DOI: http://doi.org/10.15276/hait.02.2021.7
UDC 004.932.2 + 004.85

Deep learning technology for videoframe processing in face segmentation on mobile devices

Victoria. M. Ruvinskaya[1)]
ORCID: https://orcid.org/0000-0002-7243-5535; iolnlen@te.net.ua
Yurii Y.Timkov[1)]
ORCID: https://orcid.org/0000-0001-9401-016X; timkow11@gmail.com
[1)] Odessa National Polytechnic University, 1, Shevchenko Ave. Odesa, 65044, Ukraine

ABSTRACT

The aim of the research is to reduce the frame processing time for face segmentation on videos on mobile devices using deep learning technologies. The paper analyzes the advantages and disadvantages of existing segmentation methods, as well as their applicability to various tasks. The existing real-time realizations of face segmentation in the most popular mobile applications, which provide the functionality for adding visual effects to videos, were compared. As a result, it was determined that the classical segmentation methods do not have a suitable combination of accuracy and speed, and require manual tuning for a particular task, while the neural network-based segmentation methods determine the deep features automatically and have high accuracy with an acceptable speed. The method based on convolutional neural networks is chosen for use because, in addition to the advantages of other methods based on neural networks, it does not require such a significant amount of computing resources during its execution. A review of existing convolutional neural networks for segmentation was held, based on which the DeepLabV3+ network was chosen as having sufficiently high accuracy and being optimized for work on mobile devices. Modifications were made to the structure of the selected network to match the task of two classes segmentation and to speed up the work on devices with low performance. 8-bit quantization was applied to the values processed by the network for further acceleration. The network was adapted to the task of face segmentation by transfer learning performed on a set of face images from the COCO dataset. Based on the modified and additionally trained segmentation model, a mobile app was created to record video with real-time visual effects, which applies segmentation to separately add effects on two zones - the face (color filters, brightness adjustment, animated effects) and the background (blurring, hiding, replacement with another image). The time of frames processing in the application was tested on mobile devices with different technical characteristics. We analyzed the differences in testing results for segmentation using the obtained model and segmentation using the normalized cuts method. The comparison reveals a decrease of frame processing time on the majority of devices with a slight decrease of segmentation accuracy.

Keywords: Segmentation; Video Processing; Deep Neural Networks; Deeplabv3+

For citation: **Ruvinskaya V. M., Timkov Y. Y. Deep Learning Technology for Videoframe Processing in Face Segmentation on Mobile Devices. *Herald of Advanced Information Technology*. 2021; Vol.4 No.2: 185–194. DOI: https://doi.org/10.15276/hait.02.2021.7**

1. INTRODUCTION

In today's world there is a need for fast and distance-independent communication, and electronic methods best meet this need. But such methods cannot fully convey the amount of information received by all means of human perception during live communication. Video-based communication is the most effective in this regard, which makes it popular and creates a demand for software that provides this functionality.

One method of image and video analysis is segmentation, that is, dividing an image into segments. In the context of video communication, segmentation is used to process the video with visual effects independently for different segments, namely for the foreground and background.

The purpose of segmentation is to change the representation of an image for further analysis or processing, namely to highlight certain objects and areas [1]. Technically, segmentation is the process of assigning labels to each pixel of an image so that the pixels with common visual characteristics have the same labels. Semantic:

– segmentation of images is a type of segmentation;

– which implies division of an image into areas corresponding to objects on it with simultaneous determination of the type of these objects [2, 3].

There are certain problems in programs that implement the ability to communicate using video and use the segmentation functionality:

• running real-time segmentation with high accuracy requires a lot of computing resources;

• existing segmentation functionality is meant only for high performance hardware;

• less accurate methods are used for devices with lower performance, or the loss in speed is ignored altogether.

Program methods existing at the moment can be divided into two categories: classic and machine learning based [4]. Classic methods were developed at the end of the XX century and have been widely used for the segmentation of specific simple objects,

but have a number of drawbacks that do not allow them to be effectively applied to a lot of more complex problems. Classic method examples – watershed transformation, graph partitioning methods, thresholding [5, 6], [7]. For recognition of complex objects, such methods usually require preprocessing of the input data, and the frame processing itself has a duration unacceptable for performance in real time. Machine learning-based methods include hand-made detectors combined with flat classifiers, as well as fully connected neural networks, convolutional, and fully convolutional neural networks. Methods using hand-made detectors have low expressive capabilities and require considerable preparation, so the most effective methods are those based entirely on neural networks [8].

To solve these problems, we propose to use deep convolutional neural networks, carry out transfer learning, as well as to apply technical methods for fast frame processing, which significantly increase the speed of segmentation and at the same time slightly reduce the accuracy.

2. LITERATURE REVIEW

2.1 Critical analysis of advantages and disadvantages of the existing segmentation methods

Let's take a closer look at the advantages and disadvantages of various segmentation methods.

• *Tresholding*

The simplest segmentation algorithm. The intensity of each pixel is fed into a threshold function, the result of which determines the group to which this pixel belongs [9]. The method is fast enough, but requires fine-tuning of thresholds and works only for segments with homogeneous intensity, so it is suitable for segmentation of only some specific objects.

