neural networks for specific computer vision tasks have already been built and trained on large datasets such as ImageNet (Deng et al., 2009) and MS COCO. The pre-trained weights of these datasets are accessible to the users, who can reuse these weights to better adapt the weights when training a neural network on their own dataset. An example of such learning is called transfer learning. Transfer learning works by initialising already trained model weights to adapt the weights better when training the network on another dataset. When extracting features from images, we use the neural network head and train only the new classifier to optimise the previously trained filter weights for new tasks on the second dataset; thus, the model training process requires much less time (Ramon et al., 2019). To train a Mask R-CNN neural network, we can use the pre-trained weights of two large-scale datasets, i.e., MS COCO and ImageNet.

The MS COCO dataset is one of the leading datasets for object detection and semantic segmentation. It contains annotations for various problems: (1) object detection, (2) keypoint detection, (3) stuff segmentation, (4) pan-optic segmentation, and (5) image captioning. It contains 91 image categories, of which 82 categories have more than 5,000 labelled image examples. The total number of labelled training examples of the MS COCO collection is 2.5 million on 328,000 images (Lin et al., 2014).

ImageNet is considered the largest dataset for state-of-the-art object recognition, containing more than 15 million manually annotated high-resolution images, organised according to the WordNet hierarchy into 22,000 classes. ImageNet supports: (1) object classification, (2) object detection and (3) single object localisation (Deng et al., 2009).

Mask R-CNN consists of two parts: (1) a convolutional backbone architecture used for feature extraction over an entire image, and (2) the network head for classification, bounding box recognition and mask prediction, that is applied separately for each Region of Interest (RoI) (Zhao, et al., 2018). The two backbone architectures of Mask R-CNN are ResNet101, which contains 101 convolutional layers for object detection, and a Feature Pyramid Network (FPN), which performs multiscale feature extraction from the input image (Šanca, 2020).

We used the same hyperparameters to train all the models, and we reduced the learning rate by a factor of 10 for the models whose weights were fine-tuned. The learning hyperparameters are given in Table 2, and we set them based on previous research studies. All hyperparameters listed in Table 2 are described in more detail in (Šanca, 2020).

Table 2: Hyperparameter values used in model learning

Name of the hyperparameter	Value
image shape	128
learning rate	0.001 and 0.0001
batch size	1,000
number of repetitions per epoch (steps per epoch)	1,000
number of validation steps	50
backbone architecture	ResNet-101
anchor box size	(8, 16, 32, 64, 128)
number of proposed regions per image	32
total train time	328 hours

The models were trained on a computer with the following specifications:

- CPU: Intel(R) Core (TM) i9-9900X CPU @ 3.50GHz,
- Memory: 64 GB RAM, DDR4 (4 x 16 GB DIMM DDR4 Synchronous 2400 MHz),
- Graphics card: NVidia GeForce RTX 2080 SUPER.

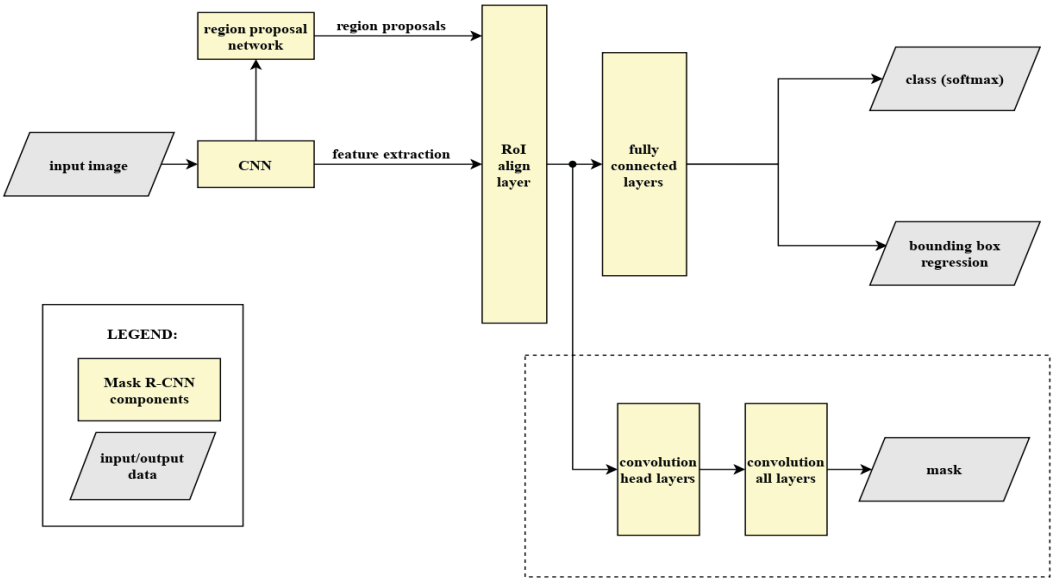


Figure 5: Mask R-CNN architecture for the building detection case. The input image passes through a convolutional neural network (CNN) that extracts features in the first step. In the second step, the region proposal network randomly suggests regions in the input image based on predictions generated according to a defined class in the dataset. The higher-level information is merged into lower-level information using the RoiAlign layer, which acts as a merging layer and is split into two parts. In the first part, the RoiAlign results are moved over the fully connected layers to predict the object class and compute the image field regression. In the second part, the RoiAlign results are moved across the convolutional layers of the neural network head and then across all the layers of the neural network, creating a pixel mask for each region of interest (ROI) and segmenting the image pixel by pixel according to the identified class.

In this study, we compare the performance of eight classification models; the results are presented in Table 3.

Table 3: Trained models on the NIR-R-G and R-G-B datasets.

Model name	Dataset	Layers used for learning	Number of epochs	Used weights	Learning rate	Learning time [h]
M1	NIR-R-G	heads	100	MS COCO	0.001	39
M2	NIR-R-G	heads	100	ImageNet	0.001	40
M3	NIR-R-G	all layers	200	MS COCO	0.0001	42
M4	NIR-R-G	all layers	200	ImageNet	0.0001	44
M5	R-G-B	heads	100	MS COCO	0.001	37
M6	R-G-B	heads	100	ImageNet	0.001	39
M7	R-G-B	all layers	200	MS COCO	0.0001	43
M8	R-G-B	all layers	200	ImageNet	0.0001	44

The number of epochs indicates the number of passes of the entire training dataset through Mask R-CNN. For models M1, M2, M5, M6 we trained only the head layer of the neural network, for models: M3, M4, M7, M8 we trained all the layers of the neural network. The learning rate represents the step size in the computation of the gradient of the loss function, which is iteratively minimised during neural network training. We choose a suitable learning rate to avoid underfitting/overfitting (Šanca, 2020). The model training process is illustrated in Figure 6.

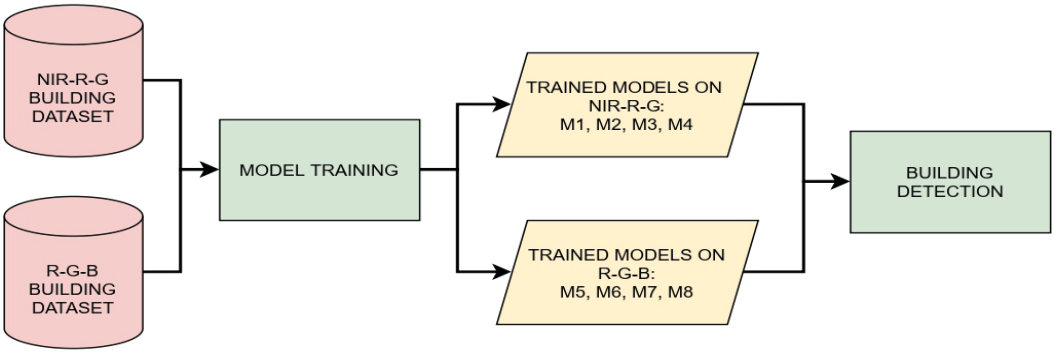


Figure 6: Flowchart of models trained for building detection. Models: M3, M4 and M7, M8 are fine-tuned models up to 100 epochs after training the neural network head. Building detection was performed separately with each model.

4 RESULTS AND EVALUATION

We evaluated the performance of transfer learning based on the loss function of the training process. We did not consider the loss function of the validation process to assess the stability of the model because on a large proportion of the training pairs, the building footprints from the building cadastre are inconsistent. The trained model successfully predicts a building where there is no footprint, which is detected by the validation process as a misclassified object. We decided to validate the performance of the models by calculating evaluation metrics (section 4.1). We validated the performance of the building classification models on a selected test area outside the training dataset area. We chose 300 examples to test the performance of the trained models, considering the criterion of roof diversity, in particular roof colour and shape, presence of solar panels on the roof, shadows, etc. We first compare the trained models with each other and then compare the obtained M3 and M4 results with the building cadastre. We present the building detection results in figures, where we compare the predictions of all the trained models on the NIR-R-G. We compare the resulting building footprints with the building cadastre.

4.1 Loss functions after transfer learning and comparison of models

We present an example of building detection with the MS COCO (R-G-B) pre-trained model, which we used to initialise our weights when training on our own dataset, in Figure 7. As expected, building detection without transfer learning on the constructed building dataset is incorrect. The MS COCO or ImageNet base model only randomly suggests spatial fields.

Mask R-CNNs multi-task loss function is the sum of the classification loss, bounding box loss and the mask loss. We are most interested in the mask loss for the building detection case, as it represents a measure of the accuracy of the building mask classification. In Figure 8, we see that the overall loss function after transfer learning of the MS COCO model is minimally better than the loss function of the ImageNet model. This is also true across the individual loss functions for classification, regression and masking. The loss function minimises during training and stabilises towards the end of it (after about 65 training epochs), which means there is no need for further training. The value of the loss function for the Mask R-CNN (Mask loss) class does not change much during the training process. This is because we only have a single class in the building dataset that is predicted at detection. Only the loss functions during the training phase are shown.

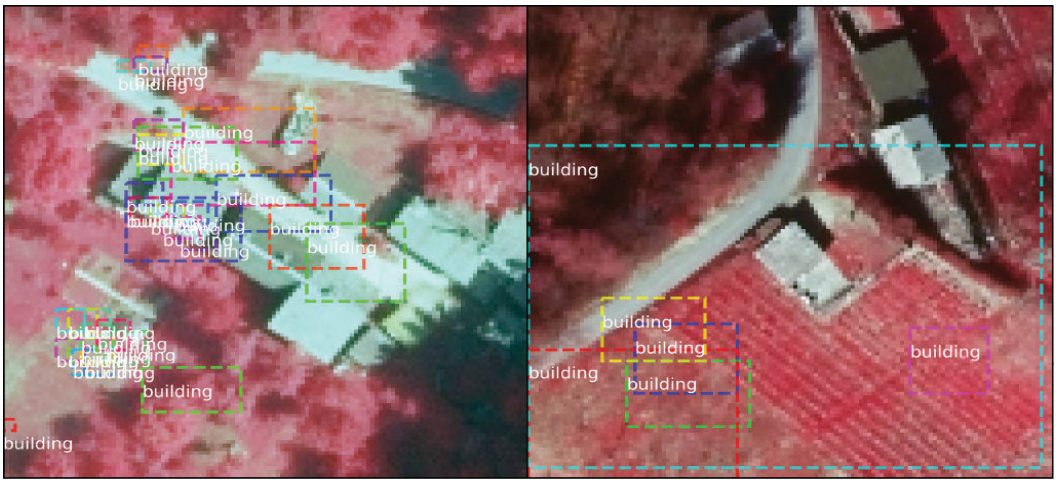


Figure 7: Example of building detection with a basic, not yet re-trained MS COCO model.

