UDC 004.8

A NEURAL NETWORK APPROACH TO SEMANTIC SEGMENTATION OF VEHICLES IN VERY HIGH RESOLUTION IMAGES

Kashtan V. Yu. – PhD, Associate Professor, Associate Professor of Department of Information Technology and Computer Engineering, Dnipro University of Technology, Dnipro, Ukraine.

Hnatushenko V. V. – Dr. Sc., Professor, Head of Department of Information Technology and Computer Engineering, Dnipro University of Technology, Dnipro, Ukraine.

Udovyk I. M. – PhD, Associate Professor, Dean of Information Technologies Department, Dnipro University of Technology, Dnipro, Ukraine.

Kazymyrenko O. V. – Postgraduate student of Department of Information Technology and Computer Engineering, Dnipro University of Technology, Dnipro, Ukraine.

Radionov Y. D. – Postgraduate student of Department of Information Technology and Computer Engineering, Dnipro University of Technology, Dnipro, Ukraine.

ABSTRACT

Context. The semantic segmentation of vehicles in very high resolution aerial images is essential in developing intelligent transportation systems. It allows for the automation of real-time traffic management and the detection of congestion and emergencies.

Objective. This work aims to develop and evaluate the effectiveness of a neural network approach to semantic segmentation in very high resolution aerial images, which provides high detail and correct reproduction of object boundaries.

Method. The DeepLab architecture with ResNet-101 as a backbone is used for gradient preservation and multiscale feature analysis. We trained on DOTA data and retrained on specialized sets with classes: vehicles, green areas, buildings, and roads. A loss function based on the Dice coefficient was applied to reduce the imbalance of classes. It effectively solves the class imbalance problem and improves the accuracy of segmenting objects of different sizes. Using ResNet-101 instead of Xception in the backbone network allows us to maintain the gradient as the network depth increases.

Results. Experimental studies have confirmed the effectiveness of the proposed approach, which achieves a segmentation accuracy of more than 90%, outperforming existing analogs. The use of multiscale feature analysis allows for preserving the texture features of objects, reducing false classifications. A comparative study with U-Net, SegNet, FCN8s, and other methods confirms the higher performance of the proposed approach in terms of mIoU (82.3%) and Pixel Accuracy (95.1%).

Conclusions. The experiments confirm the effectiveness of the proposed method of semantic segmentation of vehicles in ultra-high spatial resolution images. Using DeepLab v3+ResNet-101 significantly improves the quality of vehicle segmentation in an urbanized environment. Excellent metric performance makes it promising for infrastructure monitoring and traffic planning tasks. Further research will focus on adapting the model to new datasets.

KEYWORDS: semantic segmentation, vehicles, deep neural networks, ResNet-101, DeepLab, multi-scale analysis, very high resolution images.

ABBREVIATIONS

UAVs is a Unmanned Aerial Vehicle;
RGB is a red, green, blue;
CNN is a convolutional neural network;
FCNs is a fully convolutional network;
RPN is a region proposal network;
DOTA is a dataset for object detection;
PA is a pixel accuracy;
MA is a mean accuracy;
mIoU is a mean intersection over the union;
FP is a False Positive;
FN is a False Negative;
TN is a True Negative;
TP is a True Positive.

NOMENCLATURE

X is an input image;
$f(a)$ is a function of structural features;
P is a predicate that defines the segmentation rule;
$s_{i,j}$ is a name of the region $s_{i,j} \in S$;
$P(s_{i,j})$ is an indication of the neighborhood model;
$\nabla f(a)$ is a gradient;
x_m and x_n are elements of the pixel set X;
$F(w)$ is a neural network model;
X_{norm} is a normalized image;
μ is a mean value of the image X;
σ is a standard deviation of the image X;
$F(x, \{Wi\})$ is a mapping function that represents a sequence of layers with parameters $\{Wi\}$;
y is a residual building block;
ϵ is a small positive number added to avoid division by zero in the case of no intersection between the predicted and real segments;
$D(p,q)$ is a measure of similarity between p and q;
p is a predicted segmentation;
q is a real segmentation;
L_{Dice} is a loss function;
TP is a number of correctly classified positive pixels;
FP is a number of false positive pixels;
FN is a number of false negative pixels;
TN is a number of correctly classified negative pixels;
N is a number of image pixel categories;
TP_i is a number of correctly classified pixels of class i;
FP_i is a number of false positive pixels for class i;
FN_i is a number of false negative pixels for class i.
T_i is a total number of pixels of class i;

DOI 10.15588/1607-3274-2025-3-8

X_{ii} is a total number of pixels with actual type i and prediction type i;

X_{ji} is a total number of pixels with actual type i and prediction type j.

INTRODUCTION

Uncrewed aerial vehicles (UAVs) are an effective tool for high-precision aerial surveys, providing fast and detailed ultra-high-resolution images that can reach an accuracy of several centimeters [1]. It ensures high object detail and provides operational aerial photography with minimal resource costs. One of the parameters affecting the quality of the data is the camera angle. Vertical imaging (perpendicular to the camera's optical axis) provides high accuracy but has a limited coverage area. Low-angle images (15°–30°) expand the coverage area of the scene, improve the depth of perspective, and allow for better analysis of objects in the image. Images acquired at high tilt angles (approximately 60°) provide a much wider coverage area, including horizons, making them suitable for complex analysis of traffic flows and urban environments.

UAVs combine compactness, mobility, and efficiency, which makes it possible to obtain data in real-time and adapt the research methodology depending on the specifics of the territory or object under observation. Equipped with various sensors (RGB cameras, multispectral and hyper-spectral sensors, LiDAR, and thermal imaging systems), UAVs provide multispectral information necessary for thematic image processing and environmental change analysis. Using UAVs for automated vehicle recognition and segmentation is an urgent task in security, logistics, and traffic management [2].