• *Watershed transformation.*

The algorithm works with the image as a function of two variables - the pixel coordinates. The function value is the absolute value of the gradient. If you plot the value on the graph, then in the places of the intensity difference ridges are formed, and in homogeneous regions plains appear. After finding the minima of the function, there is a process of "filling with water", which begins with the global minimum. As soon as the water level reaches the value of the next local minimum, it begins to fill with water. When the two regions begin to merge, the partition is built to prevent the regions from merging. The water will continue to rise until the regions are separated only by artificially built partitions [10]. The main disadvantage of this algorithm is the need for preprocessing for pictures with a large number of local minima.

• *Graph partitioning method*

Partitioning methods examine the similarity between neighboring pixels and try to separate them into groups that are loosely connected to each other. In these methods, the image is represented as a weighted undirected graph. Each pixel is associated with a vertex, and the weights of the edges determine the similarity of neighboring pixels. Then the graph is cut according to the criterion created to obtain clusters, and each cluster is considered an image object. Some popular algorithms in this category are normalized graph cuts, minimum cut, minimum spanning tree segmentation [11]. Graph partitioning methods are quite flexible and universal, but the large number of graph vertices in the images negatively affects their speed.

• *Hand-made detectors together with the classifiers.*

The methods allow to label manually the elements of the image, the presence of which determines the presence of an object of a certain class, and then on the basis of the found elements to determine the class using the classifier. The obvious disadvantage is the high labor intensity of labeling elements on complex objects to achieve sufficient accuracy [1].

• *Histogram methods*

Histogram-based methods use the vertices of a histogram composed by the color (or other parameter) of the image pixels to find segments. Such methods require only one pass over the image, but show low accuracy on images homogeneous by the pixel parameter being compared, and are also subject to the problem of redundant segmentation, since they do not take into account spatial information [12].

• *K-means method*

The method is used to divide the image into k clusters. During the operation of the k-means algorithm, the number of clusters and their centers are initialized, then each pixel of the image is put in the cluster with the center nearest to it. Afterwards, the center values are updated with the average values of the pixels in that cluster. The steps are repeated until the error value is less than the required value [13]. A disadvantage of the method is the dependence of the algorithm accuracy on the choice of starting parameters – the number of clusters and their initial centers [14].

• *Viola–Jones object detection framework*

The method described in [15] includes integral image representation, Haar-like features, Adaboost training [16] and cascade classifiers for object recognition and area selection. The algorithm shows

excellent results in face recognition, but is very sensitive to the tilt and angles of the face, and requires significant time for training [17, 18]

• *Fully connected neural networks*

The use of fully connected neural networks for segmentation guarantees high accuracy in both class detection and boundary search [19].

The disadvantage of such networks is that the structure of the network allows to represent the image only in one-dimension and it does not allow to take into account the vertical position of pixels relative to each other, which leads to loss of some information about the topology of the image. Because of their number of weights, fully connected networks also require significant computational costs, which makes their use for real-time video processing impractical.

• *Deep convolutional neural networks*.

The architecture of artificial neural networks aimed at efficient pattern recognition. Such networks are unidirectional, multilayered, with the alternation of convolutional layers and subsampling layers. The deep features in convolutional networks are automatically determined in the process of training. Standard methods are used for training, most often the method of backpropagation. Such networks contain a much smaller number of weights than the fully connected ones, since one kernel weight is used entirely for the whole image [20].

Table 1 provides a brief comparison of the advantages and disadvantages of the described segmentation methods.

2.2. Peculiarities of segmentation implementation in video processing program

Let's look at the functionality and features of the implementation of segmentation in some programs that provide the ability to communicate with the help of video:

– B612 is a mobile application for creating photos with filters and effects. Real-time segmentation is used to distribute effects to segments during previews through the camera.

– Snapchat – A mobile app for messaging with pinned photos and videos. Includes a camera with augmented reality elements.

– Zoom – a video conferencing program that uses segmentation to remove and replace backgrounds.

As we can see from Table 2, software producers do not prioritize the speed of operation on low-performance devices, or significantly reduce the accuracy in order to maintain the speed.

Taking into account the ratio of accuracy and speed of convolutional neural networks with low (relative to hand-made detector-based methods) development complexity, it is reasonable to use them for face segmentation on real-time video on mobile devices.

Table 1. **A brief comparison of the advantages and disadvantages of segmentation methods**

Method	Advantages	Disadvantages
Thresholding	High speed	Necessity of manual setting of values
Watershed transformation	High accuracy for small details	Necessity of image preprocessing
Graph partitioning methods	Flexibility and universality	Speed drop for complex images
Hand-made detectors in combination with classifiers	High speed	Necessity of manual setting, low expression capabilities of the detectors
Histogram methods	One-pass image processing	Lack of consideration of spatial information
K-means method	Easy implementation, high speed	Dependence of algorithm accuracy on the choice of starting parameters
Viola-Jones object detection framework	High accuracy and speed	Training requires a lot of time, sensitivity to object location
Fully connected neural networks	High accuracy	Large expenditure of computing resources
Deep convolutional neural networks	High accuracy at relatively low computational costs	A large number of network parameters that need to be configured. (The solution - transfer learning)

Source: **compiled by the authors**

Table 2. **Functionality and features of segmentation implementation in some programs**

	Video recording	Video Broadcasting	Segmentation acceleration for low-performance devices	Maintaining accuracy for low-performance devices
B612	-	+	-	+
Snapchat	+	+	+	-
Zoom	+	+	-	+

Source: **compiled by the authors**

3. PROBLEM STATEMENT

The objective of this work is to apply modern deep learning methods with optimizations for mobile devices in order to improve the speed of frame processing during face segmentation.

