Advances in Science and Technology
Research Journal

Advances in Science and Technology Research Journal 2023, 17(6), 127–139
https://doi.org/10.12913/22998624/172771
ISSN 2299-8624, License CC-BY 4.0

Received: 2023.06.30
Accepted: 2023.09.22
Published: 2023.11.21

A Comparative Analysis of Image Segmentation Using Classical and Deep Learning Approach

Arsen Plaksyvyi[1], Maria Skublewska-Paszkowska[1*], Paweł Powroznik[1]

[1] Lublin University of Technology, Faculty of Electrical Engineering and Computer Science, Department of Computer Science, Nadbystrzycka 38D, 20-618 Lublin, Poland

* Corresponding author's e-mail: maria.paszkowska@pollub.pl

ABSTRACT

Segmentation is one of the image processing techniques, widely used in computer vision, to extract various types of information represented as objects or areas of interest. The development of neural networks has influenced image processing techniques, including creation of new ways of image segmentation. The aim of this study is to compare classical algorithms and deep learning methods in RGB image segmentation tasks. Two hypotheses were put forward: (1) The quality of segmentation applying deep learning methods is higher than using classical methods for RGB images, and (2) The increase of the RGB image resolution has positive impact on the segmentation quality. Two traditional segmentation algorithms (Thresholding and K-means) were compared with deep learning approach (U-Net, SegNet and FCN 8) to verify RGB segmentation quality. Two resolutions of images were taken into consideration: 160x240 and 320x480 pixels. Segmentation quality for each algorithm was estimated based on four parameters: Accuracy, Precision, Recall and Sorensen-Dice ratio (Dice score). In the study the Carvana dataset, containing 5,088 high-resolution images of cars, was applied. The initial set was divided into training, validation and test subsets as 60%, 20%, 20%, respectively. As a result, the best Accuracy, Dice score and Recall for images with resolution 160x240 were obtained for U-Net, achieving 99.37%, 98.56%, and 98.93%, respectively. For the same resolution the highest Precision 98.19% was obtained for FCN-8 architecture. For higher resolution, 320x480, the best mean Accuracy, Dice score, and Precision were obtained for FCN-8 network, reaching 99.55%, 99.95% and 98.85%, respectively. The highest results for classical methods were obtained for Threshold algorithm reaching 80.41% Accuracy, 58.49% Dice score, 67.32% Recall and 52.62% Precision. The results confirm both hypotheses.

Keywords: image segmentation, pattern recognition, artificial neural networks.

INTRODUCTION

The increase of computing power gave an impulse to the development of new methods of image processing. Artificial intelligence technologies, including deep learning, have become the most popular and growing in this area over the past few years [1]. Segmentation, as a method of image processing, was introduced in the 1960s., over the past decade, this technology has undergone significant changes as a result of machine development and deep learning [2]. Segmentation involves splitting images into groups of regions, also called image segments. Each of it represents a significant area of the image, thus separating objects or parts of the scene or image [3].

There are many types of image segmentation, but three are the most common: semantic [4], instance [5] and panoptic [6]. Semantic segmentation involves combining pixels based on their belonging to a semantic class. Instance segmentation introduces pixel classifications based on their belonging to an object instance. Unlike semantic segmentation, this method does not allow to determine a class, but only to select an area based on the edges of a single object. On the other hand, panoptic segmentation is a combination of the first two methods and allows to determine the area that belongs to an instance of the object and to classify the area based on the semantic suitability of the class.

Nowadays, segmentation, as well as other image processing methods, is widely used in computer vision [7-9]. This type of algorithms can be divided into several groups: threshold-based segmentation, region-based segmentation, edge-based segmentation, clustering-based segmentation, graph-based segmentation, active contour-based segmentation, and segmentation algorithms that use deep learning neural networks [10]. The first group of algorithms is the oldest and perform segmentation using a given type of algorithms involves finding similarities among pixels. A feature of the next group of algorithms is the principle of image segmentation based on the edges of individual objects. Clustering algorithms, such as the K-means algorithm, iteratively assign pixels to a certain group, creating regions composed of pixels with similar features. Deep learning image segmentation is a technique that relies on the functioning of deep neural networks. This class includes a large number of algorithms, the following are among the most popular: DeepLab - convolutional deep learning networks, Fully Convolutional Network (FCN), deep learning networks based on encoder-decoder architecture: U-Net and SegNet, Region-based Convolutional Neural Network (R-CNN) and others [11-13].

The segmentation methods and image analyses may be found in many papers. The study presented in [14] contains a detailed description of the deep convolutional neural network architecture for semantic pixel segmentation called SegNet. The authors compare the segmentation precision for different architectures for 2 datasets CamVid and SUNRGB. A comparison of the calculation time and hardware resources required for the different deep learning architectures was also analyzed, which allowed to confirm the effectiveness of the presented neural network. Another study concerns a new approach called Edge-Based Segmentation Network (ESNet) for Real-Time Semantic Semantic Segmentation in Traffic Scenes [15]. The authors describe the network and compare it with other artificial intelligence approaches such as: FCN-8, PSPNet, SegNet and DeepLab. The main goal was to demonstrate the advantages of the proposed ESNet in image segmentation in terms of computational efficiency without compromising segmentation quality. The Cityscapes dataset was chosen for this study.

The most extensive studies contain articles comparing different segmentation methods. The study described in [16] compares the effectiveness of classic segmentation algorithms with U-Net. The specificity of a given work is that it used an original dataset, which is a set of images with various configurations of geometric images. Furthermore, the given work included a single segmentation technique using a deep neural network, which did not make it possible to carry out a comparative analysis clearly and to map the results of the study to a set of photos of reality. The authors of [17] study the possible correlation between image resolution and the effectiveness of algorithms. They compared FCN and U-Net algorithms to segment road network images in two resolutions: 256x256 and 512x512. Another study presented in [18] presents a comparison of three neural network models: FCN, U-Net, and DeepLab. The authors aimed to determine the quality of segmentation and the computational efficiency of a particular method. A specific set of photographs showing a top view captured by the camera was selected for the study. The challenge for the authors was that the images contained groups or individuals taken from different angles. The authors of [19] focus on comparing two types of neural networks: convolutional neural networks (CNN) and fully convolutional neural networks (FCNN). The study was conducted in the context of classifying areas in underwater images of coral reef ecosystems into biologically relevant categories. It includes 6 patch-based CNN models (a special variant of CNN, used for image segmentation) and 4 FCNN approaches. The algorithms are compared in terms of pixel accuracy and the CNN models give better results than FCNN. In [20] the quality of segmentation methods on complex images of immunofluorescence cells is presented. It compares 5 deep learning methods and 2 classic methods in the form of the h-min based watershed algorithm and attributed relational graphs. The authors also proposed and evaluated a new strategy of adding artificial imagery to extend the training set. The effects of various factors such as image scaling, annotation quality and post-processing methods on segmentation effectiveness were also compared. Furthermore, the results were compared with those of manual image segmentation performed by experts. The study described in [21] demonstrated a different approach for comparing segmentation methods.

