TECHNICAL SCIENCES

AUTOMATED SEMANTIC SEGMENTATION OF TERRAIN IMAGES

Momot A.
National Technical University of Ukraine
"Igor Sikorsky Kyiv Polytechnic Institute, PhD, Senior Lecturer
Zemliakov O.
National Technical University of Ukraine
"Igor Sikorsky Kyiv Polytechnic Institute, Student
https://doi.org/10.5281/zenodo.15397815

Abstract

The article deals with the problem of automated semantic segmentation of cartographic images using deep neural networks. Modern approaches to segmentation are analysed, and the choice of the U-Net architecture as the optimal one for processing cartographic data is substantiated. The model was trained using pre-training on a large set of images and subsequent retraining on a specialised dataset. The quality of segmentation was evaluated by the Accuracy, F1 Score, and IoU metrics. The results obtained indicate the high efficiency of the developed model for the tasks of accurate terrain object detection and open up prospects for further improvement.

Keywords: semantic segmentation, cartographic images, deep neural networks, U-Net, machine learning, image processing.

Introduction

In today's era of rapid development of digital technologies, automation of cartographic image processing is becoming increasingly important. The growing amount of spatial data obtained through satellite imagery and aerial photography, as well as the use of unmanned aerial vehicles, requires efficient methods of analysing them. Segmentation of cartographic images is one of the critical stages of this process, as it allows to identify and classify terrain objects, which is the basis for further mapping, monitoring and planning.

Traditional approaches to segmentation based on manual processing or simple algorithms are increasingly giving way to automated methods, in particular those based on machine learning and computer vision technologies [1]. The use of deep convolutional neural networks significantly improves the accuracy and speed of map data processing, reducing dependence on the subjective human factor.

Automation of the segmentation process is of particular importance in the context of the need to promptly update cartographic information, which is critical both for civilian industries such as urban planning, agriculture, and environmental monitoring and for defence needs, in particular in the context of modern military conflicts.

At the same time, the use of neural networks for cartographic image segmentation has a number of specific challenges related to the variety of image scales, the use of conventional labels and a high level of object generalisation. Therefore, it is important to develop specialised data mining systems adapted to the peculiarities of cartographic data.

The aim of this study is to develop an automated system for segmentation of cartographic images using deep neural networks. To achieve this goal, it is necessary to analyse existing segmentation methods, justify the choice of neural network architecture, prepare and process training data, train and test the developed model, and analyse its effectiveness.

Review of recent works

In recent years, researchers have been actively working on improving methods of automated semantic segmentation of cartographic images, focusing mainly on the use of deep neural networks. Traditional segmentation methods, such as thresholding, regional fusion or active contouring algorithms, although still relevant for certain tasks, show limitations when processing images with a large number of object classes, fuzzy boundaries or high texture variability. That is why methods based on deep learning technologies are coming to the fore today [2].

A detailed overview of modern approaches to semantic segmentation of remote images is given in [3]. The authors aimed to systematise existing methods for segmenting high-resolution images using deep learning. The study considered both classical convolutional neural networks and modern approaches based on transformers. The use of semi-supervised and unsupervised learning to address the lack of annotated data typical for remote sensing tasks was analysed. The results of the analysis showed that hybrid models that combine local properties of CNNs and global features of transformers achieve the best results in complex scenes. The authors emphasise that the future of segmentation lies in the development of efficient models with low resource consumption and the ability to generalise to new data types.

The problem of limited data volumes in segmentation tasks in cartography was discussed in detail in [4]. Their work was aimed at studying the effectiveness of various deep learning strategies when the number of training examples is small. The authors analysed the possibilities of using pre-trained models on large datasets and adapting them to specific mapping tasks through fine-tuning. In addition, they considered methods of data augmentation and weak learning. The results showed that the use of pre-trained models significantly improves the quality of segmentation, even in the case of a limited number of annotated examples.

The main conclusion of the authors is the need to combine different techniques to achieve acceptable accuracy in low-data scenarios.