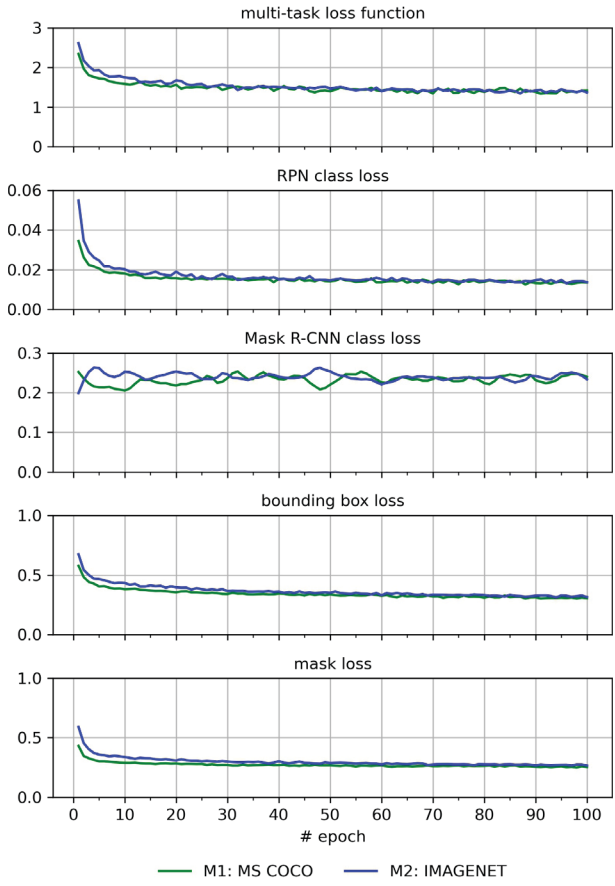


Figure 8: Training loss functions for M1 and M2 on the NIR-R-G dataset. For both models, we trained the heads of the neural network for 100 epochs.

After the fine-tuning, the values of the loss functions are further reduced, except for the loss function for the mask, which always stabilises at the start and does not improve, which means that the footprints of the identified buildings do not change noticeably in detection.

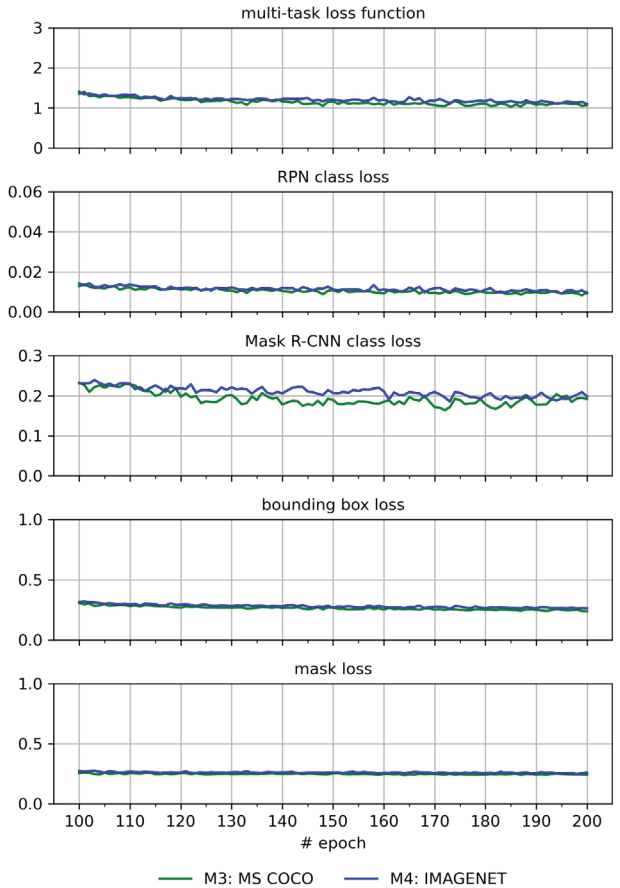


Figure 9: Training loss functions for M3 and M4 on the NIR-R-G dataset. For both models, we trained all the layers of the neural network for 100 epochs.

The results of the building detection with the trained models on BIR-G-R are shown in Figures 10, 11 and 12. The M3 and M4 models recognise smaller buildings, but the footprints of the recognised buildings remain very similar compared to M1 and M2. Fine-tuning with training all the layers of the Mask R-CNN is important in improving the prediction accuracy. Still, it is not crucial for improving the footprints of the recognised buildings compared to the footprints obtained by training the neural network head alone.

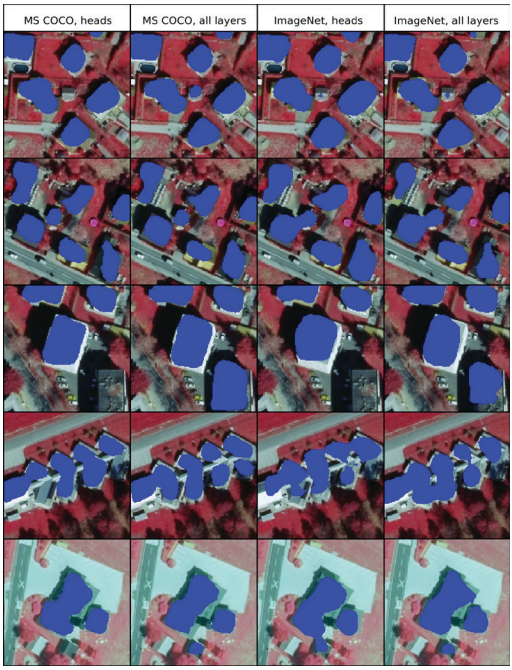


Figure 10: Comparison of the performance of the models M1, M3, M2, M4.

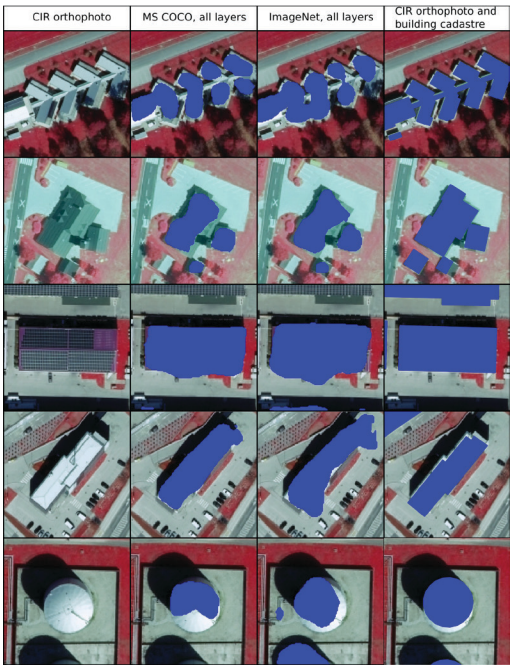


Figure 11: Detection results of buildings with M3 and M4 and comparison with the building cadastre.

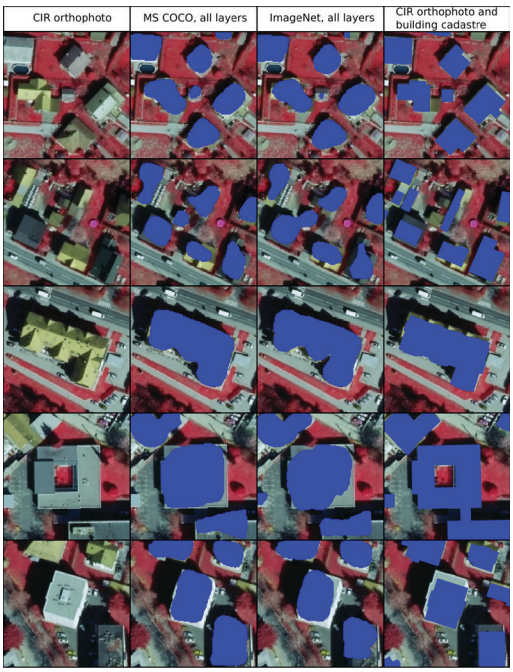


Figure 12: Detection results of buildings with M3 and M4 and comparison with the building cadastre.

Figure 13 shows the training loss functions of M5 and M6 trained on R-G-B, and Figure 14 shows the loss functions of M7, M8 trained on R-G-B. It can be seen that the loss functions of the R-G-B models are approximately the same as the loss functions of the NIR-R-G models.

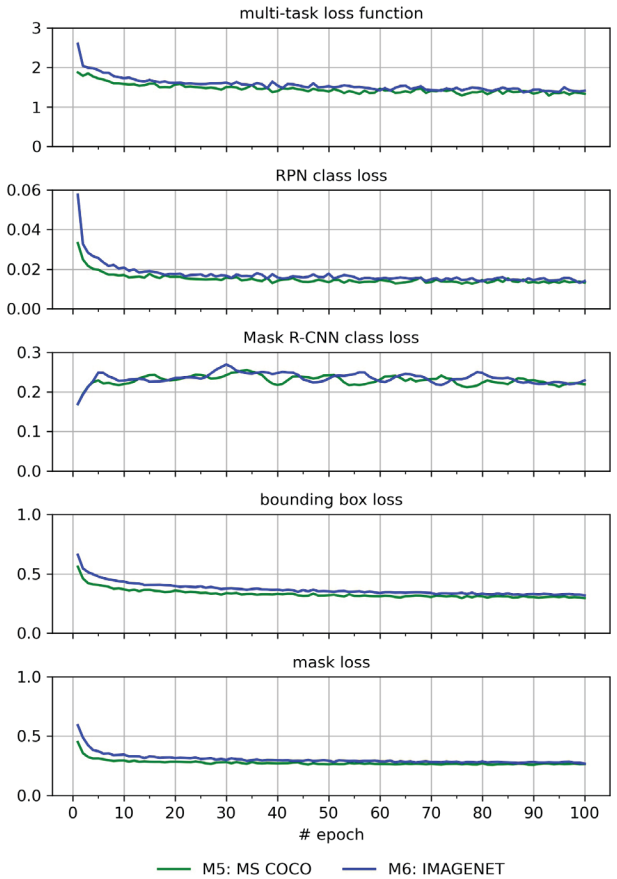


Figure 13: Training loss functions M1 (green) and M2 (blue) on the R-G-B dataset. For both models, we trained the heads of the neural network for 100 epochs.

We were also interested in the difference between the performance of the models trained on the NIR-R-G and R-G-B datasets. We show the results of building detection with the R-G-B models in Figure 15.

The differences between M1, M2 NIR-R-G and M5, M6 R-G-B are also minimal when recognising building footprints. This is also true for the fine-tuned models M3, M4 and M7, M8. Compared to the NIR-R-G orthophotos, the building detection results on R-G-B are slightly better because the base model is trained on MS COCO dataset R-G-B images of everyday life. In both cases, transfer learning from weights pre-trained on MS COCO or ImageNet datasets turns out to be an efficient solution compared to training models from scratch.

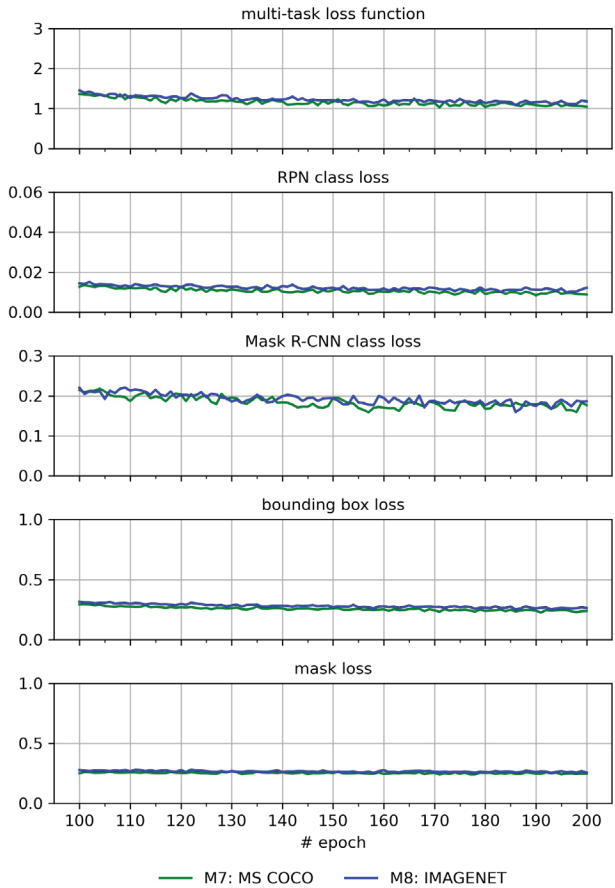


Figure 14: Training loss functions M3 (green) and M4 (blue) on the R-G-B dataset. For both models, we trained all the layers of the neural network for 100 epochs.