The authors test various modifications of the U-Net model, such as: U-Net Vgg16, U-Net InceptionResNetV2, U-Net DenseNet121 in terms of the quality of segmentation and buildings extraction from aerial photos of Chicago.

Based on the literature review it can be stated that no study has presented a comparative analysis of deep neural network and classical segmentation methods while also considering the effect of resolution changing on segmentation quality. The novelty of the study is to compare the classical segmentation algorithms with ones of deep learning approaches. An impact of image resolution for segmentation performance was also verified that is a very important aspect in this type of study. Also, the research was carried out on a set of images of cars of various shapes and colors in the RGB format, which has not been used in research so far. Therefore, the aim of the research is a comparative analysis of classical methods (K-means clustering, Threshold segmentation) and deep learning methods (U-Net, FCN, SegNet) in RGB image segmentation tasks. The study aims to confirm or reject the following hypotheses:

- H1. The quality of segmentation, understood as accuracy, precision, recall and Sorensen-Dice coefficient, applying deep learning methods is higher than using classical methods for RGB images.
- H2. The increase of the RGB image resolution has positive impact on the segmentation quality.

The Carvana dataset, used in this study, was successful dataset containing various used cars created for the purpose of selling them online. It was utilized in many research involving improving the CNN performance by adding n-sigmoid function with Squeeze-and-Excitation block [22] or improving segmentation performance by introducing Convolutional Block Attention Module (CBAM) for U-Net architecture [23]. As a response for Kaggle Carvana Image Masking Competition in 2017 [24] that dealt with extracting cars in high resolution from the background, the study was performed with U-Net and RefineNet with various pre-trained CNN architectures [25]. The process of training was accelerated as well as the network performance was improved.

MATERIAL AND METHODS

Dataset

Images from the Carvana dataset are selected for this study. The analysis will relate to the segmentation performance of the selected techniques for two different resolutions. The chosen dataset was created for the Carvana Image Masking Challenge [26] by Carvana company in 2017. This collection contains 5088 photos of cars in RGB format (Fig. 1) with a resolution of 1280x1918 pixels. 318 passenger used cars of different brands and colors were selected for the photos. They were taken using a rotating photo studio that automatically captured and processed 16 photos of the car from different angles (Fig. 2). All the photos in the set also have a corresponding binary mask, marking the area occupied by the car. In this experiment, the data set was divided into three sets: training – 3056 photos (~60%), validation – 1024 photos (~20%), and testing – 1008 photos (~20%).

Fig. 1. Images and masks from Carvana dataset [26]

Fig. 2. 360-degree vehicle photography for Carvana dataset [27]

Deep learning segmentation methods

U-Net

U-Net (Fig. 3) is a convolutional neural network architecture designed for image segmentation tasks. There are two major components to the network: encoders and decoders [28] The encoder part consists of repeated use of convolutions, followed by rectified linear unit (ReLU) and the max pooling operation which results in downsampling and increases the number of feature channels. The decoder path consists of an upsampling of the feature map, up-convolution layers followed by concatenation with the correspondingly cropped feature map, and convolutions, each followed by ReLU. The skip connections between encoder and decoder allow segmentation mask to be more accurate. This method allows to detect even very small objects in images, which is why a given architecture is commonly used to aid diagnosis, treatment planning and tracking disease progress with computer tomography image analysis.

FCN

FCN (Fig. 4) is a neural network architecture created for the solution of image segmentation tasks [29]. It is an encoder and decoder architecture, similar to U-Net. The FCN includes an encoder in the form of convolutional layers for image downsampling. The encoder usually uses the VGG16 network [30], which is designed for image recognition tasks and has 16 layers including (13 convolutional and three fully connected)

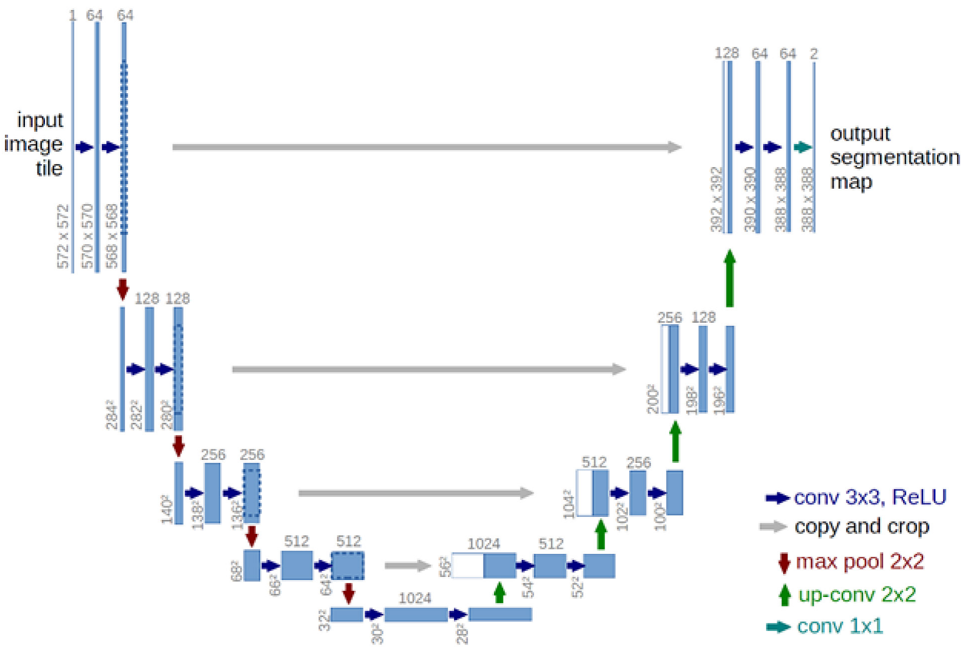

Fig. 3. U-Net architecture [28]