Traditional tracking methods based on ground-based cameras and satellite imagery have limitations associated with limited spatial coverage, high dependence on weather conditions, and delays in data updates. Using UAVs for vehicle recognition and segmentation can overcome these shortcomings by providing adaptability to the information collection process, high spatial resolution, and the ability to update data quickly.

Semantic segmentation is one of the approaches for automated analysis of UAV images. This computer vision method consists of classifying each pixel of an image according to its class [3]. Semantic segmentation allows for high-accuracy vehicle detection in complex urban and road scenes [4].

The research is relevant due to the need to develop new methods of vehicle recognition for intelligent transportation systems, including traffic monitoring, logistics process management, and road safety improvement. The use of UAV imagery in combination with deep learning architecture will increase the accuracy and speed of automated vehicle detection in real-time.

The object of study is the process of semantic segmentation of vehicles in ultra-high-resolution images.

Constructing a neural network model for semantic segmentation is complex and multi-component. It is caused by segmentation accuracy and stability of model training, which mainly depends on the amount and quality of training data, neural network architecture, choice of the loss function, and optimization strategies. In particular, it is necessary to balance computational costs and the model's generalization ability to process ultra-high-resolution images efficiently. It requires adaptation of feature extraction mechanisms and adjustment of the loss function to solve the class imbalance problem.

The subject of study is a neural network methodology for semantic vehicle segmentation based on the DeepLab + ResNet architecture with multi-scale feature extraction, loss function adaptation, and retraining on specialized datasets.

The purpose of the work is to develop and evaluate the effectiveness of a neural network approach to semantic segmentation in very high resolution aerial images, which provides high detail and correct reproduction of object boundaries.

1 PROBLEM STATEMENT

Suppose a set of image pixels $X=\{x_{i,j}\}$ is given, where each pixel $x_{i,j}$ is characterized by structural features defined by the function $f(a)$. The predicate P is also given, establishing the segmentation rule $f(a)$.

The problem of image segmentation is to find a partition of the set P into $S=\{s_{i,j}\}$, where $s_{i,j}$ is connected to non-empty subsets such that for any two pixels x_m, $x_n \in s_{i,j}$ the condition $P(x_m, x_n)$=True is fulfilled, i.e., they belong to the same segment according to the segmentation rule. The boundaries of the regions $s_{i,j}$ are determined by the contrast gradient $\nabla f(a)$ and the spatial dependencies between neighboring pixels. The background region is the set of pixels with the highest or lowest contrast relative to the segmented regions.

In general, segmentation can be considered $f(a) \to S$.

In particular, $s_{i,j}$ is the name of the region $s_{i,j} \in S$, and $P(s_{i,j})$ is an indication of the neighborhood model that characterizes the object.

2 REVIEW OF THE LITERATURE

The existing approaches to semantic vehicle segmentation can be divided into traditional methods and methods based on deep learning. Conventional methods of vehicle segmentation involve manual feature extraction and machine learning methods, such as SVM, AdaBoost, and others, for classification [5]. These methods had significant limitations, requiring extensive preprocessing to extract features and set thresholds. It makes them difficult to apply to complex scenes in aerial images containing small objects. In addition, traditional methods are usually only capable of extracting surface objects, which limits their effectiveness in analyzing more complex and variable cases.

Due to deep learning, in particular through the implementation of convolutional neural networks (CNNs) and

DOI 10.15588/1607-3274-2025-3-8

fully convolutional networks (FCNs), the situation has changed, and semantic segmentation methods have been significantly improved. The authors in [6] proposed a general multimodal deep learning system that uses five types of fusion networks to integrate features of hyperspectral imagery, LiDAR imagery, and SAR imagery to improve image segmentation performance. The Deeplab series of models [7] is based on increasing convolutional layers, which solves the problem of resolution reduction that occurs at the stage of maximum layer fusion.

Two main categories of deep learning approaches to object detection are two-stage and one-stage algorithms. Two-stage algorithms, such as Fast R-CNN [8], identify regions of interest and localize and classify objects. For example, the method proposed in [9] showed satisfactory results for flying object detection using Faster R-CNN and VGG-16, achieving an average accuracy of 66% (mAP). However, these methods have a significant computational complexity and may be less effective in detecting small objects. In [10], parallel RPN (Region Proposal Networks) networks are used to improve the detection of dense areas in aerial photographs. The CNN-based method proposed by the authors of [11] uses Xception for classification and U-Net with ResNet18 as an encoder to accurately segment ships in optical satellite images, achieving an accuracy of over 84%. However, its application to vehicle segmentation in ultra-high resolution aerospace images has several limitations: differences in object characteristics, different spatial features of images, lack of specialized training, and limitations in selecting small structural objects when using U-Net. In [12], a method for detecting vehicles in aerial photographs using a convolutional neural network with double focal loss (MFL CNN) was proposed. The authors emphasize the complexity of the vehicle detection task, in particular, due to their small size and complex image background. The paper demonstrates the advantages of the proposed approach compared to the baseline models, which is confirmed by the results on the EAGLE and XWHEEL datasets.

However, the complexity of the model and the two-stage detection process do not meet real-time requirements. At the same time, one-step algorithms, such as YOLO [13, 14, 15], demonstrate significant advantages in speed and accuracy compared to two-step methods but also have certain limitations, particularly in solving false positives and complex background conditions.

Deep learning algorithms have significantly improved the accuracy and efficiency of object detection, including vehicle detection. It can automatically learn from large data sets and does not depend on manual feature selection. However, problems remain unresolved: a large number of false positives in object detection arise because some non-vehicle objects have a similar appearance to vehicles; existing CNN-based vehicle detectors always have two outputs: the coordinates of the bounding box and the probability that an object within this box is a vehicle arises in conditions of a complex background or high density of objects in the image.