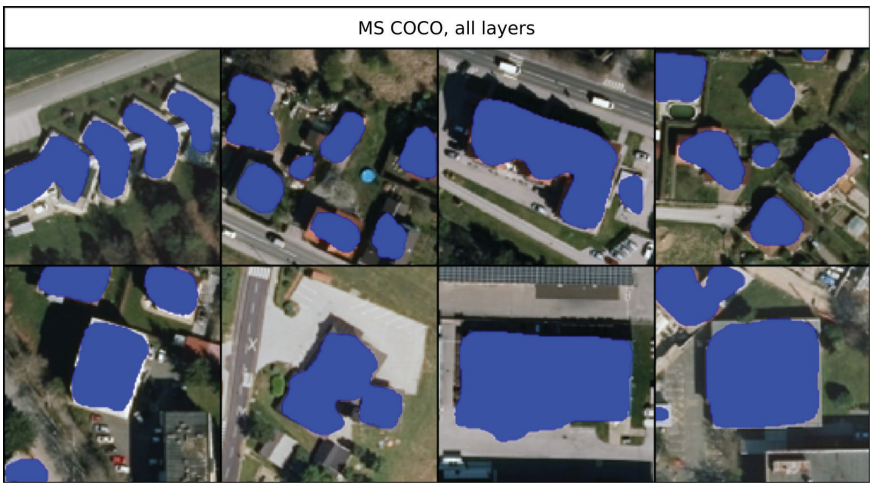


Figure 15: Example of M7 building detection on an R-G-B orthophoto.

4.2 Performance of learned models

We evaluate the performance of the trained classification models in machine learning based on testing patches. Building classification is an example of binary classification, where a confusion matrix represents the prediction performance.

		Predicted
Actual	TP – True Positive	FN – False Negative
	FP – False Positive	TN – True Negative

Correctly classified cases fall into True Positive (TP) and True Negative (TN). Misclassified cases belong to False Negative (FN) and False Positive (FP). The prediction results are used to calculate evaluation metrics to assess the performance of the building detection models. The equations used to calculate the evaluation metrics are taken from (Fetai, et al., 2021). Accuracy represents the proportion of correct predictions over all model predictions for binary classification. The equation is simplified as:

$$accuracy = \frac{TP + TN}{TP + FP + TN + FN} := \frac{TP + TN}{All\ cases} \tag{1}$$

Recall measures the proportion of correctly predicted cases out of all positive cases:

$$recall = \frac{TP}{TP + FN} \tag{2}$$

The combined precision and recall metric is defined by the F1 score and is used when you only want to show one metric for the performance of a model:

$$F1 = \frac{2 \cdot accuracy \cdot recall}{accuracy + recall} \tag{3}$$

We counted the proportion of correctly and incorrectly identified buildings for the trained models and calculated evaluation metrics. The results are shown in Table 2.

Table 4: Evaluation of the performance of the building classification models.

	Dataset	Weights	TP	TN	FP	FN	Accuracy	Recall	F1 score
M1	NIR-R-G	MS COCO	246	0	38	16	0.8200	0.9389	0.8754
M2	NIR-R-G	ImageNet	243	0	43	14	0.8100	0.9455	0.8725
M3	NIR-R-G	MS COCO	296	0	3	1	0.9867	0.9966	0.9916
M4	NIR-R-G	ImageNet	294	0	4	2	0.9800	09932	0.9866
M5	R-G-B	MS COCO	249	0	36	15	0.8300	0.9432	0.8830
M6	R-G-B	ImageNet	244	0	41	15	0.8133	0.9421	0.8730
M7	R-G-B	MS COCO	298	0	1	1	0.9933	0.9967	0.9950
M8	R-G-B	ImageNet	296	0	1	2	0.9867	0.9933	0.9900

The trained models are very similar in terms of building detection performance. The 200 epoch models (M3, M4 and M7, M8) are more successful in identifying the footprints of buildings and identifying buildings even at the edge of the sample patches. M1, M2 and M5, M6 fail to recognise mainly small buildings or buildings partially covered by vegetation.

5 CONCLUSION AND DISCUSSION

This paper presents a building detection process with Mask R-CNN, from dataset creation to building detection for eight models. We trained the models by transferring knowledge from the weights pre-trained on MS COCO and ImageNet datasets. We compared the performance of building classification using R-G-B orthophotos and NIR-R-G colour infrared orthophotos. The dataset produced for deep learning building detection has a potential for further use. It can be further extended to the whole country and possibly used for future building data retrieval and building database maintenance. The obtained results confirm that Mask R-CNN is useful and suitable for building classification as also claimed by related research (Ji et al., 2019) and (Zhou et al., 2019).

In Slovenia, this is, to our knowledge, the first example of deep learning-based building detection. It is also the first case of applying transfer learning from models trained on MS COCO and ImageNet datasets, containing only R-G-B images of everyday objects, to a building dataset made from R-G-B and NIR-R-G orthophotos. The success of transfer learning from one domain to another domain demonstrates the flexibility of this type of networks. We claim that transfer learning is an effective method for updating models that have been trained on different data. Training models from scratch is a time-consuming process that allows building better models. Applying transfer learning to such models using new data of the same colour space would allow faster and more efficient model updating and, above all, faster training.

With the experience gained, we have some suggestions for improving further research. The first example of improvement is producing building masks of more regular shapes with a footprint detection algorithm, as Zhao et al., (2018) stated. The identified building masks can be vectorised in the next step and integrated into a GIS. The dataset can be enhanced by combining a colour infrared orthophoto and a normalised digital surface model (nDSM), which would separate the roofs from the terrain in the data creation process. The dataset could be extended to include buildings across Slovenia, allowing the model to be trained over the whole country.

ACKNOWLEDGEMENTS

The research was partly carried out within the applied research project L2-1826, co-funded by the Slovenian Research Agency, the Geodetic Administration of the Republic of Slovenia and the Ministry of Defence, and the research programme P2-0406 and project J2-9251, funded by the Slovenian Research Agency. Many thanks also to the Department of Civil Engineering at Western Norway University of Applied Sciences for partially funding the research.

Literature and references:

Abdulla, W. (2017). Mask R-CNN for object detection and instance segmentation on Keras and TensorFlow.

https://github.com/matterport/Mask_RCNN, accessed 17. 3. 2020.

Boguszewski, A., Batorski, D., Ziemba-Jankowska, N., Zambrzycka, A., Dziedzic, T. (2020). LandCover.ai: Dataset for Automatic Mapping of Buildings, Woodlands and Water from Aerial Imagery. ArXiv.

Fetai, B., Račič, M., Lisec, A. (2021). Deep Learning for Detection of Visible Land Boundaries from UAV Imagery. Remote Sensing. 13 (11): 2077. DOI: https://doi.org/10.3390/rs13112077.

Chen, Q., Wang, L., Yifan, W., Guangming, G., Zhiling, W. S. (2019). Aerial Imagery for Roof Segmentation: A Large-Scale Dataset towards Automatic Mapping of Buildings. 147 (07), 42–55.

Crésson, R. (2018). Orfeo Toolbox meets TensorFlow. https://github.com/remicres/otbtf, accessed 13. 4. 2020.

Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K., Fei-Fei, L. (2009). ImageNet: a Large-Scale Hierarchical Image Database. IEEE Conference on Computer Vision and Pattern

Recognition, 248–255. DOI: https://doi.org/10.1109/CVPR.2009.5206848

Goodfellow, I., Bengio, Y., Courville, A. (2016). Deep Learning. Boston, MIT Press.

Geodetska Uprava RS (2020). Zbirka topografskih podatkov (DTM). https://www.e-prostor.gov.si/fileadmin/struktura/DTM_objektni_katalog.pdf, accsessed 13. 4. 2020.

Grigillo, D. (2010). Samodejno odkrivanje stavb na visokoločljivih slikovnih virih za potrebe vzdrževanje topografskih podatkov [Automatic building detection from high resolution imagery for maintenance of topographic data]. Doctoral dissertation. Ljubljana: University of Ljubljana. http://drugg.fgg.uni-lj.si/781/, accessed 25. 4. 2020.

Grizonnet, M., Michel, J., Poughon, V., Inglada, J., Savinaud, M., Cresson, R. (2018). Orfeo ToolBox: Open source processing of remote sensing images. Open Geospatial Data, Software and Standards. 2. DOI: https://doi.org/10.1186/s40965-017-0031-6

He, K., Gkioxari, G., Dollár, P., Girshick, R. (2017). Mask R-CNN. 2017 IEEE International Conference on Computer Vision (ICCV), 2980–2988. DOI: https://doi.org/10.1109/iccv.2017.322

Ji, S., Yanyun, S., Lu, M., Zhan, Y. (2019). Building Instance Change Detection from Large-Scale Aerial Images using Convolutional Neural Networks and Simulated Shapes. Remote Sensing, 11 (11), 1343–1363. DOI: https://doi.org/10.3390/rs11111343

Lin, T. Y., Maire, M., Belongie, S., Bourdev, L., Girshik, R., Hays, J., Perona, P., Ramanan, D., Zitnick, C. L., Dollár, P. (2014). Microsoft COCO: Common Objects in Context. Computer Vision and Pattern Recognition, pp. 740–755.

Maggiori, E., Tarabalka, Y., Charpiat, G., Alliez, P. (2017). Can semantic labeling methods generalize to any city? The Inria aerial image labeling benchmark. 2017 IEEE International Geoscience and Remote Sensing Symposium (IGARSS), 2017, 3226–3229. DOI: https://doi.org/10.1109/IGARSS.2017.8127684

Mnih, V. (2013). Machine Learning for Aerial Image Labeling. Doktorska disertacija. Toronto: University of Toronto. https://www.cs.toronto.edu/~vmnih/docs/Mnih_Volodymyr_PhD_Thesis.pdf, accessed 8. 10. 2021.

MOP (2020). Dostop do podatkov o prostorskih aktih.

https://dokumentipis.mop.gov.si/javno/veljavni/tematski_zbirni_sloji/gnrp_opis.pdf, accessed 14. 4. 2020.

MS COCO (2014). COCO Data Format. http://cocodataset.org/#format-data, accessed 1. 4. 2020.

Oštir, K. (2006). Daljinsko zaznavanje. Ljubljana, Založba ZRC, ZRC SAZU.

Račič, M. (2019). Kategorizacija uporabe zemeljske površine na podlagi multispektralnih slik [Categorisation of land use based on multispectral imagery]. Master thesis. Ljubljana: University of Ljubljana. https://repozitorij.uni-lj.si/IzpisGradiva.php?lang=slv&id=110064, accessed 17. 3. 2020.

Šanca, S. (2020). Samodejna klasifikacija stavb z globim učenjem [Automatic classification of buildings with deep learning]. Master thesis. Ljubljana: University of Ljubljana. https://repozitorij.uni-lj.si/Dokument.php?id=135433&lang=slv, accessed 20. 11. 2020.

