EFFICIENCY EVALUATION OF BACKGROUND REMOVAL ALGORITHMS

DOI: 10.36724/2072-8735-2022-16-7-52-58

Manuscript received 10 June 2022;
Accepted 30 June 2022

Anastasia A. Davydova,
Moscow Technical University of Communications and Informatics, Moscow, Russia, ***a.a.davydova@mtuci.ru***

Dmitry A. Egorov,
Moscow Technical University of Communications and Informatics, Moscow, Russia, ***d.a.egorov@mtuci.ru***

Keywords: *background removal, image, neural network, processing, dataset, encoder, decoder, model, metrics, object*

The necessity to delete background on an image occurs quite often. Technologies are closely connected to our everyday life so with its change they have to be adapted to the new realities. The increasing amount of video conferences during the COVID19 epidemic brought the concept of virtual backgrounds to the fore. Any videoconference services either Google Meet, Zoom or MS Teams and others have this function. Virtual background could be really useful for confidential protection or hiding one's chaotic surroundings. The technology of automated background removal must be developed, because first of all, people want to leave their personal life private, that is why they hide the place where they are. Secondly, the usage of this technology is applied for visual tracing and image segmentation. But more often background removal is used when making new content, it is possible to add any kind of background to an extracted fragment thereby getting a completely new image. Currently, there are a lot of programs with a help of which a background removal could be done, but manual correction still remains the most trusted method, yet it is a labour-intensive process with requires a lot of time. Frequently people need a quick solution for image editing and the services that provide such features are either paid services or they are not able to cope with the task, presenting the necessary level of quality. The article explores different modules of background removal and evaluates the efficiency of their work. The actuality of this problem is caused with the need of confidentiality protection with the increasing usage of services for videoconferences and also for making content, visual detection and operative problem solution. In this material we take a look at the methods used in the models, evaluate the work of the most popular models for background removal with the help of proposed new metrics for output results. All of this helps to define pluses and minuses of studied algorithms.

Information about authors:
Anastasia A. Davydova, *Student, Moscow Technical University of Communications and Informatics (MTUCI), Moscow, Russia*
Dmitry A. Egorov, *Senior Lecturer, Moscow Technical University of Communications and Informatics (MTUCI), Moscow, Russia*

Для цитирования:
Давыдова А.А., Егоров Д.А. Исследование эффективности моделей удаления фона // T-Comm: Телекоммуникации и транспорт. 2022. Том 16. №7. С. 52-58.

For citation:
Davydova A.A., Egorov D.A. (2022). Efficiency evaluation of background removal algorithms. *T-Comm*, vol. 16, no.7, pp. 52-58. (*in Russian*)

Introduction

The necessity to delete background on an image occurs quite often. Technologies are closely connected to our everyday life so with its change they have to be adapted to the new realities. The increasing amount of video conferences during the COVID19 epidemic brought the concept of virtual backgrounds to the fore. Any videoconference services either Google Meet, Zoom or MS Teams and others have this function. Virtual background could be really useful for confidential protection or hiding one's chaotic surroundings.

The technology of automated background removal must be developed, because first of all, people want to leave their personal life private, that is why they hide the place where they are. Secondly, the usage of this technology is applied for visual tracing and image segmentation. But more often background removal is used when making new content, it is possible to add any kind of background to an extracted fragment thereby getting a completely new image.

Currently, there are a lot of programs with a help of which a background removal could be done, but manual correction still remains the most trusted method, yet it is a labour-intensive process with requires a lot of time. Frequently people need a quick solution for image editing and the services that provide such features are either paid services or they are not able to cope with the task, presenting the necessary level of quality.

Neural network

Neural networks form the base of deep learning, a subfield of machine learning where the algorithms are inspired by the structure of the human brain, trying to identify the main decisions in a set of data. [1-5]. In this case, the neural networks are referred to a system of organic or artificial neurones nature.

Neural networks can adapt to an input data change; that is why network generates the best result with no need of output criteria adaptation.

Neural networks consist of three main components (layers): input layer, processing layer and output layer.