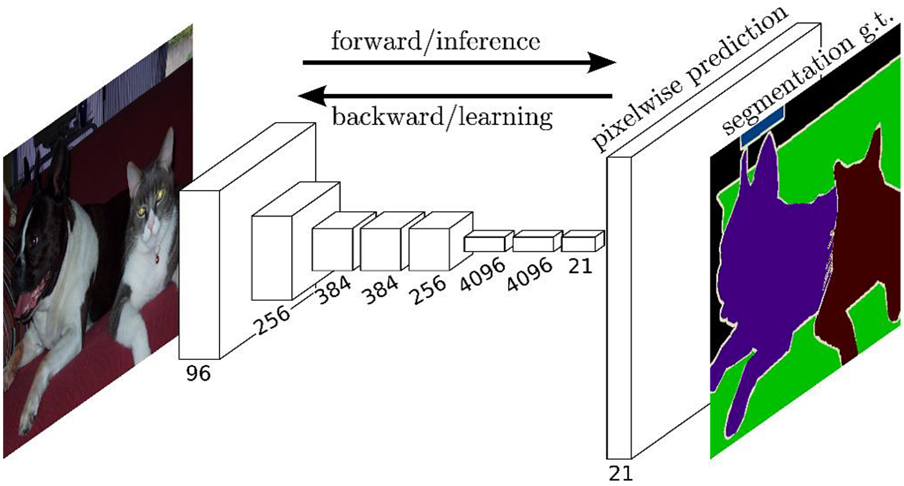

Fig. 4. FCN architecture [29]

and additional five max-pooling layers. Unlike U-Net, FCN's decoder path is not symmetrical to the encoder. The decoder contains upsampling layers that enable 32x, 16x, or 8x upsampling and allows fusing deep features from deeper convolutional layers and spatial location information from shallower layers, resulting in more precise feature output map.

SegNet

SegNet (Fig. 5) is a neural network architecture created for image segmentation [31]. The structure of this network is similar to that of U-Net and FCN networks. SegNet consists of two parts: encoder and decoder. The encoder part consists of 13 layers of convolutions corresponding to the 13 layers of the VGG16 network. Unlike FCN, SegNet does not contain three fully connected layers. Each encoder layer has a corresponding decoder layer, resulting in a symmetrical network structure. Unlike the U-Net network, SegNet does not use pooling indices, but allows the entire feature map to be sent directly to the decoder.

Classical segmentation methods

Threshold segmentation

Threshold segmentation is one of the most popular and simple image segmentation algorithms [32]. This algorithm determines new pixel values based on one or more threshold values. In binary segmentation all pixel values below the threshold value are set to black and those above to white. The specified segmentation algorithm is used for both grayscale and color images in RGB, HSF, or YUV format. In the second case, new pixel values are determined based on the values of each color channel.

K-means segmentation

The k-means algorithm is a distance-based data clustering algorithm [33]. Segmentation using the specific algorithm allows dividing an image into a defined number of groups. Subsequent pixels are assigned to each group and the average value, also called group centroid, is determined from their respective pixel values

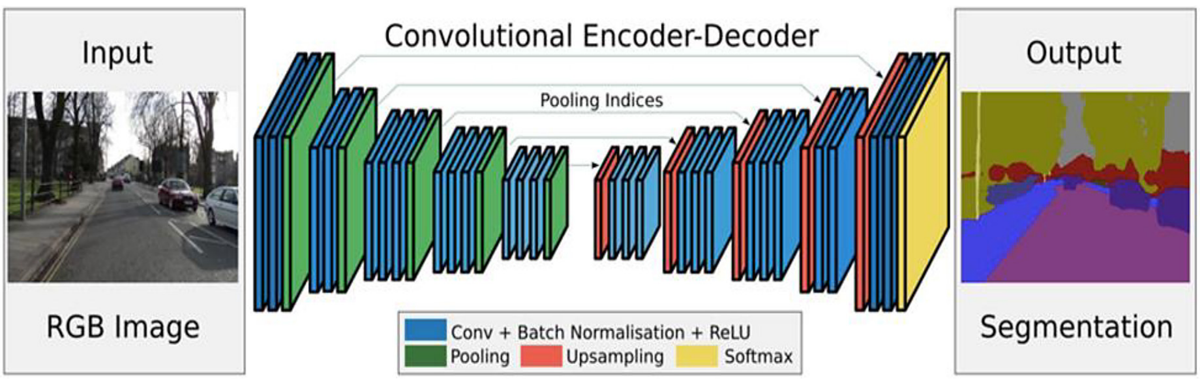

Fig. 5. SegNet architecture [31]

Table 1. Stand specification

CPU	6 cores, 12 threads
Clock speed	2.2GHz
GPU	Base Core Clock: 1290 MHz
GPU Memory	4GB
RAM	12GB
Disc	SSD 1TB

(RGB channel values, grayscale values). Pixels are added to groups based on the distance from the pixel to the centroid of each group. Therefore, the similarity to other pixels is determined by the distance.

Research methodology

Research stand

Algorithms and deep learning model accuracy tests were performed on an ASUS TUF Gaming FX504 laptop with the specifications in Table 1.

Performance measures

Two traditional segmentation algorithms (K-means and Threshold) and three deep learning algorithms (U-Net, FCN, and SegNet) were selected to check and compare the quality of segmentation as well as the stated research hypotheses. The obtained results were evaluated using four measures: Accuracy (Eq. 1), Precision (Eq. 2), Recall (Eq. 3) and the Sorensen-Dice coefficient – DSC (Eq. 4 and 5). Accuracy rate measures how often the algorithm performs image segmentation correctly. This rate is expressed as a percentage. The Sorensen-Dice score, also known as the F1 score, is a machine learning evaluation metric that determines the match between predicted segmentations and their corresponding true values.

$$Accuracy = \frac{\mathrm{TP} + \mathrm{TN}}{\mathrm{TP} + \mathrm{TN} + \mathrm{FP} + \mathrm{FN}} \tag{1}$$

where: TP – True Positive;
TN – True Negative,
FP – False Positive;
FN – False Negative.

$$Precision = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FP}} \tag{2}$$

$$Recall = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FN}} \tag{3}$$

$$DSC = \frac{2 \times |\,X \times Y|}{|\,X + Y\,|} \tag{4}$$

where: X – a set of predicted pixels;
Y – a set of reference pixels.
This metric also combines precision and recall measures:

$$DSC = \frac{2 \times (Presicion \times Recall)}{(Precision \times Recall)} \tag{5}$$

RESULTS

Segmentation quality for 160x240 resolution

Three models of neural networks have been trained on images with a resolution of 160x240 pixels in the first part of this study. Graphs of changes in loss and accuracy during the imaging training process for individual sets of images are shown in Figures 6 and 7.