Shetty, A. R., Mohan, B. (2018). Building Extraction in High Spatial Resolution Images Using Deep Learning Techniques. Computational Science and Its Applications – ICCSA 2018. Melbourne, Springer.

Veljanovski, T., Kanjir, U., Oštir, K. (2011). Objektno usmerjena analiza podatkov daljinskega zaznavanja. Geodetski vestnik, 55 (4), 665–668. DOI: https://doi.org/10.15292/geodetski-vestnik.2011.04.665-688

Wen, Q., Jiang, K., Wang, W., Liu, Q., Guo, Q., Li, L., Wang, P. (2019). Automatic Building Extraction from Google Earth Images under Complex Backgrounds Based on Deep Instance Segmentation Network. Sensors, 19 (2), 333–349. DOI: https://doi.org/10.3390/s19020333

Zakon o državnem geodetskem referenčnem sistemu (ZDGRS) [National Land Survey Reference System Act]. Official Gazette of the Republic of Slovenia No 25/2014 and 61/2017.

Zakon o evidentiranju nepremičnin (ZEN) [Real Estate Records Act]. Official Gazette of the Republic of Slovenia No 47/2006 and amendments.

Zhao, K., Kang, J., Jung, J., Sohn, G. (2018). Building Extraction from Satellite Images Using Mask R-CNN with Building Boundary Regularization. 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), Salt Lake City, 2018, pp. 2242–2424. DOI: https://doi.org/10.1109/CVPRW.2018.00045

Zhou, K., Chen, Y., Smal, I., Lindenbergh, R. (2019). Building Segmentation from Airborne VHR Images using Mask R-CNN. ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, 42 (2), 151–161. DOI: https://doi.org/10.5194/isprs-archives-XLII-2-W13-155-2019

Zhu, X., T., Devis, M., Lichao, X., Gui-Song, Z., Liangpei, X., Feng, 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: https://doi.org/10.1109/MGRS.2017.2762307

Šanca S., Oštir K., Mangafić A. (2021). Building detection with convolutional networks trained with transfer learning.
Geodetski vestnik, 65 (4), 559-593.
DOI: https://doi.org/10.15292/geodetski-vestnik.2021.04.559-593

ZAZNAVANJE STAVB Z UPORABO NEVRONSKIH MREŽ, UČENIH S PRENOSOM ZNANJA

OSNOVNE INFORMACIJE O ČLANKU:

GLEJ STRAN 559

1 UVOD

Namen klasifikacije podob je razpoznavanje in poimenovanje geografskih objektov in pojavov na zemeljskem površju (Veljanovski et al., 2011). S klasifikacijo iz podobe izluščimo bistvene razrede (objekte), ki nas zanimajo, recimo ceste, gozdove, poljščine, vodna območja, ladje, letala ali stavbe. Samodejna prepoznava in klasifikacija stavb iz letalskih posnetkov ali visokoločljivih satelitskih posnetkov je pomembna tema raziskav na področju fotogrametrije in daljinskega zaznavanja (Zhu et al., 2017). Hiter napredek na področju računalniškega vida in globokega učenja z uporabo globokih konvolucijskih nevronskih mrež ter prosti dostop do podatkov daljinskega zaznavanja omogoča razvoj novih metod za samodejno klasifikacijo stavb. Na spletu najdemo veliko nalog in tekmovanj, ki uporabljajo metode globokega učenja za samodejno klasifikacijo stavb z uporabo satelitskih ali letalskih posnetkov. Primeri takih tekmovanj so DeepGlobe Buildings Extraction Challange[7], SpaceNet Building Extraction Challenge[8], crowdAI Mapping Challenge[9] idr.

Globoko učenje za modeliranje podatkov uporablja usmerjene, večslojne nevronske mreže. Specifičen primer umetnih nevronskih mrež za klasifikacijo in segmentacijo slik so konvolucijske nevronske mreže. Poenostavljen primer konvolucijske nevronske mreže vsebuje tri osnovne sloje, ki se lahko ponavljajo. To so: (1) konvolucijski sloji (angl. *convolutional layers*), (2) združevalni sloji (angl. *pooling layers*) in (3) polno povezani sloji (angl. *fully connected layers*) (Goodfellow et al., 2016). Globoko nevronsko mrežo sestavlja veliko takih slojev, zaradi česar je sposobna prepoznati podrobne vzorce in oblike na sliki. Vsak sloj nevronske mreže ima značilno nalogo. Konvolucijski sloj je kombinacija večkratnih filtrov, ki se premikajo čez sliko, in združuje višje nivojske informacije v dvodimenzionalno aktivacijsko matriko. Prostorsko ločljivost aktivacijskih matrik postopoma zmanjšujejo združevalni sloji, da se zmanjša računska kompleksnost nevronske mreže in hkrati povečuje vidno polje posameznega filtra. Polno povezani sloj deluje kot klasifikator, ki na podlagi pridobljenih vrednosti nabora značilk vrne klasifikacijski vektor, na podlagi katerega se določi pripadnost nekemu razredu (Šanca, 2020).

Ena izmed uspešnejših konvolucijskih nevronskih mrež za detekcijo objektov je Mask R-CNN (He et al., 2017) (angl. *Mask Regional Convolutional Neural Network*), s katero lahko dosežemo visoko točnost klasifikacije stavb (Šanca, 2020). Mask R-CNN so razvili pri Facebook AI Research (FAIR) leta 2017. Gre za globoko konvolucijsko nevronsko mrežo, ki se uporablja za detekcijo objektov (angl. *object detection*), semantično segmentacijo (angl. *semantic segmentation*) ter segmentacijo primerov (angl. *instance segmentation*). Detekcijo objektov z Mask R-CNN sestavljata dva dela, v prvem se generira predlog območja, kjer

[7] *http://deepglobe.org*
[8] *https://spacenetchallenge.github.io/*
[9] *https://www.crowdai.org/challenges/mapping-challenge*

naj bi se objekt nahajal na vhodni sliki, v drugem se objekt zazna in se zanj določi verjetnost pripadnosti razredu, prikaže se položaj objekta in določi njegova maska. Več podrobnosti o Mask R-CNN najdemo v prispevkih He et al. (2017) in Abdulla (2017).

Podroben pregled teorije in uporabe globokega učenja v daljinskem zaznavanju predstavijo Zhu et al. (2017). Večina dosedanjih raziskav, ki temeljijo na uporabi globokega učenja za samodejno detekcijo stavb, kot vir podatkov uporablja satelitske posnetke visoke prostorske ločljivosti. Pri študiji samodejne detekcije stavb sta Shetty in Mohan (2018) uporabila podatkovno množico satelitskega sistema WorldView-2. Za učenje in testiranje modela z nevronsko mrežo Faster R-CNN sta uporabila pankromatske posnetke prostorske ločljivosti 0,46 metra. Ugotovila sta, da uporaba konvolucijskih nevronskih mrež omogoča detekcijo stavb različnih oblik s točnostjo 88 %, v primerjavi z metodo podpornih vektorjev (angl. *support vector machines*), kjer je bila dosegljiva točnost detekcije 88,3 %. Z uporabo velike podatkovne zbirke posnetkov Google Earth so Wen et al. (2019) raziskovali zmožnosti detekcije stavb s Faster R-CNN. Primerjali so hrbtni arhitekturi VGG in ResNet101 pri uporabi Mask R-CNN in Faster R-CNN. Pri predlagani rešitvi so Mask R-CNN nadgradili z rotacijsko matriko za detekcijo stavb. Najslabšo točnost so dosegli s Faster R-CNN-VGG (70 %), najboljšo s predlagano rešitvijo Mask R-CNN-VGG (91 %). Ostali dve rešitvi Faster R-CNN-ResNet101 in Mask R-CNN-ResNet101 dosegata identični točnosti 91 %. V nadaljevanju so primerjali tudi rezultate semantične segmentacije stavb z Mask R-CNN, kjer predlagana metoda doseže točnost 91 %, navadna metoda Mask R-CNN-Resnet101 pa točnost 91 %. Na podlagi raziskave ugotovijo, da se z uporabo velike in kompleksne podatkovne zbirke stavb in uporabo Mask R-CNN lahko dosežejo obetavni rezultati. Z uporabo podatkovne zbirke DeepGlobe so Zhao et al. (2018) predlagali izboljšano rešitev Mask R-CNN, ki so jo nadgradili z algoritmom prepoznave mej. Njihova predlagana rešitev doseže natančnejše rezultate pri prepoznavi stavb in zaznavanju obrisov prepoznanih stavb kot osnovna rešitev Mask R-CNN.

Letalski posnetki imajo v primerjavi s satelitskimi posnetki visoko prostorsko ločljivost, a manj spektralnih kanalov. Podatkovne zbirke na podlagi letalskih posnetkov omogočajo prepoznavo kompleksnejših objektov z višjo točnostjo. Primeri podatkovnih zbirk za prepoznavo stavb, izdelanih iz letalskih posnetkov, so: Massachusetts Buildings Dataset[10] (Mnih, 2013), Inria Aerial Image Labeling Dataset[11] (Maggiori et al., 2017), AIRS Automatic Mapping of Buildings Dataset[12] (Chen et al., 2019). Dosedanje raziskave z uporabo letalskih posnetkov in Mask R-CNN dosegajo visoko točnost detekcije stavb. V prispevku Ji et al. (2019) so raziskovali spremembe stavb s podatkovno zbirko Wuhan Building Change Detection. Predlagano rešitev za detekcijo stavb sestavljata dve konvolucijski nevronski mreži: (1) mreža za prepoznavo stavb, ki deluje na osnovi Mask R-CNN in MS-FCN (angl. *Multi-Scale Fully Convolutional Network*), ter (2) mreža za odkrivanje sprememb stavb. Rešitev so preizkusili na študijskem območju Christchurcha v Novi Zelandiji, kjer so primerjali območji iz let 2011 in 2016. Obe predlagani rešitvi sta dosegli visoko točnost klasifikacije, in sicer nad 89 % pri objektni klasifikaciji in nad 93 % pri pikselski klasifikaciji stavb. MS-FCN je bil v primerjavi z Mask R-CNN nekoliko boljši pri detekciji robov stavb. Z detekcijo stavb iz letalskih posnetkov z uporabo Mask R-CNN so se ukvarjali (Zhou et al., 2019). Raziskovali so zmožnost prepoznave stavb s spreminjanjem vrednosti parametra sidrnega

[10] *https://www.cs.toronto.edu/~vmnih/data/*
[11] *https://project.inria.fr/aerialimagelabeling/*
[12] *https://www.airs-dataset.com/*

polja (angl. *anchor box*) in problem segmentacije natančnih obrisov robov stavb. Primerjali so dve predlagani rešitvi: (1) Mask R-CNN-S z manjšim sidrnim poljem in (2) Mask R-CNN-L z večjim sidrnim poljem. Ugotovili so, da so rezultati detekcije stavb boljši z uporabo rešitve z manjšim sidrnim poljem, saj ta prepozna manjše stavbe in stavbe z bolj podrobno vsebino. Zaključijo s pomembno ugotovitvijo, da je Mask R-CNN primerno za detekcijo stavb, saj klasične konvolucijske nevronske mreže ne ohranijo podrobne spektralne informacije pri detekciji objektov, kar privede do slabših rezultatov. Druga pomembna ugotovitev je pomen parametra sidrnega polja, ki pomembno vpliva na kakovost prepoznave manjših objektov ter večjih, spektralno raznolikih objektov. Zbrane rezultate navedenih raziskav predstavljamo v preglednici 1.