In order to achieve this goal, it is necessary to solve the following tasks:

• analysis of existing deep neural networks in order to choose the most suitable one as a basis for segmentation;

• modification of the network architecture and its training for adaptation to the task of segmentation of a face in the frame;

• creation of a mobile application using the obtained network and implementation of techniques that increase the segmentation speed.

4. SELECTION OF A DEEP NEURAL NETWORK SUITABLE AS A BASIS FOR SEGMENTATION

Modern models based on deep neural networks contain a significant number of parameters, so training such a model from scratch requires large computational resources and a lot of time. In this case it is appropriate to use the transfer learning approach. In this approach, an already trained model is taken, and its weights are used as initialization for another task. This approach makes sense because the tasks have the same input, and more data is available for the first task than for the second [21].

In recent years, methods based on deep neural networks containing convolutional layers have shown good results in solving the segmentation problem [22]. Convolutional layers work on the kernels basis that deal with the certain image features recognition [23]. The operation of layers of such networks involves sequential scanning of the input image through a small filter of several pixels. The first such solutions involved feature extraction using convolutional layers, and further classification using full-linked layers [24]. However, this approach was inefficient for segmentation of objects of different scales.

Unlike conventional neural networks with convolutional layers, fully convolutional networks do not contain fully connected layers. The classification is also performed by convolutional layers. This architecture allows us to work with images of any size. In 2015, U-Net [25] was developed, a fully convolutional network whose architecture consisted of areas of constricting and expanding convolutional layers, which allowed context capture and then localization, and reduced the number of operations in the calculation of convolutions. The further development of U-Net was the DenseNet architecture, in which each next convolutional layer receives feature maps as input from all previous layers at once, thus reducing the growth rate of the number of channels with each layer [26]. This architecture makes it possible to significantly reduce the size of the network, and, consequently, to reduce the processing time considerably.

A completely different way of using convolutional neural networks for object segmentation is the extended convolution method, which is used in DeepLab networks. The layers of the network with such architecture use extended convolution filters, the space between the values of which is filled with zeros. Thus, the field of view is increased while computational costs remain unchanged [27].

The latest version of the DeepLabV3+ network architecture combines the idea of using extended convolutions with the encoder-decoder approach [28]. This approach allows to get rid of the problem of decreasing accuracy when dividing into segments of the original image. As the depth of the convolutional network increases, the level of abstraction of features that are recognized also increases, but their dimensionality decreases, and then the accuracy decreases when dividing the image into segments.

The block decoder uses convolution layers to increase the dimensionality of the feature maps, restoring it to its original image size.

DeepLabV3+ demonstrates efficiency on PASCAL VOC 2012 and Cityscapes datasets, achieving 89.0 % and 82.1 % mIoU for the test set without any post-processing [27]. The architectural approaches used in the model solve the main problems of segmentation when using deep convolutional neural networks, such as reduction of

ISSN 2663-0176 (Print)
ISSN 2663-7731 (Online)

spatial dimensionality, existence of an object at different scales and reduction of local accuracy.

Taking into account that the DeepLab architecture was designed to improve performance on mobile devices, and considering the above advantages of the DeepLabV3+ network architectural features, it is reasonable to use it as the basis for segmentation in this work.

5. MODIFYING THE ARCHITECTURE OF THE SELECTED NEURAL NETWORK

The DeepLabV3+ deep convolutional neural network selected as the basis for segmentation needs to be modified to successfully perform real-time facial segmentation on mobile devices.

DeepLabV3+ is designed for semantic segmentation of objects from 21 different classes: 20 foreground object classes (humans, various animals, vehicles, and others) and one background class. This number is redundant, for the task at hand it is sufficient to distinguish only the user's face and the background, so the output module of the network must be modified to produce two classes at the output.

A necessary feature for real-time execution is fast frame processing. It is possible to increase the speed of the model by reducing the number of operations in the decoder. The developers of DeepLabV3+ had a goal to achieve better accuracy results when developing the decoder, while for practical real-time applications a significant performance gain is more important than a small decrease in accuracy. So instead of restoring the size of the segmentation mask (a set of pixels where each segment corresponds to a different color) to the original image size with a 16-fold increase in the size of the feature maps in the encoder block, we will restore the mask at half the size, increasing the size by only 8 times.

The differences between the original model and the modified model are shown in Fig. 1.

Another technique that improves the speed of the network is the use of integer values instead of floating-point values using 8-bit quantization [29]. Processors of mobile devices process integer values much faster, and, as noted in the Tensorflow specification, the loss of calculation accuracy with this approach is negligible [30].

6. TRAINING PROCESS AND RESULTS

After transforming the model structure, it is needed to train it on images specific to video recording with a face in the frame. The DeepLabV3+ model contains more than 5 million parameters, so the cost of training it from scratch makes this method impractical. However, the similarity of the tasks to be solved allows the transfer-learning approach to be applied. The positive aspects of the approach are the possibility to use a small number of elements in the training sample and a significant acceleration of adaptation to a particular task. The main disadvantage is the lower accuracy of the obtained model compared to the basic model due to the fact that a small number of examples is available for training for a more specific task. But this accuracy is higher than if we build and train the model from scratch on a small number of initial examples.