To determine the quality of segmentation, four metrics were calculated: Accuracy, Precision, Recall and Dice score. The results for 3 neural network models are presented in Table 2-5, showing the minimum, maximum, mean and standard deviation values for each metric, determined on a 20-fold verification basis.

The results of the comparison of the mean values of the individual metrics calculated for the neural networks and for the conventional algorithms are gathered in Table 6.

Study of the impact of image resolution on segmentation quality

In the second part of this study three models of neural networks have been trained on images with a resolution of 320 x 480 pixels. Due to the change in image resolution, the size of the

Table 2. Accuracy for resolution 160x240 in %

Model	Minimum value	Maximum value	Mean Value	Standard deviation
U-Net	99.39518605	99.36156831	99.37869636	0.006505737
FCN-8	99.31386553	99.29425	99.30257	0.00439
SegNet	98.42643142	98.34756715	98.39049364	0.014620049

Table 3. Dice score for resolution 160x240 in %

Model	Minimum value	Maximum value	Mean Value	Standard deviation
U-Net	98.55973053	98.55955505	98.55964012	2.96021E-05
FCN-8	98.35466003	98.35455	98.35460	0.00002
SegNet	96.30705261	96.30695343	96.30699577	2.3613E-05

Table 4. Recall for resolution 160x240 in %

Model	Minimum value	Maximum value	Mean Value	Standard deviation
U-Net	98.94632578	98.91670325	98.93376738	0.006388341
FCN-8	98.56549413	98.50299	98.53718	0.01205
SegNet	98.22743008	98.12816075	98.174576	0.022404156

Table 5. Precision for resolution 160x240 in %

Model	Minimum value	Maximum value	Mean Value	Standard deviation
U-Net	98.24112535	98.10515674	98.17225146	0.034336691
FCN-8	98.22535624	98.16894	98.19469	0.01504
SegNet	94.64933175	94.35207994	94.49413146	0.055905289

Table 6. The comparison of algorithms for resolution 160x240 in %

Method	Accuracy	Dice score	Recall	Precision
U-Net	99.3787	98.5596	98.9338	98.1723
FCN-8	99.3026	98.3546	98.5372	98.1947
SegNet	98.3905	96.307	98.1746	94.4941
Threshold	75.2768	58.4872	67.3221	46.0944
K-Means	62.8176	44.8273	56.3527	30.5582