Preglednica 1: Primerjava karakteristik in metod preteklih raziskav s Faster R-CNN, Mask R-CNN in MS-FCN

Nevronska mreža	Uporabljena metoda	Podatki	Prostorska ločljivost [m]	Točnost [%]	Študija
Faster R-CNN	podporni vektorji	satelitski, pankromatski	0,46	88	Shetty in Mohan, (2018)
Faster R-CNN-VGG	rotacijska matrika za prepoznavo stavb	letalski R-G-B	0,26	70	Wen et al. (2019)
Mask R-CNN-VGG	rotacijska matrika za prepoznavo stavb	letalski R-G-B	0,26	91	Wen et al. (2019)
Faster R-CNN-Resnet101	rotacijska matrika za prepoznavo stavb	letalski R-G-B	0,26	91	Wen et al. (2019)
Mask R-CNN-Resnet101	rotacijska matrika za prepoznavo stavb	letalski R-G-B	0,26	91	Wen et al. (2019)
Mask R-CNN	algoritem regularizacije mej	satelitski R-G-B	0,5	88	Zhao et al. (2018)
Mask R-CNN	podporni vektorji	letalski R-G-B	0,3	90	Ji et al. (2019)
MS-FCN	podporni vektorji	letalski R-G-B	0,3	84	Ji et al. (2019)
Mask R-CNN-S	majhno sidrno polje	letalski R-G-B	0,5	85	Zhou et al. (2019)
Mask R-CNN-L	veliko sidrno polje	letalski R-G-B	0,5	81	Zhou et al. (2019)

Geodetski inštitut Slovenije na letni ravni izvaja samodejno klasifikacijo stavb z namenom posodabljanja prostorskih evidenc Geodetske uprave Republike Slovenije in monitoringa prostora za Ministrstvo za okolje in prostor. Klasifikacija stavb se izvaja z metodami strojnega učenja, z objektno klasifikacijo, in sicer z uporabo podpornih vektorjev (angl. *support-vector machines*) in naključnih gozdov (angl. *random forest*), kjer je ključni podatek tudi digitalni model površja. Namen prispevka je preizkus nove metode detekcije stavb z uporabo globokega učenja na dveh lastno izdelanih podatkovnih zbirkah stavb, ki sta popolnoma neodvisni od uporabe digitalnega modela površja.

V okviru raziskave smo izdelali dve podatkovni zbirki stavb, in sicer z uporabo barvnih ortofotov DOF050 (R-G-B) in z uporabo barvnih bližnjih infrardečih ortofotov DOF050IR (BIR-R-G) prostorske ločljivosti 0,5 metra iz leta 2019. Z uporabo barvnih bližnje infrardečih ortofotov dobimo drugačno predstavo o

terenu, saj so predmeti z visoko odbojnostjo v infrardečem spektru (npr. zdrava vegetacija) prikazani z rdečo barvo, medtem ko so predmeti z močno odbojnostjo v rdečem spektru prikazani zeleno in predmeti z močno odbojnostjo v modrozelenem spektru prikazani modro (Oštir, 2006). Glavna prednost uporabe barvnih bližnje infrardečih ortofotov je predvsem, da lahko že na podlagi spektralnega podpisa lažje ločimo stavbe od vegetacije.

Podatkovni zbirki stavb smo zapisali v formatu MS COCO (Lin et al., 2014), ki učne primerke stavb predstavi z zapisom JSON (angl. *JavaScript Object Notation*). Tak način označevanja je hiter in pregleden, zato smo ga uporabili pri izdelavi podatkovne zbirke stavb. Naučili smo osem različnih modelov klasifikacije stavb z uporabo predučenih uteži podatkovnih zbirk MS COCO in ImageNet (Deng et al., 2009). Uporabili smo učenje s prenosom znanja, ker ga uporablja Mask R-CNN, da se modeli, učeni na podatkovni zbirki uporabnika, hitreje generalizirajo. Uspešnost naučenih modelov smo preverili na izbranem testnem vzorcu stavb v Sloveniji in vsak model ovrednotili z izračunom evalvacijskih metrik. Zanimalo nas je predvsem, kako uspešno Mask R-CNN prepozna stavbe, kako kakovostne so prepoznane maske stavb in ali predlagana metoda ponuja možnosti za nadaljnjo uporabo.

2 ŠTUDIJSKO OBMOČJE IN OPIS PODATKOV

Ključni element uspešnega modela klasifikacije stavb z globokim učenjem je kakovostno izdelana podatkovna zbirka z velikim številom označenih učnih primerov. Študijsko območje za izdelavo dveh podatkovnih zbirk stavb je prikazano na sliki 1. Predstavlja območje pomurske in severovzhodne podravske regije. Območje velikosti 1387 km^2 vsebuje 98.425 evidentiranih stavb na dan 28. 3. 2020. Tipologija stavb je večinoma ruralna, z izjemo Murske Sobote, ki ima urbano tipologijo stavb. Strehe so različnih oblik in barv, prevladuje kritina rdeče, rjave in temno sive ali črne barve. Veliko učnih primerov streh vsebuje tudi sončne kolektorje. Območje smo izbrali zaradi dobrega poznavanja, in ker vsebuje širok nabor raznovrstnih streh, ki prispevajo k podrobnosti podatkovne zbirke stavb.

Slika 1: Prikaz študijskega območja za izdelavo podatkovne zbirke stavb.

3 METODE

V poglavju predstavimo metodologijo za izdelavo podatkovne zbirke stavb. Najprej opredelimo razred stavba, čemur sledi izdelava učnih vzorcev in njihov pregled za prikazano študijsko območje na sliki 1.

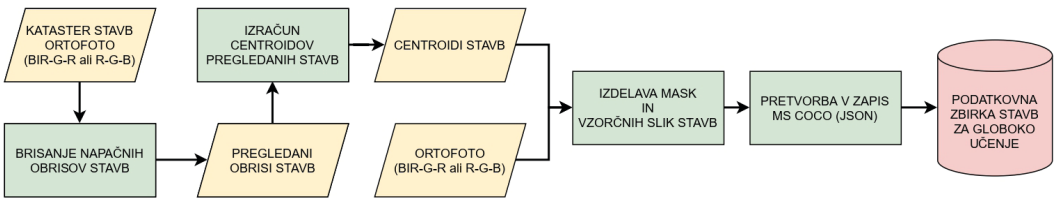
3.1 Opredelitev razreda stavba

Namen katastra stavb je evidentiranje vseh stavb v Republiki Sloveniji. Pojem stavbe in dela stavbe opredeljuje ZEN (Uradni list RS, št. 47/06). Podatki, ki se vodijo o stavbah in njihovih delih, so opisani v Zbirki topografskih podatkov (DTM) (GURS, 2020). DTM je zakonsko pokrit z 11. členom ZDGRS (Uradni list RS, št. 25/14 in 61/17). Stavbo opredelimo kot objekt, ki je trajno na enem mestu (Boguszewski et al., 2020). Primer pravilnega in upoštevanega obrisa stavbe je predstavljen na levi strani slike 2. Visoke stavbe pomenijo težavo, saj maska stavbe, pridobljena iz katastra stavb, ne sovpada s stavbo na posnetku. Takih stavb pri pripravi podatkovne zbirke nismo upoštevali.

Slika 2: Primeri pravilnih (levo) in nepravilnih (desno) učnih vzorcev stavb.

3.2 Metodologija izdelave podatkovne zbirke stavb

Korake izdelave podatkovne zbirke stavb za globoko učenje prikazuje diagram poteka na sliki 3, vsak posamezni korak je podrobneje opisan v nadaljevanju.



Slika 3: Diagram poteka izdelave podatkovne zbirke stavb za globoko učenje.

Kataster stavb, ki smo ga uporabili za izdelavo učnih vzorcev stavb, smo pred samo izdelavo počistili. Izločili smo obrise stavb, ki se po generalizirani podrobni namenski rabi prostora (MOP, 2020) na-

hajajo na: (1) območjih proizvodnih dejavnosti, (2) območjih in omrežjih prometne infrastrukture, (3) območjih komunikacijske infrastrukture in (4) območjih energetske infrastrukture. Pred izločanjem je bilo na študijskem območju 98.425 obrisov stavb. Najprej smo izločili 1271 obrisov stavb, med katere spadajo predvsem večji industrijski objekti. Te smo izločili zato, ker so objekti preveliki in jih ne moremo ustrezno predstaviti na vzorčni sliki velikosti 128 x 128 pikslov. Izločili smo še stavbe, ki se po spektralnih lastnostih enačijo s prometno infrastrukturo, predvsem s cestami in transformatorji, saj so obrisi teh premajhni, da bi algoritem masko transformatorja zaznal in ji izračunal slikovne koordinate izdelanih mask. Stavbe, ki so na območjih energetske, komunikacijske in prometne infrastrukture, so netipične. Želeli smo izdelati podatkovno zbirko, ki je čim bolj uravnotežena in ima čim manj osamelcev. Iz pregledanih in počiščenih obrisov stavb smo izračunali nove centroide ter jih uporabili pri pripravi vzorčnih slik in binarnih mask.

3.3 Pregled vzorčnih slik in mask stavb

Z ročnim pregledom učnih primerkov izboljšamo kakovost podatkovne zbirke, kar vpliva na boljše rezultate detekcije stavb. Najznačilnejše napake pri obrisih stavb so mešanje s travniki, njivami in sencami. Razlog za to je neposodobljenost katastra stavb. Za študijsko območje z 98.425 obrisi stavb smo odkrili 8226 primerov takih napak, kar predstavlja 8 % vseh obrisov. Napačnih obrisov stavb nismo upoštevali pri izdelavi vzorčnih slik in binarnih mask. Po pregledu izdelanih vzorčnih slik in binarnih mask smo pridobili 58.000 učnih primerov za izdelavo podatkovne zbirke stavb. Nekaj primerov najbolj značilnih napak prikazujemo na sliki 4.

Slika 4: Primeri napačnih obrisov stavb. Najznačilneše napake so mešanje s travniki, njivami in sencami. Tudi primeri delnega prekrivanja obrisa in stavbe niso uporabni.

Podatkovno zbirko smo razdelili v razmerju 80/20, kjer smo 80 % učnih primerov uporabili za učenje in 20 % za validacijo. Nato smo iz novo izračunanih centroidov stavb izdelali maske stavb in vzorčne slike stavb velikosti 128 x 128 pikslov. Vsaka stavba je v podatkovni zbirki predstavljena z masko in pripadajočo vzorčno sliko. Po tem smo podatkovni zbirki pretvorili v zapis MS COCO, ki, kot rečeno, za označitev učnih primerov uporablja zapis JSON. V prvem koraku smo označili lokacije vseh vzorčnih slik in pripadajočih binarnih mask glede na razred stavba, v drugem koraku smo pretvorili vzorčne slike in maske stavb v zapis JSON, ki ga uporablja podatkovna zbirka MS COCO. Podrobnosti o samem postopku je predstavil Šanca (2020).

3.4 Prenos znanja iz predučenih nevronskih mrež

Nevronske mreže so za specifične naloge računalniškega vida že izdelane in naučene na večjih podatkovnih zbirkah, kot sta ImageNet (Deng et al., 2009) in MS COCO. Te so dostopne uporabnikom, ki uteži lahko uporabijo za učenje novih modelov na podlagi lastne podatkovne zbirke. Primer takega učenja imenujemo učenje s prenosom znanja ali preneseno učenje (angl. *transfer learning*). Preneseno učenje deluje z inicializacijo že naučenih uteži modela z namenom boljšega prilagajanja uteži pri učenju na drugi podatkovni zbirki. Pri pridobivanju značilk iz slik uporabimo glavo nevronske mreže in učimo le novi klasifikator, da se predhodno naučene uteži filtrov optimizirajo za nove naloge na podlagi druge podatkovne zbirke, s tem proces učenja modelov zahteva precej manj časa (Ramon et al., 2019). Pri učenju nevronske mreže Mask R-CNN lahko uporabimo predučene uteži dveh velikih podatkovnih zbirk, to sta MS COCO in ImageNet.