3 MATERIALS AND METHODS

The neural network approach to semantic vehicle segmentation using UAV images based on the DeepLab + ResNet architecture using multi-level feature extraction is shown in Fig. 1.

The method starts with the loading of an aerospace image. Then, the input image is processed by the Backbone network, which was initialized with weights obtained during training on the DOTA dataset (Dataset for Object Detection in Aerial Images) [16] and then retrained on its specialized datasets for semantic segmentation with classes cars (individual vehicles, parking lots, roads); green area (vegetation, lawns, parks); buildings (residential and industrial buildings); roads (main highways, secondary streets, intersections). At this stage, preprocessing was performed [17]: normalization, scaling, and marking of objects to ensure correct training of the neural network by the formula (1):

$$X_{\text{norm}} = \frac{X - \mu(X)}{\sigma(X)} . \tag{1}$$

Deep neural networks with many layers connected in series are prone to the vanishing gradient problem. This problem occurs in error backpropagation when the gradients used to update the network weights decrease exponentially with the network depth approaching zero. As a result, the layers closer to the network input are practically not trained, limiting the network's ability to learn complex dependencies. The proposed methodology solves this problem using the ResNet-101 network instead of Xception as the backbone network. It allows us to maintain the gradient as the network depth increases and effectively extract features at different scales. It is achieved by adding the input to the output of one or more layers, allowing the gradients to propagate to the previous layers. The final training block (a residual building block) can be defined by the formula (2) [18]:

$$y = F(x, \{W_i\}) + x . \tag{2}$$

The encoder consists of a sequence of 1×1 convolutional layers to reduce the dimensionality of features without losing information and 3×3 convolution with ReLU activation, supplemented by MaxPooling operations for hierarchical aggregation of spatial and contextual information. The Multi-Scale Features mechanism provides multi-scale processing, including layers of global convolutional smoothing (Image Pooling) and further transformation through 1×1 convolution with ReLU activation. It allows the neural network to simultaneously analyze local and international contexts, improving the segmentation accuracy of objects of different sizes, including vehicles. However, multi-class segmentation faces the problem of class imbalance.

DOI 10.15588/1607-3274-2025-3-8

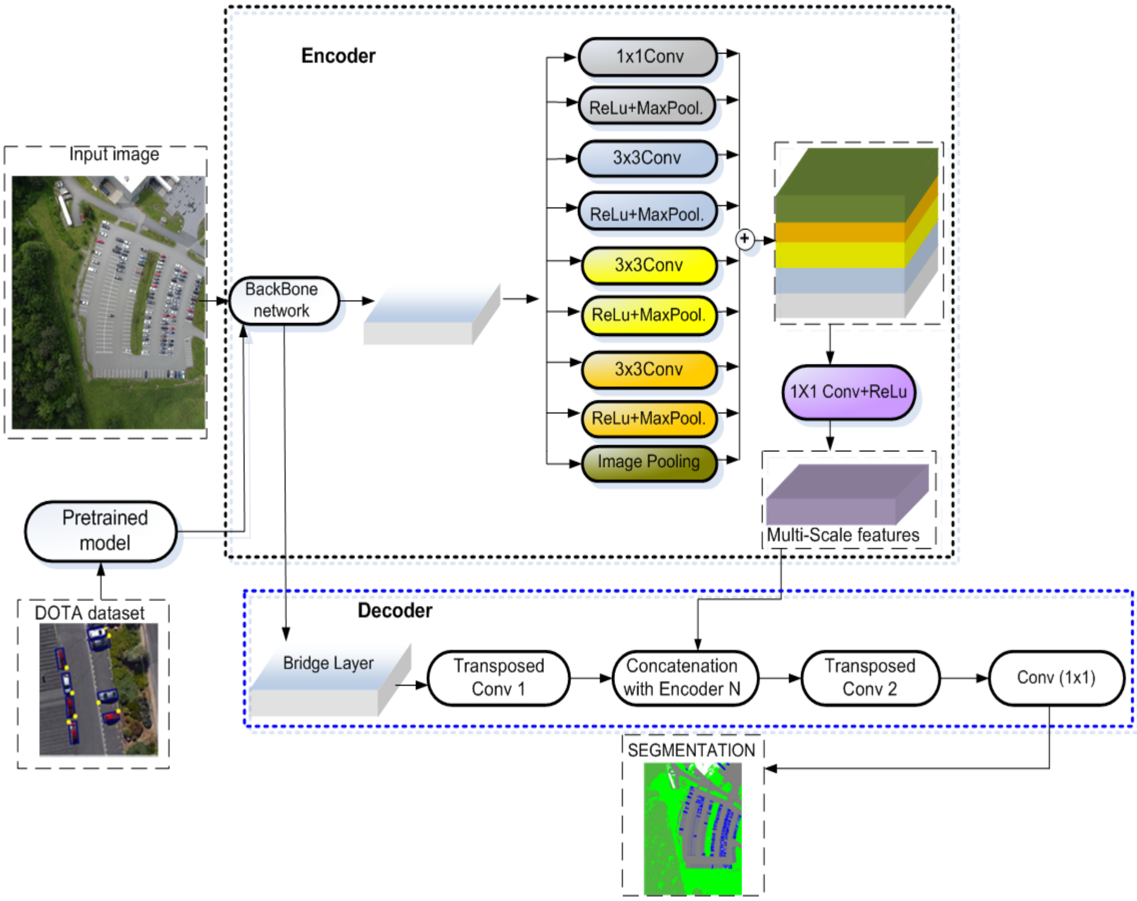


Figure 1 – Diagram of the proposed approach to semantic vehicle segmentation

It is when the number of samples of one class significantly exceeds that of others, leading to poorer recognition of less represented classes. In the worst cases, the model may completely ignore underrepresented classes if the number of their training samples is insufficient. For this reason, our method uses a customized loss function based on the Dice coefficient by the formula (3):

$$D(p,q) = \frac{2\sum_{i}^{N} p_i q_i + \in}{\sum_{i}^{N} p_i^2 + \sum_{i}^{N} q_i^2 + \in}. \tag{3}$$

The loss function is formulated using its complement by the formula (4):

$$L_{Dice} = 1 - D(p,q). \tag{4}$$

The decoder restores the spatial resolution of the segmentation image by sequentially using Transposed Convolution operations, which allow for the gradual restoration of the object structure. In addition, concatenation (skip connections) with the corresponding encoder layers is applied to preserve high-level information and improve segmentation detail. The final convolution layer (1×1 Convolution) reduces the image to the required channels for each segmentation class.