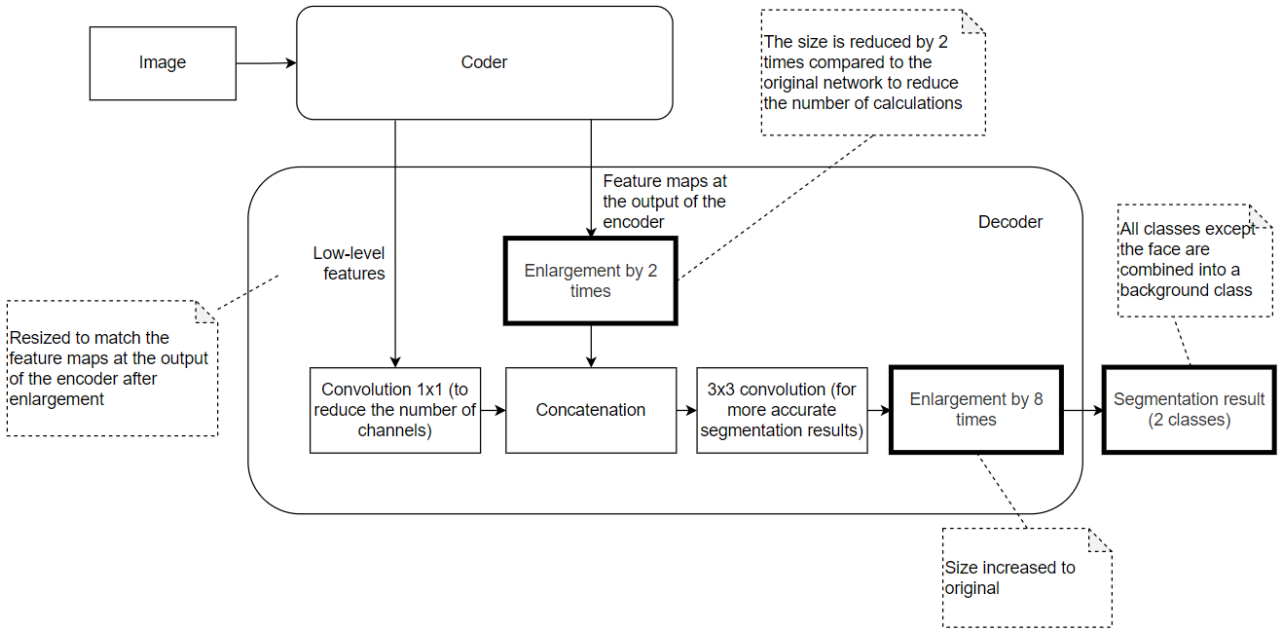


Fig. 1. **The scheme of optimized DeepLabV3+**
Source: **compiled by the authors**

To form a data for the additional training of the neural network a subset of COCO (common objects in context) dataset images was made [31].

The set consists of 5000 image pairs (source image / expected segmentation result), which are divided into training, validation and test sets in a ratio 0.7:0.15:0.15. The images are separated into face and background segments, the remaining segments in the dataset images, except for the face, are considered the background.

The metric of learning quality is mIoU (mean intersection over union) - the average ratio of the area of intersection of the obtained and expected segmentation areas to the area of their union [32].

Fig. 2 shows graphs mIoU change of the network (vertical axis) depending on the number of iterations (horizontal axis) in the process of training for training and test sets, respectively.

After training the network, the mIoU value on the test set is 72.59 %. This is less than the value of the basic DeepLabV3+ network (79.7 %), which is due to modifications aimed at improving performance. The results of frame processing time testing for different mobile devices are shown in Section 7.