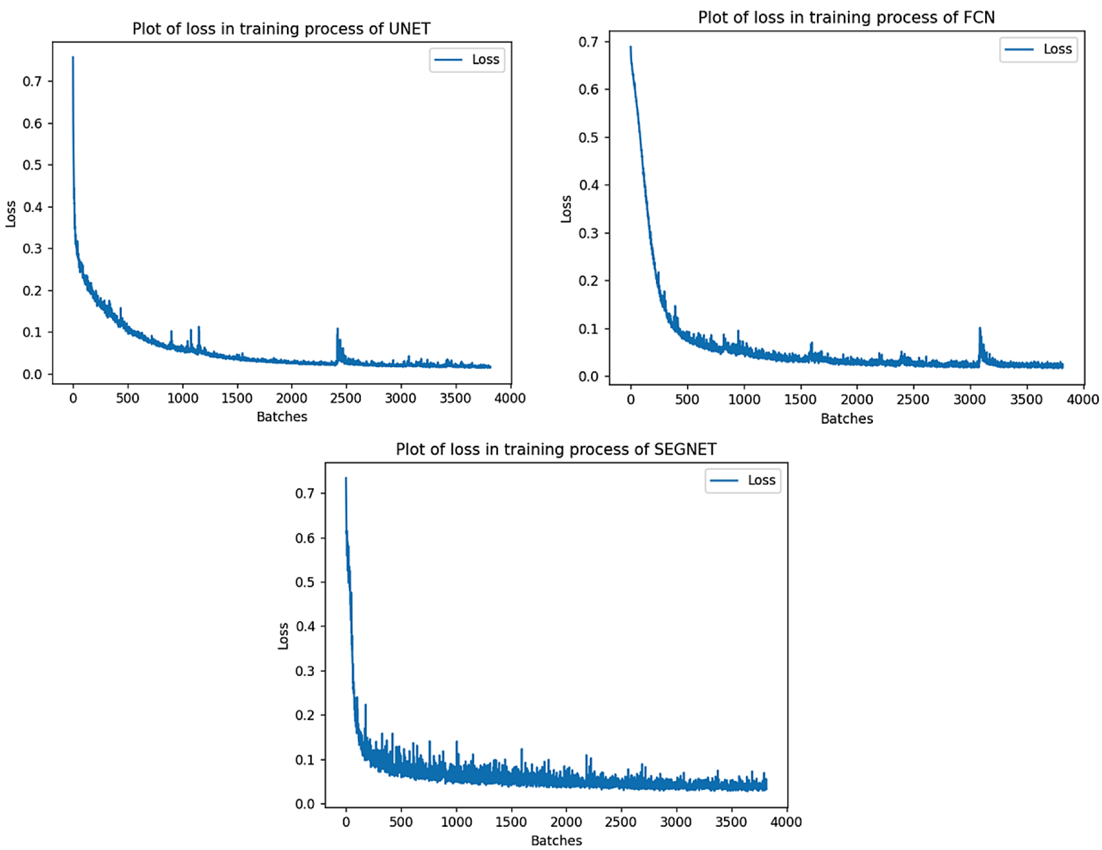

Fig. 6. Plots of loss for U-Net, FCN and SegNet models

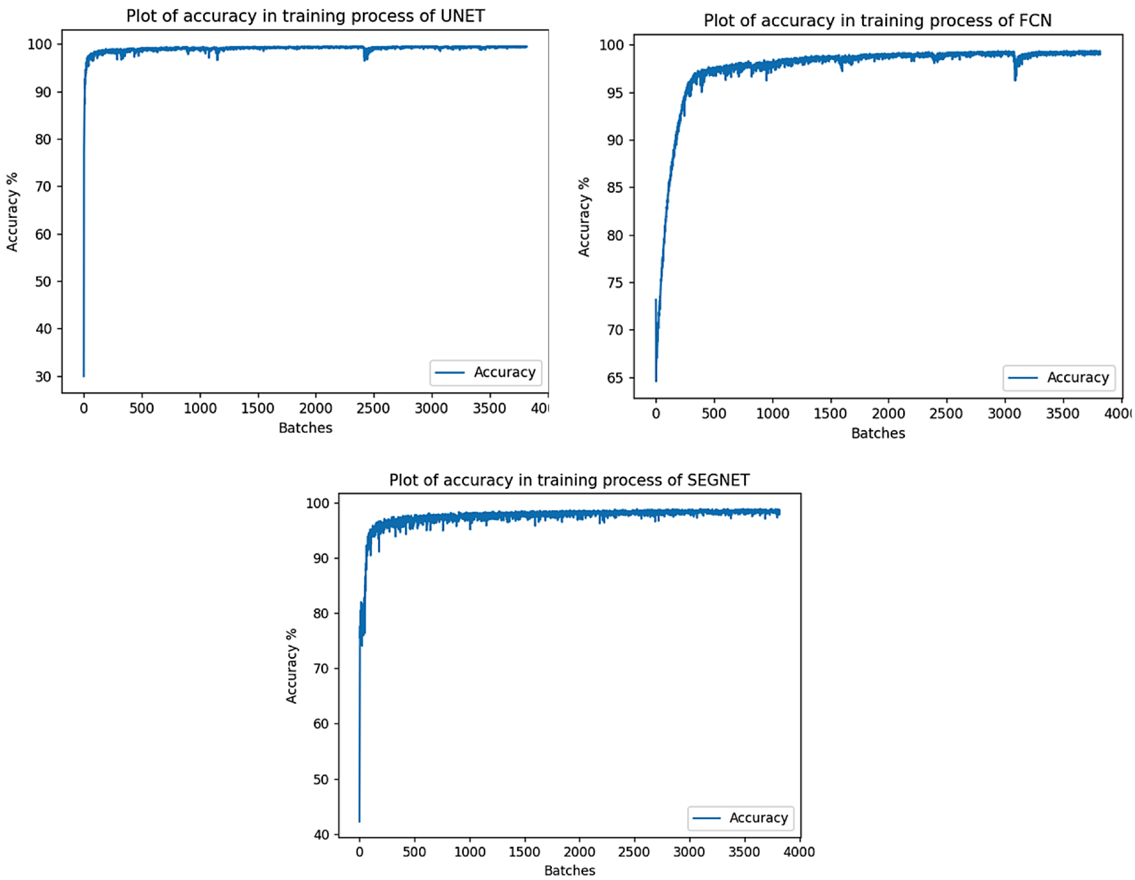

Fig. 7. Plots of accuracy for U-Net, FCN and SegNet models

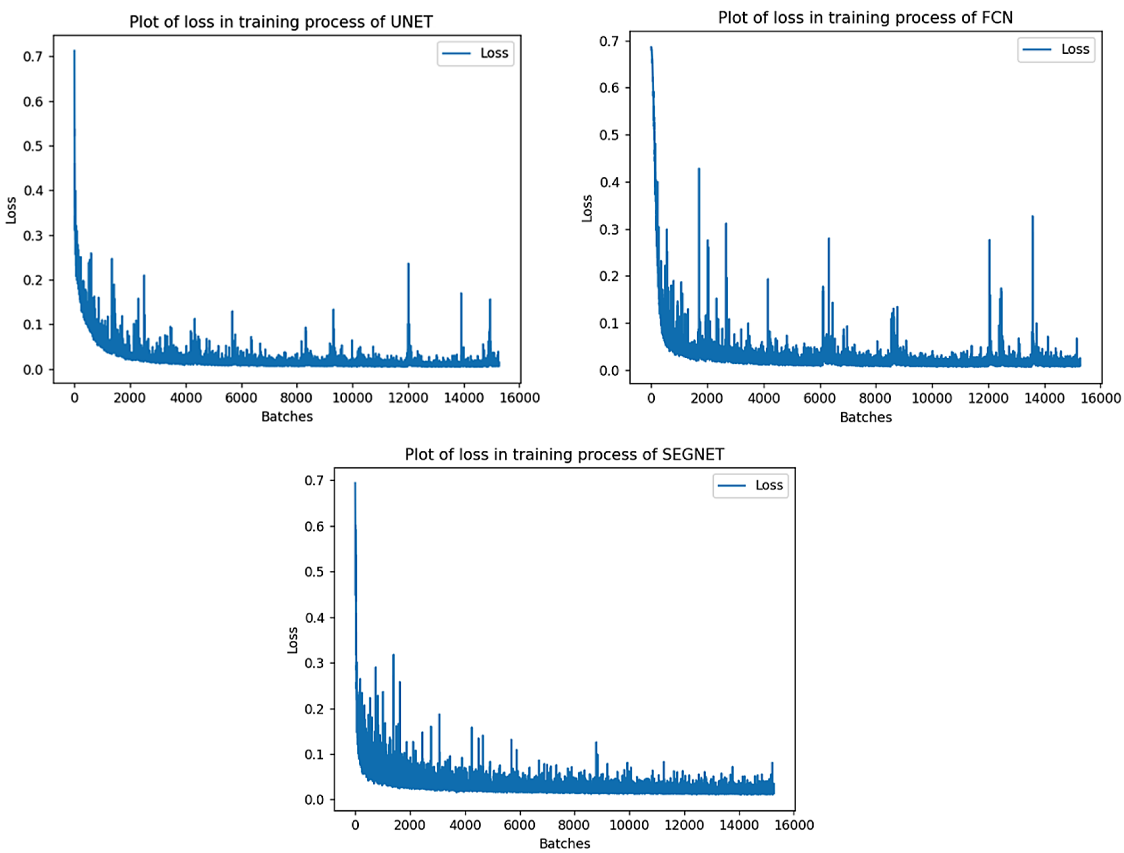

Fig. 8. Plots of loss for U-Net, FCN and SegNet models

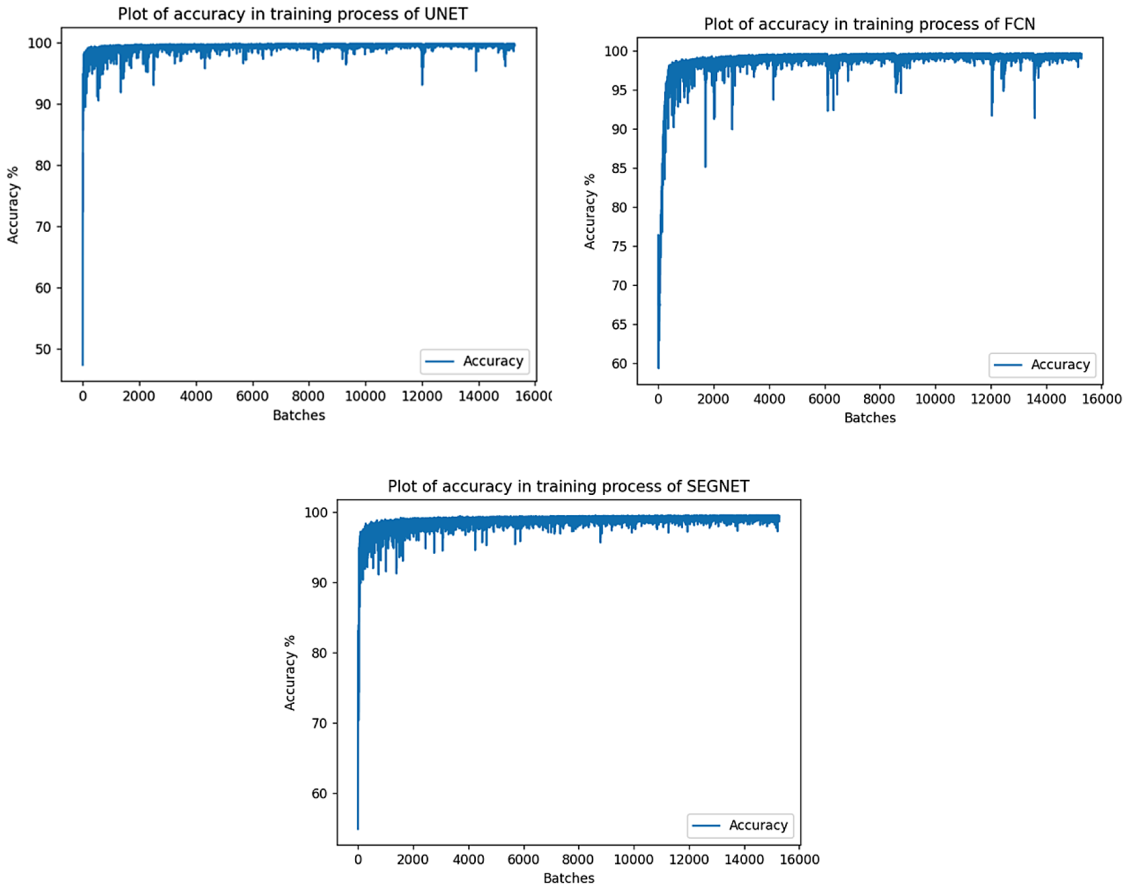

Fig. 9. Plots of accuracy for U-Net, FCN and SegNet models

Table 7. Accuracy for resolution 320x480 in %

Model	Minimum value	Maximum value	Mean Value	Standard deviation
U-Net	99.53352939	99.44035391	99.49113923	0.019060937
FCN-8	99.56414176	99.54859211	99.55717859	0.00406562
SegNet	99.24130222	99.20627384	99.22529872	0.009252595

Table 8. Dice score for resolution 320x480 in %

Model	Minimum value	Maximum value	Mean Value	Standard deviation
U-Net	98.85855103	98.84001923	98.84826775	0.003995895
FCN-8	98.96173859	98.95807648	98.95945206	0.000914993
SegNet	98.20555878	98.19795227	98.20186691	0.001835556

Table 9. Recall for resolution 320x480 in %

Model	Minimum value	Maximum value	Mean Value	Standard deviation
U-Net	99.38927243	99.34984778	99.3741179	0.009097651
FCN-8	99.09820294	99.04066392	99.070748	0.011078916
SegNet	98.85431778	98.77989702	98.81829234	0.015595463

Table 10. Precision for resolution 320x480 in %

Model	Minimum value	Maximum value	Mean Value	Standard deviation
U-Net	98.45008363	98.04797398	98.26105981	0.083711949
FCN-8	98.88313856	98.82378904	98.85661844	0.01407866
SegNet	97.63833631	97.51946833	97.5782649	0.028655288

Table 11. The comparison of algorithms for resolution 320x480 in %

Method	Accuracy	Dice score	Recall	Precision
U-Net	99.491	98.848	99.374	98.261
FCN-8	99.557	98.959	99.071	98.856
SegNet	99.225	98.202	98.818	97.578
Threshold	80.411	54.618	58.907	52.618
K-Means	78.484	54.270	55.984	49.352

Table 12. The state-of-the-art comparison

Methods	Dataset	Accuracy	Dice score	Recall	Precision	Study