The proposed DeepLab + ResNet architecture efficiently extracts multi-scale features, contributing to segmentation accuracy by preserving spatial and semantic information.

4 EXPERIMENTS

For the experiments, we used a dataset consisting of images obtained from UAVs at a height of 15 cm and corresponding reference segmentation masks. The reference masks were created manually by experts, which ensured high quality annotations. The test images are presented as JPG files of 3037 x 3672 pixels, and the annotation file is given in XML format. The annotation contains the corresponding coordinates of the four vertices of the vehicle. The dataset was divided into training, validation, and testing. The training set of 1500 images was used to train the model, the validation set of 500 images was used to set hyperparameters and monitor the training process, and the test set was used to evaluate the model's generalization ability. The training was performed until the value of the loss function stabilized on the validation set. Augmentation of the data (methods of rotation, reflection, and scaling of images) was used to improve the model's generalization ability.

Three metrics were used to evaluate the quality of segmentation: pixel accuracy (PA), mean accuracy (MA), and mean intersection over the union (mIoU).

Pixel accuracy (PA) is one of the leading indicators that determines the level of segmentation accuracy at the level of individual pixels. It is the ratio of correctly classi-

DOI 10.15588/1607-3274-2025-3-8

fied pixels to the total number of pixels in the image. Formula (5) shows the calculation of PA [19]:

$$PA = \frac{TP}{TP + FP + FN + TN}. \qquad (5)$$

Pixel accuracy allows you to evaluate how well the model copes with the classification of each pixel, which is vital for segmentation at the level of a detailed image and for the accurate selection of vehicles in satellite images.

Mean accuracy (MA) is an indicator that reflects the average classification accuracy across all categories of objects in an image. This indicator makes it possible to assess how effectively the model copes with segmenting all types of objects in the image. Formula (6) shows the calculation [19]:

$$MA = \frac{1}{N}\sum_{i=1}^{N}\frac{TP_i}{TP_i + FP_i + FN_i}. \qquad (6)$$

Average precision provides a generalized measure of segmentation performance across all classes and indicates how well the model performs with different types of objects in the image.

The mean intersection of union (mIoU) is the most widely used indicator for assessing the quality of segmentation, as it allows us to determine the degree of coincidence between the segmentation results and the actual pixel values in the image. The mean intersection of union allows us to consider not only the accuracy for individual classes but also the overall level of segmentation, considering all categories of objects. Formula (7) shows the calculation [19]:

$$mIoU = \frac{(\frac{X_{ii}}{T_i} + \sum_{j=1}^{N}(X_{ji} - X_{ii}))}{N}. \qquad (7)$$

The mIoU metric is one of the best indices for a comprehensive assessment of segmentation results, as it allows for accuracy for both positive and negative pixels and provides a balanced evaluation based on all classes. Since this indicator considers intersections and merges between segmented courses, it gives a more objective assessment of the model quality, which is especially important for tasks with several class objects (for example, vehicles, roads, and other elements in images). In the experimental studies, the above metrics are used to compare the effectiveness of different segmentation models and to evaluate the results obtained using the proposed DeepLab + ResNet architecture. In particular, in the context of research on ultra-high spatial resolution images, the evaluation using PA, MA, and mIoU allows for a detailed analysis of the quality of vehicle segmentation.

5 RESULTS

Table 1 shows the results of correctly and incorrectly classified pixels.

Table 1 – Number of correctly and incorrectly classified pixels for different models

Model	TP	TN	FP
DeepLab v3	8200	9500	1200
U-Net	8100	9400	1300
SegNet	7000	9200	1600
FCN8s	6800	9100	1700
ENet	6600	8900	2000
Proposed method	8600	9700	900

Table 2 shows the results for the Loss metric.

Table 2 – Loss function values during training and validation

Model	Epochs	Loss (training)	Loss (validation)
DeepLab v3	100	0.7	0.8
U-Net	100	0.8	0.9
SegNet	100	0.4	0.5
FCN8s	100	0.5	0.6
ENet	100	0.6	0.7
Proposed method	100	0.3	0.4

Table 3 shows the results of training and validation accuracy for different models.

Table 3 – Training accuracy and validation results for different models

Model	Epochs	Accuracy (training)	Accuracy (validation)
DeepLab v3	100	0.91	0.88
U-Net	100	0.9	0.85
SegNet	100	0.85	0.8
FCN8s	100	0.8	0.65
ENet	100	0.75	0.7
Proposed method	100	0.95	0.9

Table 4 shows the results for the Pixel Accuracy (PA) metric.