Another important area of segmentation technology development is the use of transformers, which is described in detail in [5]. The purpose of their study was to analyse the effectiveness of architectures based on the self-attention mechanism in remote sensing tasks. The authors studied both pure transformers and hybrid CNN+Transformer models, emphasising their ability to better take into account the global context of the scene compared to classical convolutional networks. The study demonstrated that transformers are particularly effective for classifying complex multispectral data and long-term context analysis of large images. However, the authors also pointed out the high computational cost of transformers, which requires further improvement of their architecture for practical applications.

The aspect of implementing semantic segmentation algorithms in real time was covered in [6]. They conducted a comparative analysis of the effectiveness of various lightweight segmentation models suitable for use on devices with limited computing resources. The aim of the study was to evaluate the trade-off between processing speed and segmentation accuracy when applying models to remote sensing tasks. The authors found that models such as ENet and BiSeNet demonstrate a satisfactory balance between performance and quality, while heavier architectures that provide higher accuracy are unable to operate in real time without a significant reduction in resolution. The main conclusion of the work is the need for further optimisation of architectures to ensure segmentation at the edge device level.

Thus, current research in the field of semantic segmentation of map data focuses on the development of hybrid architectures, improving the efficiency of learning with limited data, optimising models for real-time operation, and extending the application to three-dimensional spatial data. These trends form the basis for further improvements in automated mapping information processing systems.

Description of deep learning model

Choosing a neural network architecture is one of the key stages of building an automated map image segmentation system. It determines not only the accuracy of object recognition on the map but also the speed of data processing, resistance to image variations, and the ability to generalise the results to new data sets. Given the specifics of mapping information, which is often characterised by complex object boundaries, heterogeneous textures, and a wide range of scales, the neural network architecture must meet a number of stringent requirements: high segmentation resolution, preservation of spatial context, and efficient processing of fine details.

Among a wide range of modern architectures, such as Fully Convolutional Networks (FCN), DeepLabv3+, LinkNet, Attention U-Net, and others, the U-Net architecture may be the most suitable for solving the task at hand [7]. Originally developed for biomedical segmentation tasks, this model has proven its effectiveness in numerous industries that require accurate object extraction in complex images. The simplified architecture of the U-Net model is shown in Fig. 1. The input to the network is the original image X, and the output is the corresponding predicted segmentation mask.

The main advantage of the U-Net architecture is its symmetrical encoder-decoder structure with skip connections between the respective layers. The encoder gradually reduces the image resolution by extracting more and more complex features, while the decoder restores the spatial resolution using information from the corresponding encoder layers. The presence of direct connections avoids the loss of important spatial information, which is critical for accurate segmentation of fine details and object boundaries in mapping images. Thanks to this feature, U-Net demonstrates excellent results even in cases of small amounts of training data, which is a common situation when working with specialised map sets.

Another important advantage of U-Net is the relative simplicity of training. Compared to more complex architectures, such as DeepLabv3+, which require fine-tuning of atrus convolution and spatial pyramidal smoothing parameters, U-Net requires fewer computing resources and converges faster during training. This makes it particularly suitable for projects with limited hardware capabilities or requirements for rapid data processing.

In addition, modifications of the basic U-Net architecture, such as Attention U-Net, U-Net++, MultiResUNet and Double U-Net, open up wide opportunities for further adaptation of the model to specific features of cartographic images. For example, the use of attention mechanisms allows the model to focus on the most informative areas of the image, and the extension of standard connections with dense blocks or multi-level paths improves the segmentation of objects with complex morphology [8].

The high efficiency of U-Net in the tasks of map data segmentation is confirmed by the results of numerous modern studies. In particular, in [9], it was demonstrated that the basic U-Net model provides competitive results of satellite image segmentation in urban area classification tasks. The authors showed that the model maintains high accuracy on fine structures, such as roads and rivers, where other methods often lose details.

Another study [10] considered the use of U-Net for multispectral segmentation of landscape objects. The researchers noted the advantage of U-Net in restoring the exact boundaries of objects with minimal shape distortion, which is especially important for accurate cartographic modelling.