U-Net with CBAM	Kaggle Carvana	NA	98.00-98.71%	NA	NA	[23]
U-Net with n-sigmoid activation function	Kaggle Carvana	99.67%	NA	NA	NA	[22]
U-Net	Kaggle Carvana	99.30%	NA	NA	NA	[22]
U-Net ResNet	Kaggle Carvana	NA	99.69	NA	NA	[25]
Inception-V1	VeRi	80.08	81.00	NA	NA	[34]
Inception-V1	VehicleID	84.40	86.90	NA	NA	[34]
VggNet-16	VeRi	80.17	80.43	NA	NA	[34]
VggNet-16	VehicleID	85.10	89.17	NA	NA	[34]
ResNet-50	VeRi	80.12	82.46	NA	NA	[34]
ResNet-50	VehicleID	84.70	87.31	NA	NA	[34]
MobileNet	VeRi	69.54	71.42	NA	NA	[34]
MobileNet	VehicleID	80.29	86.40	NA	NA	[34]
GhostNet	VeRi	80.03	80.14	NA	NA	[34]
GhostNet	VehicleID	84.65	88.38	NA	NA	[34]
MicroNet	VeRi	77.90	79.28	NA	NA	[34]
MicroNet	VehicleID	83.52	88.03	NA	NA	[34]
AlexNet	VeRi	80.51	80.09	NA	NA	[34]
AlexNet	VehicleID	87.90	91.29	NA	NA	[34]
Triplet+AlexNet	VeRi	71.23	75.31	NA	NA	[34]
Triplet+AlexNet	VehicleID	76.09	84.90	NA	NA	[34]
ASDFL	VeRi	82.08	83.23	NA	NA	[34]
ASDFL	VehicleID	88.70	92.24	NA	NA	[34]
Multi-View	VeRi	82.64	NA	NA	NA	[35]
Multi-View	VehicleID	66.06	NA	NA	NA	[35]
DFN	VeRi	88.14	NA	NA	NA	[36]
DFN	VehicleID	77.02	NA	NA	NA	[36]
BIR	VeRi	90.46	NA	NA	NA	[37]
BIR	VehicleID	77.17	NA	NA	NA	[37]
SGAT	VeRi	89.69	NA	NA	NA	[38]
SGAT	VehicleID	78.12	NA	NA	NA	[38]
U-Net 160x240	Kaggle Carvana	99.3787	98.5596	98.9338	98.1723	Own
U-Net 320x480	Kaggle Carvana	99.4911	98.8483	99.3741	98.2611	Own
FCN-8 160x240	Kaggle Carvana	99.3026	98.3546	98.5372	98.1947	Own
FCN-8 320x480	Kaggle Carvana	99.5572	98.9595	99.0707	98.8566	Own
SegNet 160x240	Kaggle Carvana	98.3905	96.307	98.1746	94.4941	Own
SegNet 320x480	Kaggle Carvana	99.2253	98.2019	98.8183	97.5783	Own
Threshold 160x240	Kaggle Carvana	75.2768	58.4872	67.3221	46.0944	Own
Threshold 320x480	Kaggle Carvana	77.1690	57.8845	59.0558	49.1739	Own
K-Means 160x240	Kaggle Carvana	62.8176	44.8273	56.3527	30.5582	Own
K-Means 320x480	Kaggle Carvana	55.8672	40.7987	49.6109	24.2774	Own