Table 4 – PA metric results

Model	PA (%)
DeepLab v3	91.8
U-Net	90.1
SegNet	81.2
FCN8s	86.4
ENet	74.8
Proposed method	95.1

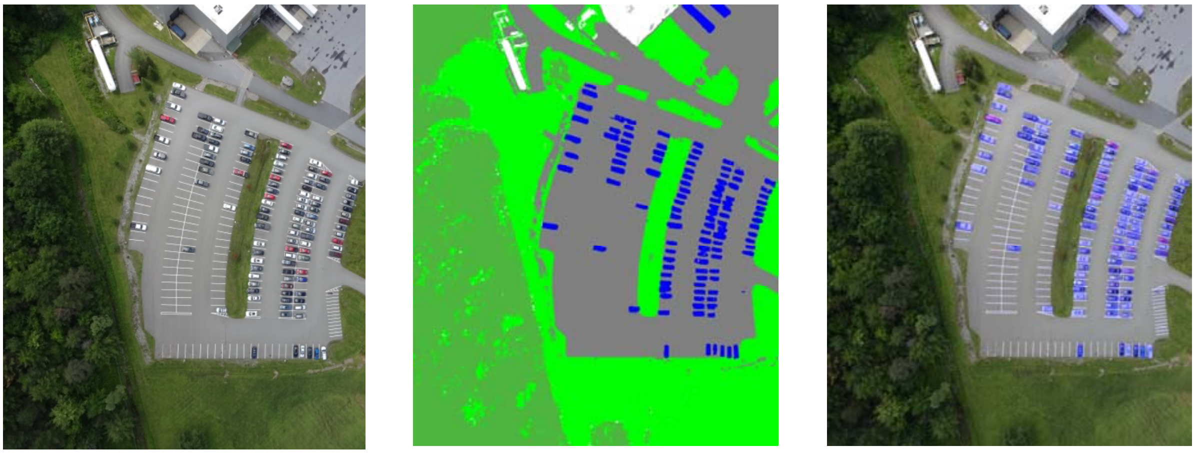
Table 5 shows the results for the Mean Intersection over Union (mIoU) metric.

Table 5 – Results of the mIoU metric

Model	mIoU (%)
DeepLab v3	74.0
U-Net	73.3
SegNet	56.7
FCN8s	56.7
ENet	70.8
Proposed method	82.3

Figure 2 shows the results of UAV image segmentation obtained using the proposed method. The image consists of three parts: the original image (Fig. 2a), a segmented image (Fig. 2b), and detected vehicles (Fig. 2c) with color coding of different classes of objects.

DOI 10.15588/1607-3274-2025-3-8



Legend trees; grass; roads; vehicles; unclassified

a b c

Figure 2 – UAV images: a – original dataset; b – result of proposed neural network approach to semantic vehicle segmentation; c – result of segmentation of the class "vehicles" on the original image using the proposed neural network approach

6 DISCUSSION

The results of the experimental study demonstrate the superiority of the proposed neural network approach over existing segmentation approaches in terms of key quality assessment metrics. The analysis of Loss, Accuracy, PA, and mIoU indicators confirms the effectiveness of training and high segmentation accuracy.

The results presented in Table 1 allow us to evaluate the effectiveness of various segmentation methods based on the TP, TN, FP, and FN metrics. The proposed method demonstrates the highest TP (8600) and TN (9700) values, indicating its ability to identify target objects and accurately classify the background. In addition, the proposed method has the lowest FP (900) and FN (600) values, which indicates a reduced number of false positive and false negative classifications. It confirms its high accuracy in detecting target objects while minimizing segmentation errors. Compared to other models, such as DeepLab v3, U-Net, and SegNet, the proposed method shows a better balance between correct and false classifications, making it practical for the semantic segmentation of transport vehicles.

The Loss metric reflects the discrepancy between the predicted and actual values, i.e., the lower the Loss value, the better the model fits the training data. According to Table 2, the proposed method demonstrates the lowest Loss values at the training (0.3) and validation (0.4) stages. It indicates that the proposed neural network architecture minimizes errors during training and generalizes acquired knowledge well to new and unknown data (validation sample). The low difference between the Loss values on the training and validation samples indicates the stability of the learning process and the absence of overtraining.

The Accuracy metric measures the percentage of correctly classified pixels and is an essential indicator of the model's performance in segmentation tasks. The Validation accuracy reflects the ability of the model to generalize the acquired knowledge to new knowledge, which is vital for assessing its generalization ability and resistance to customization. According to Table 3, the proposed method demonstrates the highest validation accuracy (90%), indicating its ability to classify pixels in new images effectively. It also achieves high accuracy on the training set (95%), indicating good model convergence. DeepLab v3 (88%) and U-Net (85%) show slightly lower validation accuracy results but still demonstrate relatively effective generalization. The SegNet (80%), FCN8s (65%), and ENet (70%) models have significantly lower validation accuracy values, indicating limited generalization ability and higher validation error.

In Table 4, the proposed method achieves the highest Pixel Accuracy (95.1%), which is higher than the results of all other considered models, including PSANet (94.8%), DANet (94.6%), and OCNet (92.1%). The high PA accuracy indicates the model's ability to effectively identify objects in the image, minimizing background noise classification errors and false vehicle detections.

The mIoU metric is one of the key indicators for assessing image segmentation quality. It determines the degree of correspondence between the predicted and actual object segments by calculating the ratio of their intersection area to the area of their union. A high mIoU value indicates the model's ability to accurately identify object boundaries, reducing the number of misclassified pixels to ensure high segmentation accuracy. The proposed method reached 82.3%, which is significantly higher than DeepLab v3 (74.0%), ENet (70.8%), and U-Net (73.3%). At the same time, SegNet and FCN8s have the same value (56.7%), which indicates their limited ability to separate objects accurately.

Experimental results show that the proposed method demonstrates high efficiency in vehicle segmentation in UAV images, achieving the best results in terms of PA and mIoU metrics and a low Loss value. It is a testament to its ability to classify peak villages and segment objects accurately.

The high values of PA and mIoU achieved by the proposed method can be explained by using multiscale features and transposed convolutions, which allow for effec-

DOI 10.15588/1607-3274-2025-3-8

tive detection and segmentation of objects of different sizes and shapes. The low value of Loss indicates practical model training.