Podatkovna zbirka MS COCO je vodilna podatkovna zbirka za detekcijo in segmentacijo objektov, vsebuje anotacije za različne probleme: (1) detekcijo objektov (angl. *object detection*), (2) detekcijo ključnih točk (angl. *keypoint detection*), (3) segmentacijo objektov (angl. *stuff segmentation*), (4) pan-optične segmentacije (angl. *panoptic segmentation*) in (5) opisovanja slik (angl. *image captioning*). Vsebuje 91 slikovnih kategorij, od teh ima 82 kategorij več kot 5000 označenih primerov slik. Skupno število označenih učnih primerov zbirke MS COCO je 2,5 milijona na 328.000 slikah (Lin et al., 2014).

ImageNet velja za največjo podatkovno zbirko nasploh, vsebuje več kot 15 milijonov ročno označenih slik visoke ločljivosti, ki so kategorizirane po hierarhiji WorldNet v 22.000 razredov. ImageNet omogoča: (1) klasifikacijo objektov (angl. *object classification*), (2) detekcijo objektov (angl. *object detection*) in (3) lokalizacijo posameznih objektov (angl. *single object localisation*) (Deng et al., 2009).

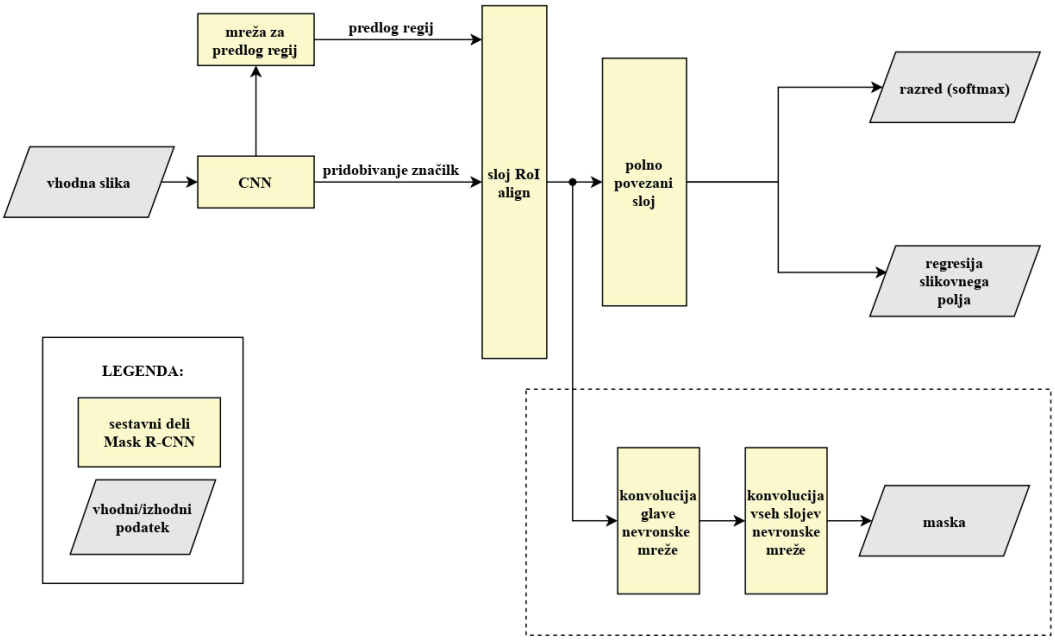
Mask R-CNN sestavljata dva dela: (1) konvolucijska mreža, namenjena pridobivanju značilk iz slik, in (2) glava nevronske mreže, namenjena klasifikaciji, prepoznavi slikovnega polja in napovedi maske prepoznanega objekta, ki se napove ločeno za vsak objekt ali interesno območje (angl. *Region of Interest*) (Zhao et al., 2018). Mask R-CNN sestavljata dve hrbtni arhitekturi, in sicer ResNet101, ki vsebuje 101 konvolucijskih slojev za klasifikacijo objektov in FPN (angl. *Feature Pyramid Network*), ki iz podobe pridobiva značilke različnih meril (Šanca, 2020). Pri učenju vseh modelov smo uporabili enake hiperparametre, edino hitrost učenja smo zmanjšali za 10-kratno vrednost pri modelih, katere uteži smo fino učili. Hiperparametre za učenje podajamo v preglednici 2, nastavili smo jih na podlagi pregledanih preteklih raziskavah. Vsi našteti hiperparametri v preglednici 2 so podrobneje opisani v (Šanca, 2020).

Preglednica 2: Uporabljene vrednost hiperparametrov pri učenju modelov

Ime hiperparametra	Vrednost
velikost slike za učenje (angl. image shape)	128 x 128 px
hitrost učenja (angl. learning rate)	0,001 in 0,0001
velikost serij (angl. batch size)	1000
število ponovitev na epoho (angl. steps per epoch)	1000
število ponovitev validacije (angl. validation steps)	50
hrbtna arhitektura (angl. backbone architecture)	ResNet-101
velikost sidrnih polj (angl. anchor box)	(8, 16, 32, 64, 128)
število predlaganih regij na posamezno sliko	32
skupni čas učenja (angl. total train time)	328 ur

Modele smo učili na računalniku z naslednjimi specifikacijami:

- centralna procesna enota: Intel(R) Core(TM) i9-9900X CPU @ 3.50GHz,
- pomnilnik: 64 GB RAM, DDR4 (4X 16 GB DIMM DDR4 Synchronous 2400 MHz),
- grafična kartica: NVIDIA GeForce RTX 2080 SUPER.



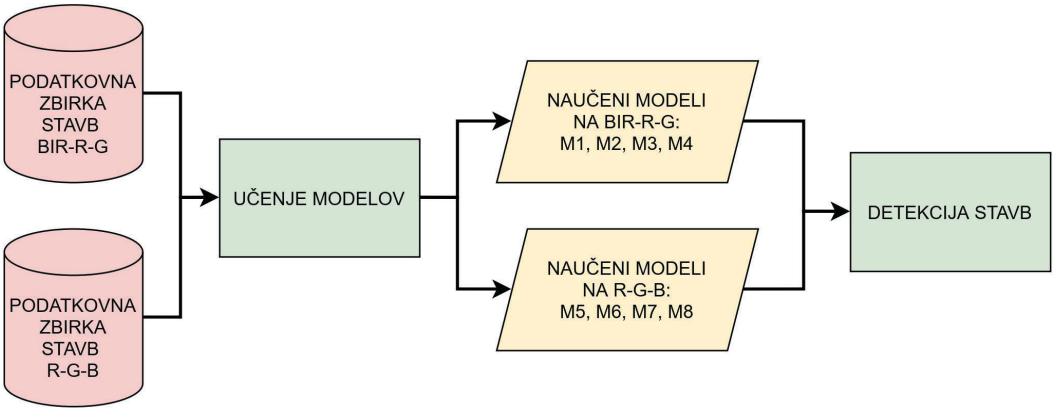
Slika 5: Arhitektura Mask R-CNN za primer detekcije stavb. V prvem koraku vhodna slika potuje skozi konvolucijsko nevronsko mrežo (CNN), ki pridobiva značilke. V drugem koraku mreža za predlog regij naključno predlaga regije na vhodni sliki, na podlagi katere se ustvarijo napovedi glede na opredeljen razred v podatkovni zbirki. Višjenivojske informacije se združujejo v nižjenivojske z uporabo sloja RoiAlign, ki deluje kot združevalni sloj in se razdeli v dva dela. V prvem delu se rezultati RoiAlign premikajo čez polno povezane sloje, na podlagi katerih se napove razred objekta in izračuna regresija slikovnega polja. V drugem delu se rezultati RoiAlign premikajo čez konvolucijske sloje glave nevronske mreže in potem čez vse sloje nevronske mreže, tako se ustvari pikselska maska za vsako regijo interesa (angl. *region of interest*) in se slika segmentira po pikslih glede na prepoznan razred.

V raziskavi primerjamo uspešnost osmih učenih modelov za klasifikacijo stavb, predstavljeni so v preglednici 3.

Preglednica 3: Naučeni modeli na podatkovni zbirki BIR-R-G in R-G-B

Ime modela	Podatkovna zbirka	Uporabljeni sloji za učenje	Število epoh	Uporabljene uteži	Hitrost učenja	Čas učenja [h]
M1	BIR-R-G	glavni sloji	100	MS COCO	0,001	39
M2	BIR-R-G	glavni sloji	100	ImageNet	0,001	40
M3	BIR-R-G	vsi sloji	200	MS COCO	0,0001	42
M4	BIR-R-G	vsi sloji	200	ImageNet	0,0001	44
M5	R-G-B	glavni sloji	100	MS COCO	0,001	37
M6	R-G-B	glavni sloji	100	ImageNet	0,001	39
M7	R-G-B	vsi sloji	200	MS COCO	0,0001	43
M8	R-G-B	vsi sloji	200	ImageNet	0,0001	44

Število epoh pomeni število ponovitev prehoda celotne podatkovne zbirke čez Mask R-CNN. Pri modelih M1, M2, M5, M6 smo učili glavo nevronske mreže (angl. *head layers*), pri modelih M3, M4, M7, M8 vse sloje nevronske mreže (angl. *all layers*). Hitrost učenja predstavlja velikost koraka pri izračunu gradienta funkcije izgube, ki se med učenjem nevronske mreže iterativno minimizira. Pri učenju si izberemo primerno hitrost učenja, da se izognemo premajhnemu ali prekomernemu prileganju uteži (angl. *underfitting/overfitting*) (Šanca, 2020). Postopek učenja modelov prikazujemo na sliki 6.



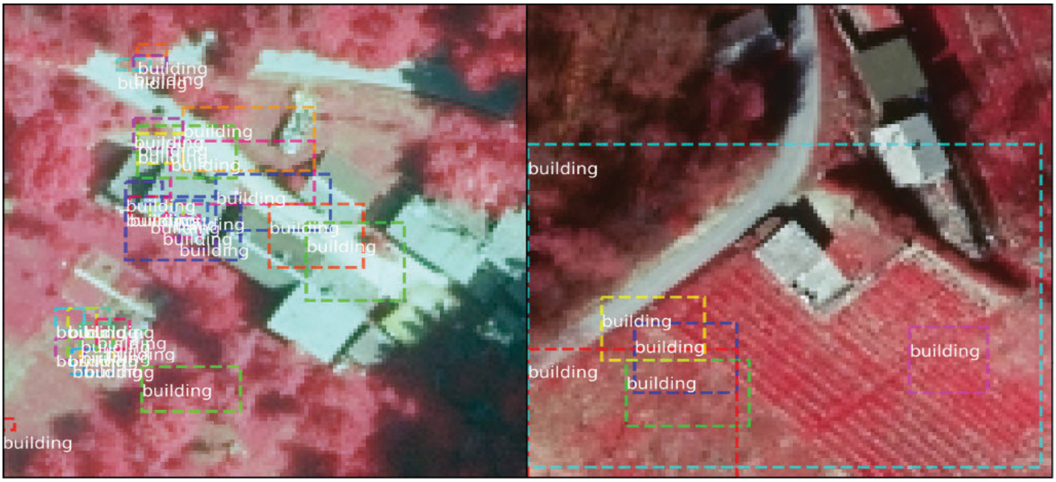
Slika 6: Diagram poteka učenja modelov za detekcijo stavb. Modeli M3, M4 in M7, M8 so fino učeni modeli še 100 epoh po učenju glave nevronske mreže. Detekcijo stavb smo izvedli ločeno z vsakim modelom.