Thus, the choice of the U-Net architecture for the development of an automated cartographic image segmentation system is reasonable and optimal. The combination of accuracy, efficiency, ease of setup and wide support in the scientific community makes this architecture the most suitable for solving the problem in this study.

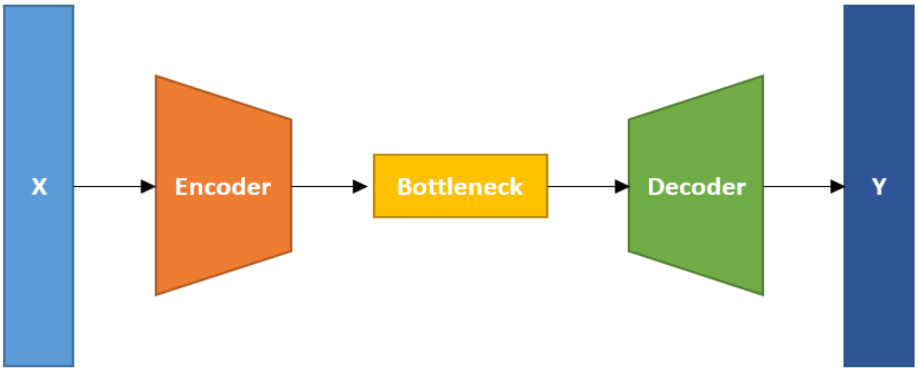

Figure 1. Simplified U-Net architecture

Results and discussion

The open dataset Semantic Drone Dataset [11] is used to train the neural network. This dataset focuses on the semantic understanding of urban scenes to improve the safety of autonomous flights and landing procedures. The images show more than 100 terrains taken by UAVs from an altitude of 5 to 30 metres above the ground. The photos are marked with 22 classes of objects, such as pedestrians, trees, grass, buildings, streets, etc. A high-resolution camera was used to capture images with a size of 6000x4000 pixels (24 Mpx). The training set contains 400 public images, and the test set consists of 200 private images. Examples of images from the training dataset are shown in Figure 2.

Figure 2. Examples of two images from the training dataset

The neural network model was written in Python using the TensorFlow framework. The Transfer learning approach was used to speed up the experiment. The U-Net network used in the study was pre-trained on the ImageNet dataset. The Adam optimiser and the SparseCategoricalCrossentropy loss function were used to train the network on the data from the Semantic Drone Dataset. The training lasted for 100 epochs with a battle size of 32. The metrics used were Accuracy and F1 Score to evaluate the classification accuracy and Intersection over Union (IoU) to determine the segmentation quality. Table 1 includes the reduction of the loss function and changes in the accuracy metrics during training.

Table 1. Training results

Epoch	Loss	Accuracy	F1 Score	IoU
20	1.20	0.56	0.58	0.51
40	0.94	0.71	0.69	0.55
60	0.58	0.79	0.74	0.63
80	0.54	0.83	0.78	0.69
100	0.35	0.89	0.82	0.77

As a result of training, the network showed a 92.5% correct answer rate on the test set. The IoU on the test set was 0.79. The trained model accurately identifies the main classes of objects and their boundaries. The separation of roads, buildings, and green spaces is particularly well visible. There are minor differences in the details between the true and predicted masks, but the overall structure is preserved. In Figure 3, you can see part of a street, a house, and people.

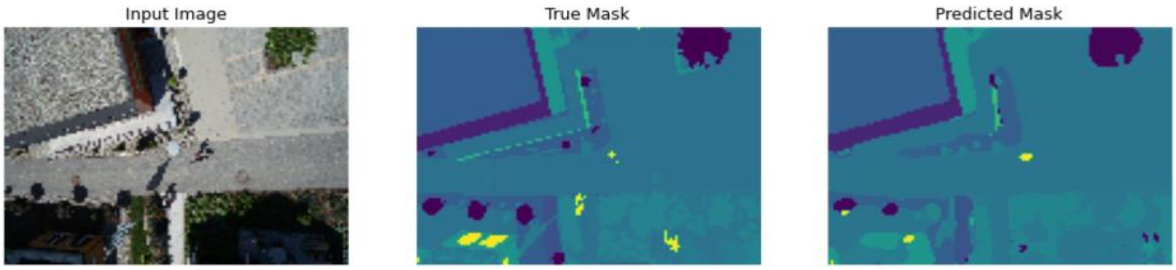

Figure 3. Comparison of the segmentation result with the test street image

Figure 4 shows the square from the maximum flight height of the drone. The model identified the main buildings well (purple) but has some inaccuracies in the details compared to the true mask. The overall segmentation structure is preserved.