A visual analysis of the results confirms the effectiveness of the proposed method in the task of semantic segmentation of vehicles. As shown in Figure 2, the proposed method provides clear and accurate detection of vehicles, which is confirmed by the quality of the binary mask and color segmented image.

In particular, vehicles are correctly identified on the binary mask without significant gaps or false positive segmentations. In addition, the segmented image demonstrates high accuracy in separating classes of objects such as roads, green spaces, and buildings. An essential factor is that the proposed method effectively distinguishes objects with similar spectral characteristics, which is often a problem for traditional approaches.

Compared to existing models, the proposed method demonstrates better preservation of object contours and minimization of noise in segmentation. It is essential for applications requiring a high level of detail in the results, such as traffic monitoring, parking zone analysis, and urban planning.

The experimental results show that the proposed method demonstrates high efficiency in vehicle segmentation in UAV images, achieving the best results in terms of PA and mIoU metrics and a low value of Loss. It reflects its ability to classify peak villages and segment objects accurately.

The high values of PA and mIoU achieved by the proposed method can be explained by using multiscale features and transposed convolutions, which allow for effective detection and segmentation of objects of different sizes and shapes. The low value of Loss indicates the model's practical training.

CONCLUSIONS

The paper proposes a neural network approach to semantic segmentation of vehicles in ultra-high spatial resolution images. Using the DeepLab + ResNet architecture with multiscale feature extraction and a loss function based on the Dice coefficient allows for achieving high accuracy of vehicle segmentation, particularly in the context of multi-class segmentation, where it is essential to solve the problem of class imbalance effectively.

Experimental studies have shown that the proposed method achieves high segmentation accuracy, outperforming the results of other well-known architectures such as U-Net, SegNet, FCN8s, and ENet. In particular, the analysis of the Loss metric showed that the proposed method demonstrates the lowest values at the training and validation stages, which indicates the stability of the training process and the efficient generalization of the model. Similarly, the Accuracy validation results confirmed the proposed method's high efficiency, which reached 95% accuracy on the training set and 90% on the validation set, which exceeds the results of other models. It indicates the effectiveness of pre-training on specialized datasets, adaptation of the loss function, and application of multiscale feature extraction mechanisms.

The scientific novelty of the results is that a neural network approach was proposed for the semantic segmentation of vehicles in ultra-high spatial resolution images. This approach is based on the DeepLab + ResNet architecture with multi-level feature extraction. The use of retraining on specialized datasets, adaptation of the loss function, and the application of mechanisms for multiscale feature extraction and concatenation (feature fusion) allows for achieving significantly higher accuracy and efficiency compared to other known models such as U-Net, SegNet, FCN8s, and ENet. The method also considers the problem of class imbalance in multi-object segmentation, for which a customized loss function based on the Dice coefficient was proposed, which increases the efficiency of recognizing classes with low representation.

The practical significance of the proposed approach lies in its ability to provide accurate and efficient vehicle segmentation in UAV images for real-time traffic monitoring, congestion detection, and emergencies.

Prospects for further research include optimizing the neural network architecture, expanding the dataset, using additional data sources, developing methods for real-time operation, adapting to different lighting and weather conditions, segmenting video streams, and 3D segmentation. These research areas will improve the accuracy and efficiency of vehicle segmentation in ultra-high spatial resolution images and expand the possibilities of its application in various industries.

REFERENCES

1. Yongtao Yu., Tiannan Gu., Haiyan G., Dilong Li, Shenghua J. Vehicle detection from high-resolution remote sensing imagery using convolutional capsule networks, *IEEE Geosci. Remote Sens. Lett.*, 2019, Vol. 16, No. 12, pp. 1894–1898. DOI:10.1109/LGRS.2019.2912582.
2. Byun S., Shin I.-K., Moon J., Kang J., Choi S.-I. Road traffic monitoring from UAV images using deep learning networks, *Remote Sens*, 2021, No. 13, P. 4027. DOI: 10.3390/rs13204027.
3. Khrissi L., El Akkad N., Satori H., Satori K. Clustering method and sine cosine algorithm for image segmentation, *Evol. Intell.*, 2022, No. 15, pp. 669–682. DOI: 10.1007/s12065-020-00544-z.
4. Osco L., Junior J., Ramos A., de Castro Jorge L., Fatholahi S., de Andrade Silva J., Matsubara E., Pistori H., Gonçalves W., Li J. A review on deep learning in UAV remote sensing, *International Journal of Applied Earth Observation and Geoinformation*, 2021. DOI:102:102456.
5. Kemker R., Salvaggio C., Kanan C. Algorithms for semantic segmentation of multispectral remote sensing imagery using deep learning, *ISPRS Journal of Photogrammetry and Remote Sensing*, 2018, Vol. 145. DOI:10.1016/j.isprsjprs.2018.04.014.
6. Hong D., Danfeng H., Gao L., Yokoya N., Jing Y., Chanussot J., Du Q., Zhang B. More diverse means better: multimodal deep learning meets remote-sensing imagery classification, *IEEE Transactions on Geoscience and Remote Sensing*, 2020, Vol. 59 (5), pp. 4340–4354. DOI:10.1109/TGRS.2020.3016820.
7. Binge C., Chen X., Lu Y. Semantic segmentation of remote sensing images using transfer learning and deep convolutional neural network with dense connection, *IEEE Access,*