4 REZULTATI IN NJIHOVO OVREDNOTENJE

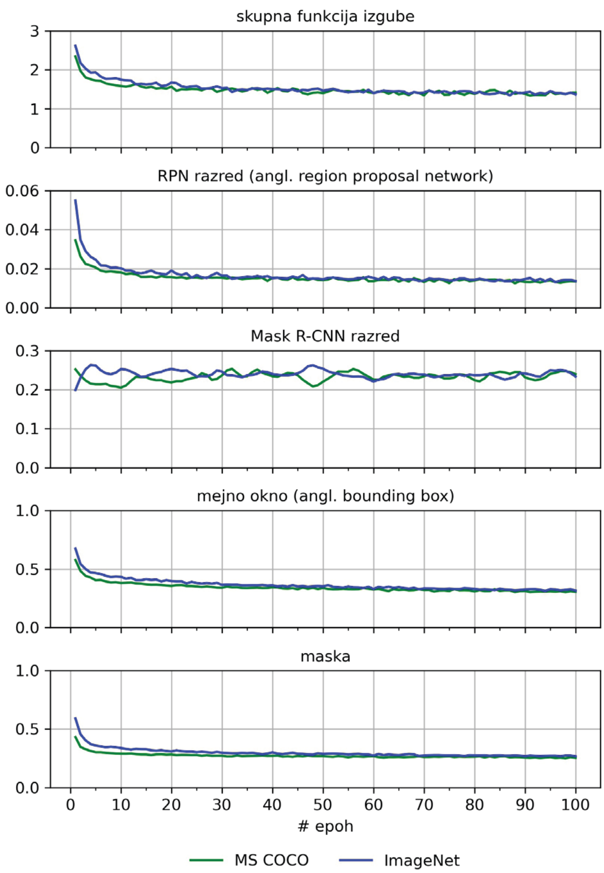
Uspešnost prenosa znanja smo ocenili na podlagi skupne funkcije izgube v procesu učenja. Skupne funkcije izgube validacijskega procesa nismo upoštevali za ocenjevanje stabilnosti modela, saj je na velikem deležu učnih parov obris stavb iz katastra stavb nekonsistenten, zato učen model uspešno napove stavbo, kjer ni obrisa, kar validacijski proces zazna kot napačno klasificiran objekt. Odločili smo se, da bomo validacijo uspešnosti modelov opravili z izračunom evalvacijskih metrik (poglavje 4.1). Uspešnost modelov klasifikacije stavb smo preverili na izbranem testnem območju izven območja podatkovne zbirke za učenje. Izbrali smo 300 primerov za testiranje uspešnosti učenih modelov, pri tem smo upoštevali kriterij raznolikosti streh, predvsem barvo in obliko strehe, prisotnost sončnih kolektorjev na strehi, sence ipd. Najprej primerjamo naučene modele med sabo in potem še dobljene rezultate M3 in M4 s katastrom stavb. Rezultate detekcije stavb predstavimo na slikah, kjer primerjamo napovedi vseh naučenih modelov na BIR-R-G. Dobljene obrise stavb primerjamo s katastrom stavb.

4.1 Funkcije izgube po prenosu znanja in primerjava modelov

Primer detekcije stavb s predučenim modelom MS COCO (R-G-B), ki smo ga uporabili za inicializacijo naših uteži pri učenju na lastni podatkovni zbirki, predstavljamo na sliki 7. Kot pričakovano je detekcija stavb brez prenosa znanja na izdelani podatkovni zbirki stavb napačna. Osnovni model MS COCO ali ImageNet le naključno predlaga prostorska polja.



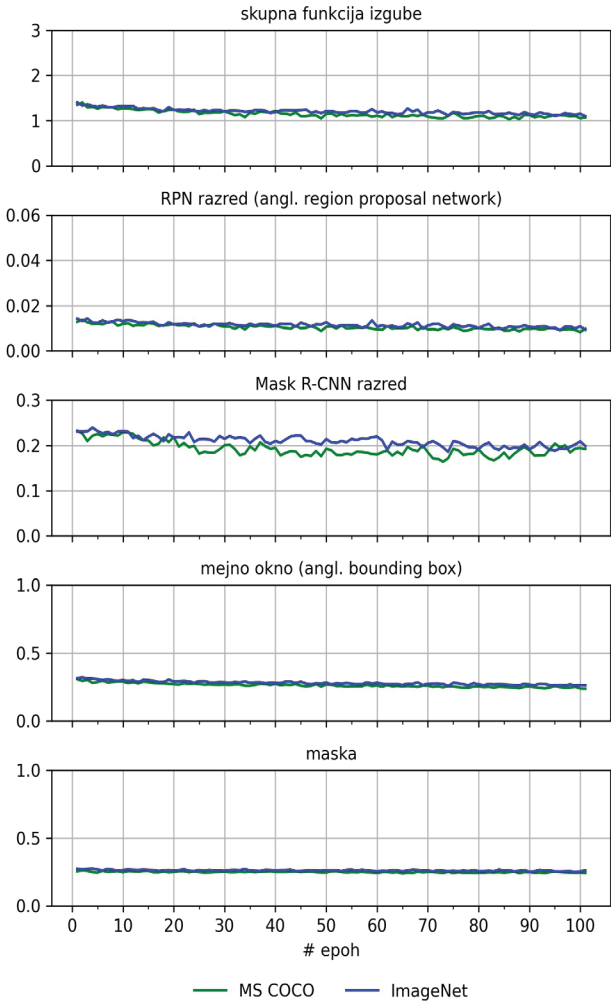
Slika 7: Primer detekcije stavb z osnovnim, še ne naučenim modelom MS COCO.



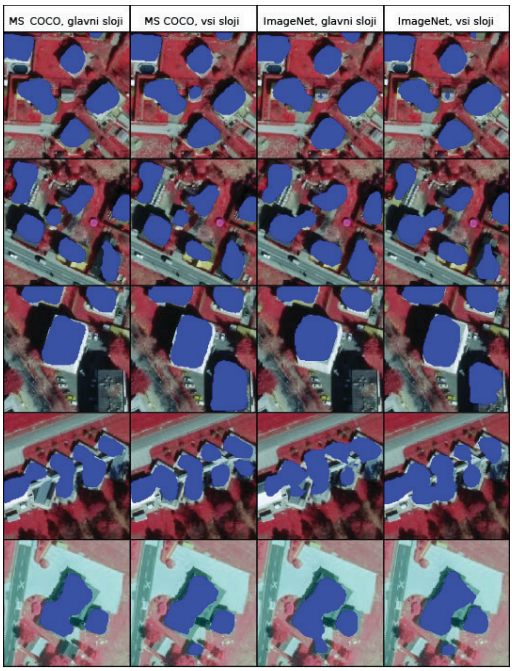
Slika 8: Prikaz funkcij izgube učenja za M1 in M2 na podatkovni zbirki BIR-R-G. Pri obeh modelih smo učili glavne sloje nevronske mreže v trajanju 100 epoh.

Skupna funkcija izgube Mask R-CNN je vsota klasifikacijske izgube, izgube slikovnega polja in izgube maske. Za primer detekcije stavb nas najbolj zanima izguba maske, saj ta predstavlja mero za natančnost klasifikacije maske stavbe. Na sliki 8 vidimo, da je skupna funkcija izgube po prenosu znanja modela MS COCO minimalno boljša od funkcije izgube modela ImageNet. To velja tudi po posameznih funkcijah izgube za klasifikacijo, regresijo in masko. Pri učenju se funkcija izgube minimizira in proti koncu učenja stabilizira (približno po 65. epohi učenja), kar pomeni da ni potrebe po daljšemu učenju. Vrednost funkcije izgube za razred Mask R-CNN (angl. *mask loss*) se med učenjem ne spreminja veliko. Razlog za to je, da imamo v podatkovni zbirki stavb le en sam razred, ki se napoveduje pri detekciji. Prikazane so funkcije izgube le v fazi učenja.

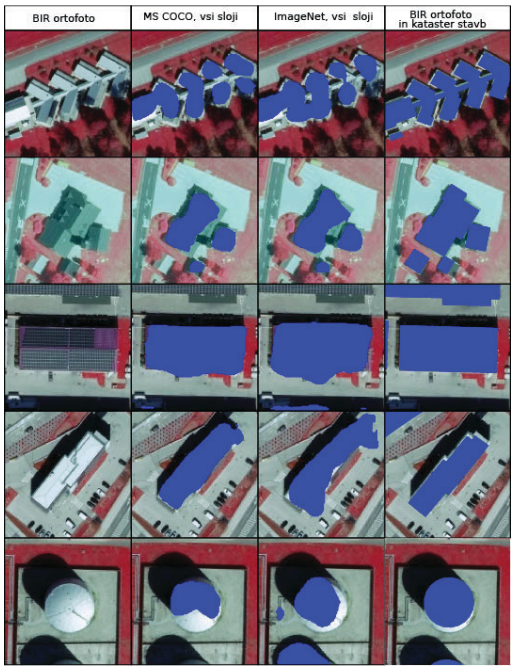
Pri finem učenju uteži se vrednosti funkcij izgube dodatno zmanjšajo, razen funkcije izgube za masko, ki že na začetku stabilizira in se med učenjem ne izboljšuje več, kar pri detekciji pomeni, da se obrisi prepoznanih stavb izrazito ne spreminjajo.



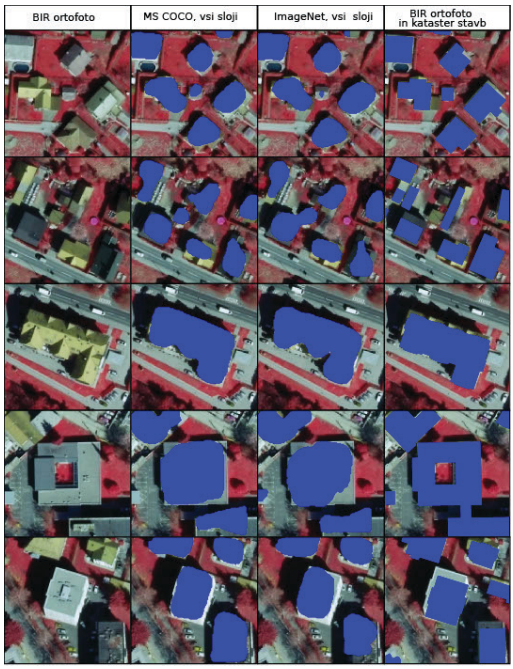
Slika 9: Prikaz funkcij izgube učenja za M3 in M4 na podatkovni zbirki BIR-R-G. Pri obeh modelih smo učili vse sloje nevronske mreže v trajanju 200 epoh.



Slika 10: Primerjava uspešnosti naučenih modelov: M1, M3, M2, M4.



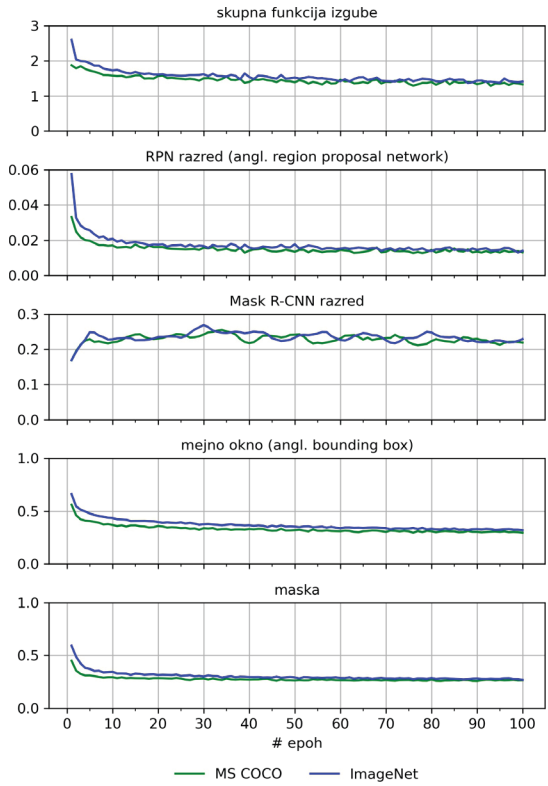
Slika 11: Rezultati detekcije stavb z M3 in M4 ter primerjava s katastrom stavb.