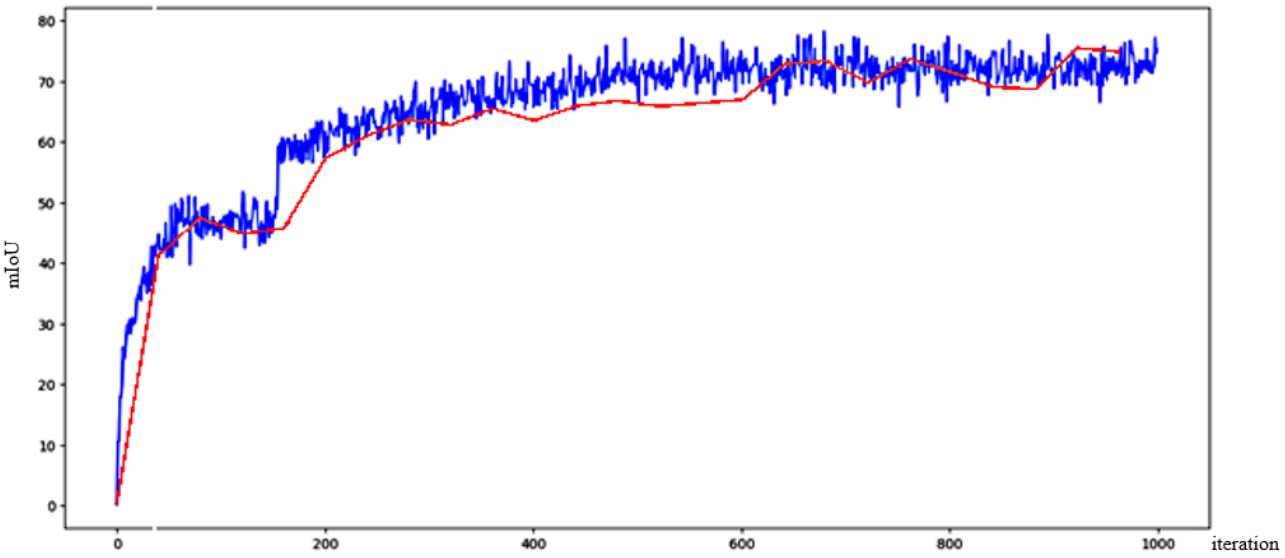


Fig. 2. **mIoU in training and test set**
Source: **compiled by the authors**

Table 3. **Characteristics of tested devices**

Name of the device	Samsung Galaxy J3	OnePlus 3T	Samsung Galaxy A8	Google Pixel 2 XL	Huawei P10
Central processing unit frequency, GHz	1,4	1,6	2,2	2,35	1,84
Number of central processing unit cores	4	4	8	8	8
Graphics processing unit frequency, MHz	700	653	800	710	1037
AI benchmark score	2,8	5,4	26,1	19,6	28,7

Source: **compiled by the authors**

ISSN 2663-0176 (Print)
ISSN 2663-7731 (Online)

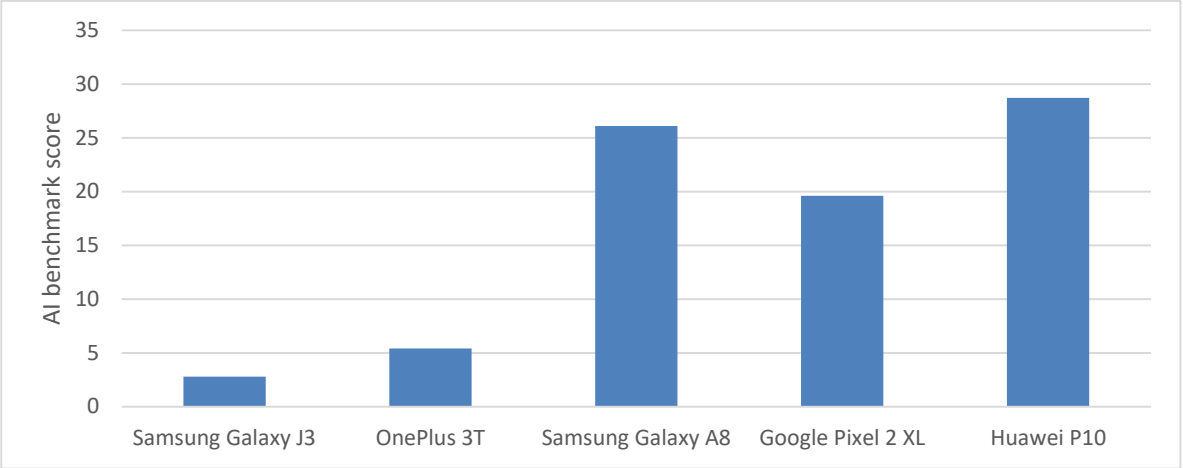


Fig. 3. **Comparison chart of AI benchmark scores for tested devices**
Source: **compiled by the authors**

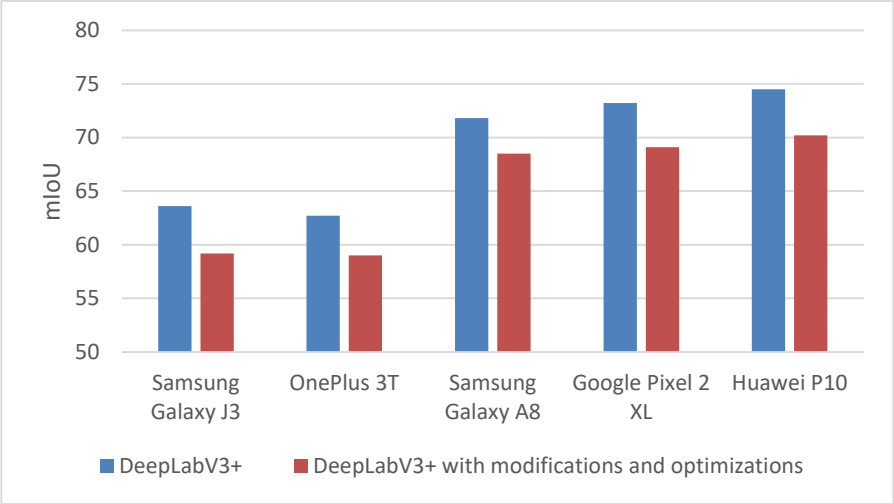


Fig. 4. **The mIoU values for testing the original and optimized versions of the network**
Source: **compiled by the authors**