DOI 10.15588/1607-3274-2025-3-8

2020, No. 8, pp. 116744–116755. DOI:10.1109/Access.6287639.
8. Singh C., Mishra V., Jain K., Shukla A. FRCNN-based reinforcement learning for real-time vehicle detection, tracking and geolocation from UAS, *Drones*, 2022, No. 6, P. 406. DOI: 10.3390/drones6120406.
9. Saqib M., Khan S., Sharma N., Blumenstein M. A study on detecting drones using deep convolutional neural networks, *In Proceedings of the 2017 14th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS),* 2017, pp. 1–5. DOI: 10.1109/AVSS.2017.8078541.
10. Kong X., Zhang Y., Tu S., Xu C., Yang W. Vehicle detection in high-resolution aerial images with parallel RPN and density-assigner, *Remote Sens*, 2023, No. 15, P. 1659. DOI: 10.3390/rs15061659.
11. Hordiiuk D., Oliinyk I., Hnatushenko V., Maksymov K. Semantic segmentation for ships detection from satellite imagery, *2019 IEEE 39th International Conference on Electronics and Nanotechnology (ELNANO),* 2019. DOI:10.1109/elnano.2019.8783822.
12. Borovyk D., Fedoniuk R., Oliinyk A., Subbotin S., Kolpakova T. Detection of vehicles in aerial photographs using convolutional neural network, *Smartindustry*, 2024. https://ceur-ws.org/Vol-3699/paper12.pdf
13. Bochkovskiy A., Wang C., Liao H. YOLOv4: optimal speed and accuracy of object detection, *arXiv 2020*, arXiv:2004.10934, 2020. DOI: 10.48550/ARXIV.2004.10934.
14. Radionov Y., Kashtan V., Hnatushenko V., Kazymyrenko O. Aircraft detection with deep neural networks and contour-based methods, *Radio Electronics, Computer Science, Control*, 2024, №4(71), pp. 121–129. DOI:10.15588/1607-3274-2024-4-12.
15. Zhang Y., Guo Z., Wu J., Tian Y., Tang H., Guo X. Real-time vehicle detection based on improved YOLO v5, *Sustainability*, 2022, No. 14(19), P. 12274. DOI:10.3390/su141912274.
16. DOTA [Electronic resource]. Access mode: https://captain-whu.github.io/DOTA/index.html
17. Kashtan V., Hnatushenko V., Shedlovska Y. Processing technology of multispectral remote sensing images, *2017 IEEE International Young Scientists Forum on Applied Physics and Engineering (YSF),* 2017. DOI:10.1109/ysf.2017.8126647.
18. He K., Zhang X., Ren S., Sun J. Deep residual learning for image recognition, *in Proc. IEEE Conf. Comput. Vis. Pattern Recognit,* 2016, pp. 770–778. DOI: 10.48550/arXiv.1512.03385.
19. Chengzhi Y., Hongjun G. A Method of Image Semantic Segmentation Based on PSPNet, *Mathematical Problems in Engineering*, 2022, pp. 1–9. DOI:10.1155/2022/8958154.

Received 03.04.2025.
Accepted 23.06.2025.

УДК 004.8

НЕЙРОМЕРЕЖЕВИЙ ПІДХІД ДО СЕМАНТИЧНОЇ СЕГМЕНТАЦІЇ ТРАНСПОРТНИХ ЗАСОБІВ НА ЗОБРАЖЕННЯХ НАДВИСОКОГО ПРОСТОРОВОГО РОЗРІЗНЕННЯ

Каштан В. Ю. – канд. техн. наук, доцент, доцент кафедри інформаційних технологій та комп’ютерної інженерії Національного технічного університету «Дніпровська політехніка», Дніпро, Україна.

Гнатушенко В. В. – д-р техн. наук, професор, завідувач кафедри інформаційних технологій та комп’ютерної інженерії, Національного технічного університету «Дніпровська політехніка», Дніпро, Україна.

Удовик І. М. – канд. техн. наук, доцент, деканка факультету інформаційних технологій Національного технічного університету «Дніпровська політехніка» , Дніпро, Україна.

Казимиренко О. В. – аспірант кафедри інформаційних технологій та комп’ютерної інженерії Національного технічного університету «Дніпровська політехніка», Дніпро, Україна.

Радіонов Є. Д. – аспірант кафедри інформаційних технологій та комп’ютерної інженерії Національного технічного університету «Дніпровська політехніка», Дніпро, Україна.

АНОТАЦІЯ

Актуальність. Семантична сегментація транспортних засобів на аерокосмічних зображеннях надвисокого просторового розрізнення є важливим завданням для розвитку інтелектуальних транспортних систем, дозволяє автоматизувати управління дорожнім рухом у реальному часі, виявляти затори та аварійні ситуації.

Мета роботи – розробка та оцінка ефективності нейромережевого підходу для сегментації транспортних засобів на аерокосмічних зображеннях надвисокого розрізнення, що забезпечує високу деталізацію та коректне відтворення границь об’єктів.

Метод. Використано архітектуру DeepLab із ResNet-101 як Backbone для збереження градієнтів і багатомасштабного аналізу ознак. Проведено навчання на даних DOTA та донавчання на спеціалізованих наборах із класами: транспортні засоби, зелені зони, будівлі, дороги. Для зменшення дисбалансу класів застосовано функцію втрат на основі коефіцієнта Dice. Це дозволяє ефективно вирішити проблему дисбалансу класів та покращити точність сегментації об’єктів різних розмірів. Використання ResNet-101 замість Xception у магістральній мережі дозволяє зберегти градієнт при збільшенні глибини мережі.

Результати. Експериментальні дослідження підтвердили ефективність запропонованого підходу, що досягає точності сегментації понад 90%, перевершуючи існуючі аналоги. Використання багатомасштабного аналізу ознак дозволяє зберігати текстурні особливості об’єктів, зменшуючи хибні класифікації. Порівняльний аналіз із методами U-Net, SegNet, FCN8s та іншими підтверджує вищу продуктивність запропонованого підходу за метриками mIoU (82.3%) та Pixel Accuracy (95.1%).