Slika 12: Rezultati detekcije stavb z M3 in M4 ter primerjava s katastrom stavb.

Rezultati detekcije stavb z učenimi modeli na BIR-G-R so prikazani na slikah 10, 11 in 12. Modela M3 in M4 prepoznata manjše stavbe, a obrisi prepoznanih stavb v primerjavi z M1 in M2 ostajajo zelo podobni. Fino učenje uteži z učenjem vseh slojev Mask R-CNN se je izkazalo kot pomembno pri izboljšanju točnosti napovedi, ampak ni ključnega pomena za izboljšanje obrisov prepoznanih stavb v primerjavi z obrisi, ki jih dobimo le z učenjem glave nevronske mreže.

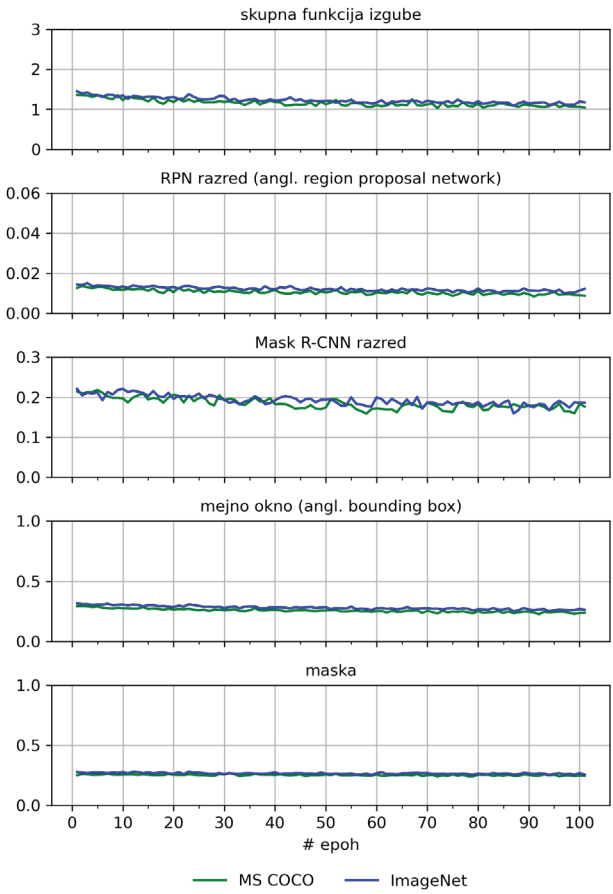
Na sliki 13 so prikazane funkcije izgube učenja M5 in M6, učenih na R-G-B, in na sliki 14 funkcije izgube M7, M8, učenih na R-G-B. Vidimo, da so funkcije izgube R-G-B modelov približno enake kot funkcije izgube BIR-R-G modelov.



Slika 13: Prikaz funkcij izgube učenja M1 (zelena) in M2 (modra) na podatkovni zbirki R-G-B. Pri obeh modelih smo učili glavne sloje nevronske mreže 100 epoh.

Zanimala nas je tudi razlika med uspešnostjo modelov, ki smo jih naučili na podatkovni zbirki BIR-R-G in R-G-B. Rezultate detekcije stavb z modeli R-G-B prikazujemo na sliki 15.

Razlike med modeloma M1, M2 BIR-R-G in modeloma M5, M6 R-G-B so minimalne tudi pri prepoznavanju obrisov stavb. To velja tudi za fino učene modele M3, M4 in M7, M8. V primerjavi z ortofoti BIR-R-G so rezultati detekcije stavb na R-G-B nekoliko boljši, ker je osnovni model od MS COCO, učen na barvnih slikah iz vsakdanjega življenja (R-G-B). V obeh primerih učenja se izkaže, da je prenos znanja iz predučenih uteži podatkovne zbirke MS COCO ali ImageNet učinkovita rešitev v primerjavi z učenjem modelov iz začetka.



Slika 14: Prikaz funkcij izgube učenja M3 (zelena) in M4 (modra) na podatkovni zbirki R-G-B. Pri obeh modelih smo učili vse sloje nevronske mreže 200 epoh.

Slika 15: Primer detekcije stavb z M7 na R-G-B ortofotu.

4.2 Uspešnost naučenih modelov

Uspešnost učenih modelov klasifikacije v strojnem učenju ovrednotimo na podlagi vzorčnih slik za testiranje. Klasifikacija stavb predstavlja primer binarne klasifikacije, pri kateri uspešnost napovedi predstavimo z matriko zamenjav.

	Napoved	
Resnica	TP – pravilno pozitivne	FN –napačno negativne
	FP – napačno pozitivne	TN – pravilno negativne

Pravilno klasificirani primeri spadajo v pravilno pozitivne TP (angl. *true positive*) in pravilno negativne TN (angl. *true negative*). Napačno klasificirani primeri spadajo med napačno negativne FN (angl. *false negative*) in napačno pozitivne FP (angl. *false positive*). Rezultate napovedi uporabimo za izračun evalvacijskih metrik, s katerimi ovrednotimo uspešnost modelov detekcije stavb. Uporabljene enačbe za izračun evalvacijskih metrik smo povzeli po Fetai et al. (2021). Točnost (angl. *accuracy*) predstavlja delež pravilnih napovedi glede na vse napovedi modela, pri binarni klasifikacije se enačba poenostavi:

$$točnost = \frac{TP + TN}{TP + FP + TN + FN} := \frac{TP + TN}{število\ vseh\ testnih\ \ primerov} \tag{1}$$

Priklic (angl. *recall*) pove delež pravilno napovedanih primerov glede na vse pozitivne primere:

$$priklic = \frac{TP}{TP + FN} \tag{2}$$

Združeno metriko natančnosti in priklica definira mera F1 (angl. *F1 score*), uporabimo jo, ko želimo prikazati samo eno mero za uspešnost modela:

$$F1 = \frac{2 \cdot točnost \cdot priklic}{točnost \cdot priklic} \tag{3}$$

Za učene modele smo prešteli delež pravilno in napačno prepoznanih stavb in izračunali evalvacijske metrike. Rezultate prikazujemo v preglednici 4.

Preglednica 4: Ovrednotenje uspešnosti klasifikacije stavb po modelih

	Podatkovna zbirka	Uteži	TP	TN	FP	FN	Točnost	Priklic	Mera F1
M1	BIR-R-G	MS COCO	246	0	38	16	0,8200	0,9389	0,8754
M2	BIR-R-G	ImageNet	243	0	43	14	0,8100	0,9455	0,8725
M3	BIR-R-G	MS COCO	296	0	3	1	0,9867	0,9966	0,9916
M4	BIR-R-G	ImageNet	294	0	4	2	0,9800	0,9932	0,9866
M5	R-G-B	MS COCO	249	0	36	15	0,8300	0,9432	0,8830
M6	R-G-B	ImageNet	244	0	41	15	0,8133	0,9421	0,8730
M7	R-G-B	MS COCO	298	0	1	1	0,9933	0,9967	0,9950
M8	R-G-B	ImageNet	296	0	1	2	0,9867	0,9933	0,9900

Naučeni modeli so si po uspešnosti detekcije stavb zelo podobni. Modeli, učeni 200 epoh (M3, M4 in M7, M8), so bolj uspešni predvsem pri prepoznavi natančnih obrisov stavb in pri prepoznavi stavb tudi na robu vzorčnih slik. M1, M2 in M5, M6 ne prepoznajo majhnih stavb ali stavb, ki so delno prekrite z vegetacijo.

5 SKLEP IN RAZPRAVA

V članku smo predstavili postopek detekcije stavb z Mask R-CNN od izdelave podatkovne zbirke do detekcije obrisov stavb za osem modelov. Modele smo učili s prenosom znanja iz predučenih uteži podatkovnih zbirk MS COCO in ImageNet. Primerjali smo uspešnost klasifikacije stavb pri uporabi ortofotov R-G-B in barvno bližnje infrardečih ortofotov BIR-R-G. Izdelana podatkovna zbirka, namenjena prepoznavi stavb z globokim učenjem, ponuja možnosti za nadaljnjo uporabo, ker se lahko dodatno razširi na območje celotne države in se mogoče v prihodnosti uporablja za iskanje in vzdrževanje podatkov o stavbah. Dobljeni rezultati potrjujejo, da je Mask R-CNN uporaben in primeren za klasifikacijo stavb, kar trdijo tudi sorodne raziskave (Ji et al., 2019, in Zhou et al., 2019).

V Sloveniji je to po našem poznavanju prvi primer detekcije stavb z globokim učenjem in prav tako prvi primer uporabe prenosa znanja iz podatkovnih zbirk MS COCO in ImageNet, ki vsebujeta le R-G-B slike vsakdanjih predmetov, v podatkovno zbirko stavb, izdelano iz R-G-B in BIR-R-G ortofotov, katere slike so predstavljene s povsem drugačne perspektive. Uspešnost prenosa učenja iz ene domene v drugo kaže na fleksibilnost tovrstnih mrež. Na podlagi tega lahko trdimo, da je prenos učenja učinkovita metoda za posodabljanje modelov, ki so učenih na različnih podatkih. Učenje modelov od začetka je dolgotrajen proces, ki omogoča izgradnjo boljših modelov. Uporaba prenosa znanja na takih modelih z uporabo novih podatkov enakega barvnega prostora bi omogočila hitro in učinkovitejše posodabljanje modelov in predvsem hitrejše učenje.

S pridobljenimi izkušnjami imamo nekaj predlogov za izboljšave nadaljnje raziskave. Prvi primer izboljšav je izdelava prepoznanih mask stavb pravilnejših oblik z algoritmom detekcije obrisov, kot navajajo Zhao et al. (2018). Prepoznane maske stavb lahko v naslednjem koraku vektoriziramo in s tem omogočimo integracijo v geografski informacijski sistem. Podatkovno zbirko lahko nadgradimo s kombinirano uporabo barvnega bližnje infrardečega ortofota in normiranega digitalnega modela površja (nDMP), s katerim bi že pri pripravi podatkov ločili strehe od reliefa. Podatkovno zbirko lahko razširimo tako, da bi vsebovala stavbe celotne Slovenije, s čimer bi model učili na območju celotne države.

ZAHVALA:

Raziskava je bila delno opravljena v okviru aplikativnega raziskovalnega projekta L2-1826, ki ga sofinancirajo Javna agencija za raziskovalno dejavnost Republike Slovenije, Geodetska uprava Republike Slovenije in Ministrstvo za obrambo, ter raziskovalnega programa P2-0406 in projekta J2-9251, ki ju financira Javna agencija za raziskovalno dejavnost Republike Slovenije. Hvala tudi Oddelku za gradbeništvo Høgskulen på Vestlandet za delno sofinanciranje raziskave.

Literatura in viri:

Glej str. 575.

Šanca S., Oštir K., Mangafić A. (2021). Zaznavanje stavb z uporabo nevronskih mrež, učenih s prenosom znanja. Building detection with convolutional networks trained with transfer learning. Geodetski vestnik, 65 (4), 559-593.
DOI: https://doi.org/10.15292/geodetski-vestnik.2021.04.559-593

Simon Šanca, mag. inž. geod. geoinf.
Høgskulen på Vestlandet
Inndalsveien 28, NO-5063 Bergen, Norveška
e-naslov: simon.sanca@hvl.no

Alen Mangafić, mag. prost. načrt.
Geodetski inštitut Slovenije
Jamova cesta 2, SI-1000 Ljubljana, Slovenija
e-naslov: alen.mangafic@gis.si

prof. dr. Krištof Oštir, univ. dipl. inž. fiz.
Univerza v Ljubljani, Fakulteta za gradbeništvo in geodezijo
Jamova cesta 2, SI-1000 Ljubljana, Slovenija
e-naslov: kristof.ostir@fgg.uni-lj.si