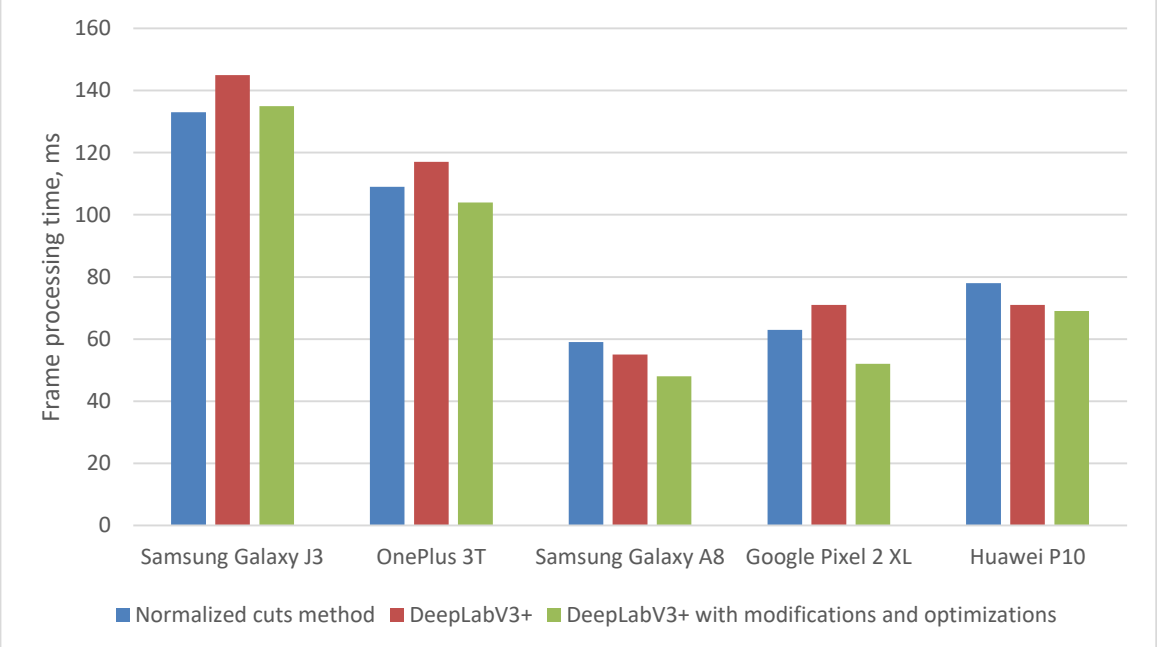


Fig. 5. **Frame processing time for testing the original and optimized versions of the network, as well as for segmentation by the normalized cuts method**
Source: **compiled by the authors**

Table 4. **Frame processing time for different devices**

Device	Samsung Galaxy J3	OnePlus 3T	Samsung Galaxy A8	Google Pixel 2 XL	Huawei P10
Normalized cuts method	133	109	59	63	78
Convolutional neural network DeepLab V3+	145	117	55	71	71
Convolutional neural network DeepLab V3+ after modification and optimization	135	104	48	52	69

Source: **compiled by the authors**

8. CONCLUSIONS

The DeepLabV3+ convolutional deep neural network is used as the basic one to solve the segmentation problem. The change of weights in the model for adaptation to the task of face segmentation is performed by the method of transfer learning. A selection of methods was made to increase the speed of segmentation without significant degradation of the accuracy. It was decided to reduce the size of feature maps as the training progresses, and to use integers instead of real numbers with the help of eight-bit quantization.

Based on a modified and trained convolutional deep neural network, a mobile application was created to impose visual effects on the image background and face.

As a result of comparison of frame processing time using the methods based on the classical segmentation method (normalized cuts) and using deep learning, it was found that the solution based on the developed deep convolutional neural network works 2-8 ms faster depending on device performance. At the same time, the value of mIoU of the neural network for segmentation decreases by less than 5% in reference to the basic network, which does not lead to noticeable visual differences in real-time segmentation on mobile devices.

Possible directions for the continuation of the work are to further improve the speed performance of the neural network by removing redundant layers, as well as the development of functionality to broadcast video to the network and third-party applications.

REFERENCES

1. Shapiro, Linda G. & Stockman, George C. “Computer Vision”. *Publ. Prentice-Hall*. New Jersey: USA. ISBN 0-13-030796-3, 2001.p. 279–325.

2. Dazhou Guo, Yanting Pei, Kang Zheng, Hongkai Yu., Yuhang Lu & Song Wang. “Degraded Image Semantic Segmentation with Dense-Gram Networks”. *IEEE Transactions on Image Processing*. 2019; Vol. 29: 782–795.

3. Tymchenko, B. I. “Neural Network Methods for Planar Image Analysis in Automated Screening Systems”. *Applied Aspects of Information Technology*. *Publ. Science і Technical*. Odesa: Ukraine. 2021; Vol. 4 No. 1: 71–79. DOI: https://doi.org/10.15276/aait.01.2021.6.

4. Amza, Catalin. “A Review on Neural Network-Based Image Segmentation Techniques”. *De Montfort University, Mechanical and Manufacturing Engg. The Gateway Leicester*. United Kingdom: 2012. p. 1–23.

5. Fu, K. S. & Mui, J. K. “A Survey on Image Segmentation”. *School of Electrical Engineering, Purdue University. Pattern Recognition*. 1980; Vol. 13: 3–16.

6. Haralick, Robert M. & Shapiro, Linda G. “Image Segmentation Techniques”. *Computer Vision, Graphics and Image Processing.* 1985; Vol. 29: 100–132.

7. Pal, Nikhil R. & Pal, Sankar K. “A Review on Image Segmentation Techniques”. *Pattern Recognition.* 1993; Vol. 26 No. 9: 1277–1294.

8. Zhong-Qiu Zhao, Peng Zheng, Shou-tao Xu & Xindong Wu. “Object Detection with Deep Learning: A Review”. *IEEE Transactions on Neural Networks and Learning Systems*. 2019; Vol. 30 Issue 11: 3212–3232.

9. Salem Saleh Al-amri1, Kalyankar, N.V. & Khamitkar S.D. “Image Segmentation by Using Thresholds Techniques”. *Journal of Computing*. 2010; Vol. 2 Issue 5: 83–86.