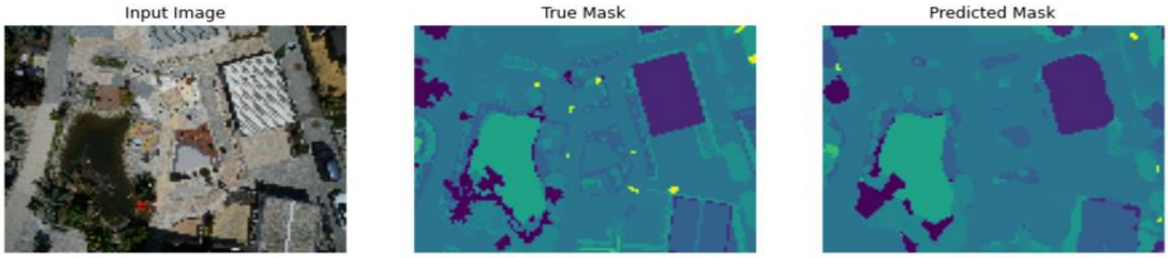

Figure 4. Comparison of the segmentation result with the test image of the square

The analysis of the obtained results shows that the use of pre-training on the large ImageNet dataset has significantly accelerated the model convergence on a specialised set of cartographic images. After 40 epochs of training, a significant decrease in the loss function and a steady increase in the accuracy and F1 Score metrics were observed, indicating an effective transfer of knowledge from the general classification task to the specific task of semantic segmentation.

Of particular note is the fact that the model demonstrates high accuracy in identifying large and well-structured objects, such as buildings and road surfaces. At the same time, the segmentation of small objects, such as pedestrians or thin road infrastructure elements, is less accurate. This is an expected result, as training was performed on reduced-size copies of the original images to optimise computing resources, which partially leads to the loss of fine details [12].

Comparison of the predicted masks with the actual markings shows that the main model errors are not so much related to incorrect object classification as to inaccuracies within the segments. In some places, the model "blurs" the contours of objects or combines neighbouring classes into a single area, especially in cases of complex scenes with a large number of intertwined elements. This result indicates the potential benefit of further using attention mechanisms or multi-level augmentation in the learning process to improve boundary localisation.

The evaluation of IoU values for individual classes shows that the best results are achieved for objects with clear texture features, such as grass or asphalt roads. In contrast, objects with similar visual characteristics, such as bushes and trees, are more likely to be confused with each other. This once again highlights the importance of a comprehensive training set, including enriching it with examples that present complex segmentation cases.

In general, the results indicate the high efficiency of using the U-Net architecture for the tasks of automated semantic segmentation of cartographic images in urban environments. It is worth noting that further accuracy improvement is possible by introducing deeper network modifications, improving the data preparation strategy and applying specialised post-processing techniques for segmentation results.

Conclusion

In this paper, a method for automated semantic segmentation of cartographic images based on deep convolutional neural networks is developed and investigated. The study analyses the current state of development in the field of semantic segmentation. Based on a detailed literature review and specific requirements for map data processing, the article justifies the choice of U-Net architecture as the most suitable for the task. Its ability to preserve spatial details, efficient encoder-decoder structure, and high segmentation accuracy, even with a limited amount of training data, provided optimal results.

The model was trained using the open data set "Semantic Drone Dataset". The use of the transfer learning approach and a pre-trained encoder based on ImageNet significantly accelerated the learning process and increased the model's ability to generalise. During the training, a steady increase in key metrics such as Accuracy, F1 Score and IoU was observed, which reached 89% and 0.77 on the validation set, respectively. On the test set, the model demonstrated an overall accuracy of 92.5% and an IoU of 0.79, which confirms the effectiveness of the developed approach.