individual data batches has also been changed from 8 to 2 images. Graphs of changes in loss and accuracy during the imaging training process for individual sets of images are shown in Figures 8 and 9. In view of the fast attainment of optimal accuracy and loss values, as in the first study, a limit to 10 epochs has been set.

To determine the impact of image resolution on segmentation quality, 4 metrics were calculated: Accuracy, Precision, Recall and Dice score (Table 7-10). They were compared with the results of the first study. The results of the comparison of the mean values of the individual metrics calculated for the neural networks and for the conventional algorithms are gathered in Table 11. The results of the comparison of the mean values of the individual metrics calculated for the neural networks and for the conventional algorithms for two resolutions are shown in Table 12.

DISCUSSION

The aim of this study is a comparative analysis of classical methods (K-means clustering, Threshold segmentation) and deep learning methods (U-Net, FCN, SegNet) in RGB image segmentation tasks. Images from the Carvana dataset have been selected for this study. The analysis relates to the performance of the selected techniques for two various resolutions: 160x240 and 320x480. Based on the obtained results, the correctness of the H1 hypothesis was proven, while the H2 hypothesis was true only for deep learning networks methods. In case of classical algorithms, changing the resolution to a higher one resulted in an increase in segmentation quality for Accuracy and Dice score for both algorithms and a decrease in another measures.

The segmentation accuracy of images from the Carvana dataset was 99.38%, 99.30%, 98.39% for the U-Net, FCN-8 and SegNet models, respectively, 75.17% and 62.82% for threshold and Kmeans algorithms (Table 6), which clearly confirms the advantage of the algorithms using neural network models. Changing the resolution resulted in an increase in the accuracy metric for all analyzed models (Table 11). It means that the increasing the image resolution has impact on the improvement of the segmentation quality. However, in classical methods, for Precision and Recall lower values were obtained.

The comparison of the obtained results in this study for the deep learning approaches applied for vehicle datasets with the state-of-the-art are presented in Table 12. Three datasets were taken into consideration: Kaggle Carvana, Veri and VehicleId. On the dataset utilized in this paper, Kaggle Carvana, the U-Net and its modifications were verified [22-23, 25]. The obtained Dice score results for CBAM solution were in range of 98.00-98.71%, which are slightly lower than for analyzed U-Net and FCN-8 for 320x480 image resolution, obtained in this study. The results achieved for U-Net and FCN-8 for 160x240 image resolution are comparable with the CBAM performance. The approach of combining U-Net together with the ResNet gave the highest Dice score on the level of 99.69% [25]. The Accuracy obtained for U-Net architecture in [22] gave the comparable results to ones obtained in this study.

In [34] an ASDFL approach was compared with eight neural networks using two vehicle datasets. The obtained Accuracy and Dice score did not exceed 93%. The Multi-View approach [35], the discriminative fine-grained network (DFN) [36] and the Structured Graph Attention Network (SGAN) [38] did not reach 90%. The framework with background interference removal (BIR) mechanism [37] achieved up to 90.46%. The performance of the methods presented in [34-38] is lower according to the networks proposed in this study. They are much more adequate for segmentation purposes. The research results confirm the results of the analyzed papers [16] and [17] and clearly show the performance advantage of the neural networks over the classical algorithms in image segmentation tasks.

CONCLUSIONS

The study presented in this paper allows for a thorough analysis of the performance of the image segmentation solutions for Carvana dataset. Both classical methods and those based on deep learning approaches were taken into consideration. K-means and Threshold were applied as classical segmentation algorithms, while U-Net, FCN-8 and SegNet were chosen as deep learning solutions. The most important measures such as Accuracy, Dice score, Recall and Precision were applied in order to verify the segmentation performance. All analysed deep learning methods reached very high results for both image resolutions, 160x240 and 320x480. They stated to be more appropriate choice for image segmentation in comparison to classical algorithms. Moreover, the state-of-the-art analysis showed that the proposed deep learning approaches in this study are one the best tools for segmentation purposes of vehicles.