10. Beucher, S. & Lantuejoul, C. “Use of Watersheds in Contour Detection”. *International Workshop on Image Processing: Real-Time Edge and Motion Detection/Estimation.* Rennes. Septemper 17–21, 1979.

11. Shi, J. & Malik, J. “Normalized Cuts and Image Segmentation”. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. 2000; Vol. 22 No. 8: 888–905.

ISSN 2663-0176 (Print)
ISSN 2663-7731 (Online)

12. Giuliana Ramella & Gabriella Sanniti di Baja. “Color Histogram-Based Image Segmentation”. *Computer Analysis of Images and Patterns - 14th International Conference, CAIP 2011*. Seville: Proceedings, Part I. August 29-31, 2011.

13. Dhanachandra, N., Manglem K., & Yambem Jina Chanu. “Image Segmentation using K-means Clustering Algorithm and Subtractive Clustering Algorithm”. *Procedia Computer Science;* 2015; Vol. 54: 764–771.

14. Komleva, N. O., Liubchenko, V. V. & Zinovatna, S. L. “Methodology of Information Monitoring and Diagnostics of Objects Represented by Quantitative Estimates Based on Cluster Analysis”. *Applied Aspects of Information Technology*. *Publ. Science і Technical*. Odesa: Ukraine. 2020; Vol.3 No.1: 376–392. DOI: https://doi.org/10.15276/aait.01.2020.1.

15. Viola, Paul & Jones, Michael. “Rapid Object Detection using a Boosted Cascade of Simple Features”. *Computer Vision and Pattern Recognition*, 2001.

16. Freund, Yoav & Schapire, Robert E. “A decision-theoretic generalization of on-line learning and an application to boosting”. *Journal of Computer and System Sciences.* 1997*;* No.55: 119–139.

17. Gupta, Varsha & Sharma, Dipesh. “A Study of Various Face Detection Methods”. *International Journal of Advanced Research in Computer and Communication Engineering*. 2014; Vol. 3 Issue 5: 6694–6697.

18. Tania Akter Setu & Dr. Md. Mijanur Rahman. “Human Face Detection and Segmentation of Facial Feature Region”. *Global Journal of Computer Science and Technology: G Interdisciplinary*. 2016; Vol.16 Issue 1 Version 1.0: 1–9.

19. Bengio, Y. & Lecun, Y. “Convolutional Networks for Images. Speech and Time-Series”. *The handbook of brain theory and neural networks*. October, 1998. p. 255–258.

20. Long, Jonathan, Shelhamer, Evan & Darrell, Trevor. “Fully Convolutional Networks for Semantic Segmentation”. UC Berkley, 2015. – Available from: https://arxiv.org/pdf/1411.4038.pdf. – [Accessed 25th Dec 2021].

21. Nikolenko A., Kadurin Ye. & Arkhangel'skaya S. “Glubokoye obucheniye. Pogruzheniye v mir neyronnykh setey”. [Deep Learning. Dive into the World of Neural Networks.] *Publ. Izd*. SPb. Piter: Russian Federation. 2018, 449 p.

22. Tymchenko, B. I., Marchenko, Philip O. & Spodarets, D. V. “Segmentation of Cloud Organization Patterns from Satellite Images Using Deep Neural Networks”. *Herald of Advanced Information Technology. Publ. Science і Technical*. Odesa: Ukraine. 2020; Vol.3 No.1: 352–361. DOI: https://doi.org/10.15276/hait.03.2020.03.2020.3.

23. Sheremet, O. I., Korobov, O. Ye., Sadovoi, O. V. & Sokhina, Yu. V. “Intelligent System Based on a Convolutional Neural Network for Identifying People without Breathing Masks”. *Applied Aspects of Information Technology. Publ. Science і Technical*. Odesa: Ukraine. 2020; Vol.3 No.3: 133–134. DOI: https://doi.org/10.15276/aait.03.2020.2.

24. Sultana, F., Sufian, A. & Dutta, P. “Evolution of Image Segmentation using Deep Convolutional Neural Network: A Survey”. Dept. of Computer Science, University of Gour Banga, Dept. of Computer & System Sciences, Visva-Bharati University, 2020.

25. Ronneberger O., Fischer P. & Brox T. “U-Net: Convolutional Networks for Biomedical Image Segmentation”. Computer Science Department and BIOSS Centre for Biological Signalling Studies. University of Freiburg, Germany. 2015.

26. Gao Huang, Zhuang Liu, Laurens van der Maaten & Kilian Q. Weinberger. “Densely Connected Convolutional Networks”. Cornell University. 2018.

27. Liang-Chieh Chen, George Papandreou, Iasonas Kokkinos, Kevin Murphy & Alan L. Yuille. “DeepLab: Semantic Image Segmentation with Deep Convolutional Nets, Atrous Convolution and Fully Connected CRFs”. 2016.

28. Liang-Chieh Chen, Yukun Zhu, Papandreou, G., Schroff, F. & Adam, H. “Encoder-Decoder with Atrous Separable Convolution for Semantic Image Segmentation”. – Available from: https://arxiv.org/pdf/1802.02611.pdf. – [Accessed: 18th Dec 2021].

29. Banner, R., Hubara, I., Hoffer, E. & Soudry, D. “Scalable Methods for 8-bit Training of Neural Networks”. Israel Institute of Technology. Haifa: Israel. 2018.