Experimental results showed that the model is particularly good at segmenting large and clearly structured objects, such as buildings, roads, and green spaces. At the same time, the segmentation of small or visually similar objects requires further improvement. This indicates the prospects of using data augmentation methods, attention mechanisms, and multilayer feature processing to improve the quality of segmentation.

In general, the study demonstrated that deep learning methods, in particular the U-Net architecture, are highly effective tools for automating the analysis of map images. Prospects for further research include testing modifications of U-Net with attention mechanisms, extending the model to multispectral and hyperspectral data, and optimising the network for use on embedded systems in real-time.

References:

1. R. Galagan, S. Andreiev, N. Stelmakh, Y. Rafalska, and A. Momot, "Automation of polycystic ovary syndrome diagnostics through machine learning algorithms in ultrasound imaging," Applied Computer Science, vol. 20, pp. 194-204, 2024. DOI: 10.35784/acs-2024-24.
2. A. Momot, V. Kretsul, O. Muraviov, and R. Galagan, "Automated defect detection in printed circuit boards based on the YOLOv5 neural network," *The Paton Welding Journal*, no. 4, pp. 46-52, 2024. DOI: 10.37434/tpwj2024.04.07.
3. L. Huang, B. Jiang, S. Lv, Y. Liu, and H. Zhang, "Deep-Learning-Based Semantic Segmentation of Remote Sensing Images: A Survey," IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, vol. 16, pp. 1-28, 2023. DOI: 10.1109/JSTARS.2023.3335891.
4. A. Yu et al., "Deep Learning Methods for Semantic Segmentation in Remote Sensing with Small Data: A Survey," Remote Sensing, vol. 15, no. 20, p. 4987, 2023. DOI: 10.3390/rs15204987.
5. A. A. Aleissaee et al., "Transformers in Remote Sensing: A Survey," arXiv preprint arXiv:2209.01206, 2022.
6. C. Broni-Bediako, J. Xia, and N. Yokoya, "Real-Time Semantic Segmentation: A Brief Survey & Comparative Study in Remote Sensing," arXiv preprint arXiv:2309.06047, 2023.
7. O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional Networks for Biomedical Image Segmentation," in *Medical Image Computing and Computer-Assisted Intervention - MICCAI 2015*, vol. 9351, N. Navab, J. Hornegger, W. M. Wells, and A. F. Frangi, Eds. Cham: Springer, 2015, pp. 234-241. DOI: 10.1007/978-3-319-24574-4_28
8. A. Momot, M. Zaboluieva, and R. Galagan, "Automated segmentation of ultrasound medical images using the Attention U-Net model," *Norwegian Journal of Development of the International Science*, no. 128, pp. 56-60, 2024. DOI: 10.5281/zenodo.10817342.
9. X. Liu, J. Zhang, Y. Wang, and L. Zhang, "Remote Sensing Image Segmentation with Improved U-Net for Urban Feature Extraction," IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, vol. 14, pp. 10253-10265, 2021. DOI: 10.1109/JSTARS.2021.3094122.
10. Y. Zhong, L. Ma, H. Li, and W. Chen, "Deep Learning Based Semantic Segmentation of High-Resolution Remote Sensing Images Using Multiscale and Multidepth Features," Remote Sensing, vol. 12, no. 15, p. 2425, 2020. DOI: 10.3390/rs12152425.
11. A. S. Alam, "Semantic Drone Dataset," Kaggle, 2019. [Online]. Available: https://www.kaggle.com/datasets/awsaf49/semantic-drone-dataset. [Accessed: Apr. 27, 2025].
12. Y. V. Steshenko, A. S. Momot, A. G. Protasov, and O. V. Muraviov, "Automation of the process of segmentation of images of metal surface defects using the neural network U-Net," *Technical Diagnostics and Non-Destructive Testing*, no. 2, pp. 34-40, 2023. DOI: 10.37434/tdnk2023.02.05.