Висновки. Експерименти підтверджують ефективність запропонованого методу семантичної сегментації транспортних засобів на зображеннях надвисокого просторового розрізнення. Використання DeepLab v3+ ResNet-101 значно покращує якість сегментації транспортних засобів в урбанізованому середовищі. Високі метричні показники роблять його перспективним для застосування у задачах інфраструктурного моніторингу та планування дорожнього руху. Подальші дослідження будуть зосереджені на адаптації моделі до нових наборів даних.

КЛЮЧОВІ СЛОВА: семантична сегментація, транспортні засоби, глибокі нейронні мережі, ResNet-101, DeepLab, багатомасштабний аналіз, зображення надвисокого розрізнення.

DOI 10.15588/1607-3274-2025-3-8

ЛІТЕРАТУРА

1. Yongtao Yu. Vehicle detection from high-resolution remote sensing imagery using convolutional capsule networks / [Yu. Yongtao, Gu. Tiannan, G. Haiyan et al.] // IEEE Geosci. Remote Sens. Lett. – 2019. – Vol. 16, No. 12. – P. 1894–1898. DOI:10.1109/LGRS.2019.2912582.
2. Road traffic monitoring from UAV images using deep learning networks / [I.-K. Shin, J. Moon, J. Kang, S.-I. Choi] // Remote Sens. – 2021. – No. 13. – P. 4027. DOI: 10.3390/rs13204027.
3. Clustering method and sine cosine algorithm for image segmentation / [L. Khrissi, N. El Akkad, H. Satori, K. Satori] // Evol. Intell. – 2022. – No. 15. – P. 669–682. DOI: 10.1007/s12065-020-00544-z.
4. A review on deep learning in UAV remote sensing / [L. Osco, J. Junior, A. Ramos et al.] // International Journal of Applied Earth Observation and Geoinformation. – 2021. DOI:102:102456.
5. Kemker R. Algorithms for semantic segmentation of multispectral remote sensing imagery using deep learning / R. Kemker, C. Salvaggio, C. Kanan // ISPRS Journal of Photogrammetry and Remote Sensing. – 2018. – Vol. 145. – DOI:10.1016/j.isprsjprs.2018.04.014.
6. More Diverse Means Better: Multimodal deep learning meets remote-sensing imagery classification / [D. Hong H. Danfeng, L. Gao et al.] // IEEE Transactions on Geoscience and Remote Sensing. – 2020. – Vol. 59 (5). – P. 4340–4354. DOI:10.1109/TGRS.2020.3016820.
7. Binge C. Semantic segmentation of remote sensing images using transfer learning and deep convolutional neural network with dense connection / C. Binge, X. Chen, Y. Lu // IEEE Access. – 2020. – No. 8. – P. 116744–116755. DOI:10.1109/Access.6287639.
8. FRCNN-based reinforcement learning for real-time vehicle detection, tracking and geolocation from UAS / [C. Singh, V. Mishra, K. Jain, A. Shukla] // Drones. – 2022. – No. 6. – P. 406. – DOI: 10.3390/drones6120406.
9. A study on detecting drones using deep convolutional neural networks / [M. Saqib, S. Khan, N. Sharma, M. Blumenstein] // In Proceedings of the 2017 14th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS). – 2017. – P. 1–5. DOI: 10.1109/AVSS.2017.8078541.
10. Kong X. Vehicle detection in high-resolution aerial images with parallel RPN and density-assigner / [X. Kong, Y. Zhang, S. Tu et al.] // Remote Sens. – 2023. – No. 15. – P. 1659. DOI: 10.3390/rs15061659.
11. Semantic segmentation for ships detection from satellite imagery / [D. Hordiiuk, I. Oliinyk, V. Hnatushenko, K. Maksymov] // 2019 IEEE 39th International Conference on Electronics and Nanotechnology (ELNANO). – 2019. DOI:10.1109/elnano.2019.8783822.
12. Borovyk D. Detection of vehicles in aerial photographs using convolutional neural network / [D. Borovyk, R. Fedoniuk, A. Oliinyk et al.] // Smartindustry. – 2024. https://ceur-ws.org/Vol-3699/paper12.pdf
13. Bochkovskiy A. YOLOv4: optimal speed and accuracy of object detection / A. Bochkovskiy, C. Wang, H. Liao // arXiv 2020, arXiv:2004.10934. – 2020. DOI: 10.48550/ARXIV.2004.10934.
14. Aircraft detection with deep neural networks and contour-based methods / [Y. Radionov, V. Kashtan, V. Hnatushenko, O. Kazymyrenko] // Radio Electronics, Computer Science, Control. – 2024. – №4(71). – P. 121–129. DOI:10.15588/1607-3274-2024-4-12.
15. Zhang Y. Real-time vehicle detection based on improved YOLO v5 / [Y. Zhang, Z. Guo, J. Wu et al.] // Sustainability. – 2022. – No. 14(19). – P. 12274. DOI:10.3390/su141912274.
16. DOTA [Electronic resource]. – Access mode: https://captain-whu.github.io/DOTA/index.html
17. Kashtan V. Processing technology of multispectral remote sensing images / V. Kashtan, V. Hnatushenko, Y. Shedlovska // 2017 IEEE International Young Scientists Forum on Applied Physics and Engineering (YSF). – 2017. DOI:10.1109/ysf.2017.8126647.
18. Deep residual learning for image recognition / [K. He, X. Zhang, S. Ren, J. Sun] // in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. – 2016. – P. 770–778. DOI: 10.48550/arXiv.1512.03385.
19. Chengzhi Y. A Method of Image Semantic Segmentation Based on PSPNet / Y. Chengzhi, G. Hongjun // Mathematical Problems in Engineering. – 2022. – P. 1–9. DOI:10.1155/2022/8958154.

DOI 10.15588/1607-3274-2025-3-8