30. “TensorFlow Lite 8-bit quantization specification”. – Available from: https://www.tensorflow.org/lite/performance/quantization_spec. – [Accessed 25th Dec 2021].

31. “COCO – Common Objects in Context Dataset”. – Available from: http://cocodataset.org/#explore. – [Accessed: 25th Dec 2021].

32. Hamid Rezatofighi, Nathan Tsoi, JunYoung Gwak, Amir Sadeghian, Reid, I. & Savarese, S. “Generalized Intersection over Union: A Metric and a Loss for Bounding Box Regression”. Stanford University Computer Science Department. United States: 2019.

33. Ignatov, A., Timofte, R., Kulik, A., Seungsoo Yang, Ke Wang, Baum, F., Wu, M., Lirong Xu & Luc Van Gool. “AI Benchmark: All About Deep Learning on Smartphone’s in 2019”. – Available from: https://arxiv.org/pdf/1910.06663.pdf. – [Accessed: 18th Dec 2021].

Conflicts of Interest: the authors declare no conflict of interest

Received 24.11.2020
Received after revision 14.05.2021
Accepted 14.06.2021

DOI: http://doi.org/10.15276/hait.02.2021.7
УДК 004.932.2 + 004.85

Технології глибинного навчання для обробки відеокадрів при сегментації обличчя на мобільних пристроях

Вікторія Михайлівна Рувінська[1)]
ORCID: https://orcid.org/0000-0002-7243-5535; iolnlen@te.net.ua
Юрій Юрійович Тімков[1)]
ORCID: https://orcid.org/0000-0001-9401-016X; timkow11@gmail.com
[1)] Одеський національний політехнічний університет, проспект Шевченка, 1. Одеса, 65044, Україна

АНОТАЦІЯ

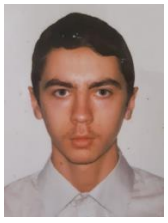
Метою дослідження є зменшення часу обробки кадрів при сегментації обличчя на відео на мобільних пристроях за допомогою технологій глибинного навчання. В роботі проведено аналіз переваг і недоліків існуючих методів сегментації, а також їх застосовності для вирішення різних завдань. Виконано порівняння існуючих реалізацій сегментації обличчя в реальному часі в найбільш популярних мобільних додатках, які надають функціонал додавання візуальних ефектів на відео. В результаті визначено, що класичні методи сегментації не володіють відповідним поєднанням точності і швидкості роботи, а також вимагають ручного налаштування під конкретну задачу, тоді як методи сегментації на базі нейронних мереж визначають глибинні ознаки автоматично і мають високу точність при прийнятному часі роботи. Для використання обрано метод на базі глибинних згорткових нейронних мереж, оскільки, крім переваг інших методів на базі нейронних мереж, він не вимагає настільки значних витрат обчислювальних ресурсів під час виконання. Проведено огляд існуючих згорткових нейронних мереж для сегментації, виходячи з якого для застосування в роботі обрано мережу DeepLabV3 + як ту, що має досить високу точність і при цьому оптимізована для роботи на мобільних пристроях. У структуру обраної мережі внесені модифікації з метою відповідності завданню сегментації на два класи і для прискорення роботи на пристроях з низькою продуктивністю. Для подальшого прискорення роботи до оброблюваних мережею значень застосована восьмібітна квантизація. Адаптація мережі під задачу сегментації обличчя виконана за допомогою перенесення навчання, проведеного на вибірці зображень з обличчями з датасета COCO. На базі зміненої і навченої моделі сегментації створено мобільний додаток для запису відео з візуальними ефектами в реальному часі, яке застосовує сегментацію для роздільного накладення ефектів на дві зони - обличчя (фільтри кольору, зміна яскравості, анімовані ефекти) і фон (розмиття, приховування, заміна на інше зображення). Проведено тестування часу обробки кадрів в додатку на мобільних пристроях з різними технічними характеристиками. Проаналізовано відмінності в показниках тестування при сегментації за допомогою отриманої моделі і з використанням сегментації методом нормального розрізу графа. В результаті порівняння виявлено зниження часу обробки кадрів на більшості пристроїв при незначному зменшенні точності сегментації.

Ключові слова: Сегментація; обробка відео; глибинні нейронні мережі; Deeplabv3+

ABOUT THE AUTHORS

Victoria M. Ruvinskaya – PhD (Eng), Professor of the Department of Sysem Software, Institute of Computer Systems. Odessa National Polytechnic University, 1, Shevchenko Ave. Odesa, 65044, Ukraine
ORCID: http://orcid.org/0000-0002-7243-5535, iolnlen@te.net.ua
Research field: Computer Science; Data Science; Deep Learning; Computer Vision

Вікторія Михайлівна Рувінська – канд. техніч. наук, проф. кафедри Системного програмного забезпечення забезпечення Інституту комп’ютерних систем Одеського національного політехнічного університету, проспект Шевченка, 1. Одеса, 65044, Україна



Yurii Yu. Timkov – Student of the System Software Department, Institute of Computer Systems. Odessa National Polytechnic University, 1, Shevchenko Ave. Odesa, 65044, Ukraine
ORCID: https://orcid.org/0000-0001-9401-016X, timkow11@gmail.com
Research field: Computer Science; Computer Vision

Юрій Юрійович Тімков – студент каф. Системного програмного забезпечення Інституту комп’ютерних систем Одеського національного політехнічного університету, проспект Шевченка, 1. Одеса, 65044, Україна