REFERENCES

1. Garcia-Garcia A., Orts-Escolano S., Oprea S., Villena-Martinez V., Martinez-Gonzalez P., Garcia-Rodriguez J. A survey on deep learning techniques for image and video semantic segmentation. *Applied Soft Computing* 2018; 70: 41-65.
2. Zhang, Y.-J. *Advances in image and video segmentation*, IGI Global, 2006.
3. Stockman G., Shapiro L. G. *Computer vision*, Prentice Hall PTR, 2001.
4. Li B., Shi Y., Qi Z., Chen Z. A survey on semantic segmentation. *In Proc. 2018 IEEE International Conference on Data Mining Workshops, Singapore* 2018, 1233-1240.
5. Hafiz A. M., Bhat G. M. A survey on instance segmentation: state of the art. International journal of multimedia information retrieval 2020; 9(3): 171-189.
6. Kirillov A., He K., Girshick R., Rother C., Dollár, P. Panoptic segmentation. In *Proc*. of the IEEE/CVF conference on computer vision and pattern recognition*, Long Beach, USA* 2019, 9404-9413.
7. Tilton, J. C. Image Segmentation Analysis for NASA Earth Science Applications. Capital Science 2010. 2010.
8. Aljabri M., AlGhamdi M. A review on the use of deep learning for medical images segmentation. Neurocomputing 2022.
9. Mousavirad S. J., Ebrahimpour-Komleh H. Image segmentation as an important step in image-based digital technologies in smart cities: a new nature-based approach. Information Innovation Technology in Smart Cities 2018; 75-89.
10. Gonzalez R. C., WoodsR. E., Eddins S. L. Digital Image Processing Using Matlab. Digital Image Processing Using Matlab, 2004.
11. Yu, H., Yang Z., Tan L., Wang Y., Sun W., Sun M., Tang, Y. Methods and datasets on semantic segmentation: A review. Neurocomputing 2018; 304: 82-103.
12. Liu D., Soran B., Petrie G., Shapiro L. A review of computer vision segmentation algorithms. Lecture notes 2012, 53.
13. Smołka, J. Fast watershed-based dilation. Advances in Science and Technology Research Journal, 2014; 8(23): 41-44.
14. Badrinarayanan V., Kendall A., Cipolla R. Segnet: A deep convolutional encoder-decoder architecture for image segmentation. IEEE transactions on pattern analysis and machine intelligence 2017; 39.12: 2481-2495.
15. Lyu H., Fu H., Hu X., Liu L. Esnet: Edge-based segmentation network for real-time semantic segmentation in traffic scenes. In Proc: 2019 IEEE International Conference on Image Processing Taipei, Taiwan 2019, 1855-1859.
16. Karabağ C., Verhoeven J., Miller N. R., Reyes-Aldasoro C. C. Texture segmentation: An objective comparison between five traditional algorithms and a deep-learning U-Net architecture. Applied Sciences 2019; 9(18): 3900.
17. Ozturk O., Saritürk B., Seker D. Z. Comparison of fully convolutional networks (FCN) and U-Net for road segmentation from high resolution imageries. International journal of environment and geoinformatics 2020; 7(3): 272-279.
18. Ahmed I., Ahmad M., Khan F. A., Asif M. Comparison of deep-learning-based segmentation models: Using top view person images. IEEE Access 2020; 8: 136361-136373.
19. King A., Bhandarkar S. M., Hopkinson B. M. A comparison of deep learning methods for semantic segmentation of coral reef survey images. In Proc. of the IEEE conference on computer vision and pattern recognition workshops, Salt Lake City, UT, USA 2018, 1394-1402.
20. Kromp F., Fischer L., Bozsaky E., Ambros I. M., Dörr W., Beiske, K., Ambros P. F., Hanbury A., Taschner-Mandl, S. Evaluation of deep learning architectures for complex immunofluorescence nuclear image segmentation. IEEE Transactions on Medical Imaging 2021; 40(7): 1934-1949.
21. Erdem F., Avdan U. Comparison of different U-net models for building extraction from high-resolution aerial imagery. International Journal of Environment and Geoinformatics 2020; 7(3): 221-227.
22. Mulindwa D. B., Du S. An n-Sigmoid Activation Function to Improve the Squeeze-and-Excitation for 2D and 3D Deep Networks. Electronics 2023; 12(4): 911.
23. Fang X. Research on the Application of Unet with Convolutional Block Attention Module to Semantic Segmentation Task. In: Proc. of the 2022 5th International Conference on Sensors, Signal and Image Processing, Nanjing Chin 2022, 13-16.
24. Shaler B., Gill D., Mark M., McDonald P., Cukierski W. Carvana Image Masking Challenge. Kaggle. https://kaggle.com/competitions/carvana-image-masking-challenge [access September 2023].
25. Xu J., Guo H., Kageza A., Wu S., AlQarni S. Removing Background with Semantic Segmentation Based on Ensemble Learning. In 2018 11th EAI

International Conference on Mobile Multimedia Communications: 187-197.

26. Carvana Image Masking Challenge. Retrieved May, 2023, https://www.kaggle.com/c/carvana-image-masking-challenge [access 26.05.2023]
27. Carvana Upgrades Its Own Industry-Changing Virtual Vehicle Tool Experience with New Automotive Imaging Technology. https://investors.carvana.com/news-releases/2020/08-19-2020-140123662 [access September 2023].
28. Ronneberger O., Fischer P., Brox T. U-net: Convolutional networks for biomedical image segmentation. In: Proc. 18th International Conference Medical Image Computing and Computer-Assisted Intervention–MICCAI, Munich, Germany, 2015, 234-241.
29. Long J., Shelhamer E., Darrell T. Fully convolutional networks for semantic segmentation. In Proc. of the IEEE conference on computer vision and pattern recognition, Boston, Massachusetts 2015, 3431-3440.
30. Simonyan K., Zisserman A. Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv 2014*,* 1409.1556.
31. Badrinarayanan V., Kendall A., Cipolla R. Segnet: A deep convolutional encoder-decoder architecture for image segmentation. *IEEE transactions on pattern analysis and machine intelligence* 2017; 39(12): 2481-2495.
32. Niu Z., Li H. Research and analysis of threshold segmentation algorithms in image processing. In: Proc of Journal of Physics: Conference Series, Ningbo, China 2019, 1237(2), 022122.
33. Shan P. Image segmentation method based on K-mean algorithm. EURASIP Journal on Image and Video Processing, 2018; (1): 1-9.
34. Qin R., Lv H., Zhang Y., Huangfu L., Huang S. AS-DFL: An adaptive super-pixel discriminative feature-selective learning for vehicle matching. *Expert Systems* 2023; 40(2): e13144.
35. Teng S., Zhang S., Huang Q., Sebe N. Multi-view spatial attention embedding for vehicle re-identification. IEEE Transactions on Circuits and Systems for Video Technology 2020; 31(2): 816–827.
36. Wang Q., Min W., He D., Zou S., Huang T., Zhang Y., Liu R. Discriminative fine-grained network for vehicle re-identification using two-stagere-ranking. Science China Information Sciences 2020; 63(11): 1–12.
37. Wu M., Zhang Y., Zhang T., Zhang W. Background segmentation for vehicle re-identification. In: Proc of the International Conference on Multimedia Modeling, Daejeon, South Kore 2020 88–99.
38. Zhu Y., Zha Z.-J., Zhang T., Liu J., Luo J. A structured graph attention network for vehicle reidentification. In: Proc. of the 28th ACM International Conference on Multimedia, Seattle, WA, USA 2020, 646–654.14 of 15QINET AL.