In the input layer, the received into neural network data is weighted upon different criteria. In the processing layer or how it as it is more commonly called the hidden layer there are units and connections between these units, similar to neurons and synapses in the brain of living organisms.

Currently, there are a lot of different kinds of neural networks, but basically, there are three types: artificial neural networks (ANN), convolutional neural networks (CNN), recurrent neural networks (RNN).

In this paper, only convolutional neural networks will be considered, as only this network is adapted for visual data analysis and identification, such as digital images or photos.

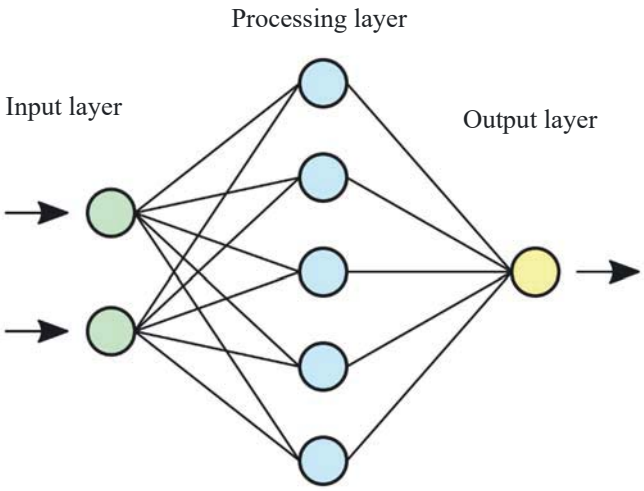

Fig. 1. Principal neural network.

Specifications of the considered models

First, consider the Background Removal model for Video Conferences by BigVision.

Unlike other developers, BigVision does not provide information about the operation and structure of their model, we only have information about the structure of the neural network on which this model is based and this is DeepLab v3 [6]. DeepLab v3 is one of the latest neural networks, on the basis of which all subsequent models are developed.

DeepLabv3 is a semantic segmentation architecture. For object segmentation on several scales, there are developed modules which use cascading or parallel convolution to capture the multiscale context by applying multiple Atrous evaluating blocks [8, 9, 10, 11].

The DeepLabV3 model has the following architecture:

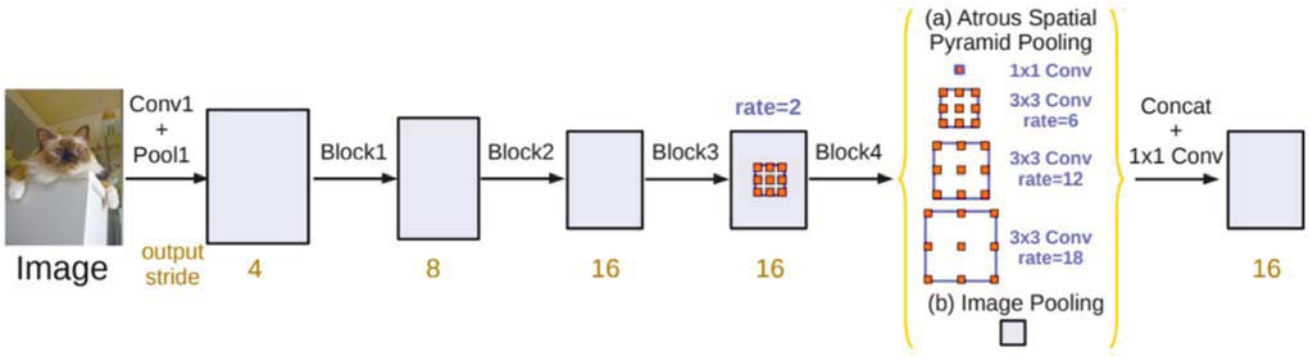

Fig. 2. DeepLabV3 model architecture

Functions are extracted from the backbone network (VGG, DenseNet, ResNet). To control the size of the feature maps, an Atrous convolution is used in the last few blocks of the backbone. In addition to the features extracted from the backbone, an ASPP network [12] is added to classify each pixel. The output of the ASPP network goes through a 1 x 1 convolution to get the actual size of the image, which will be the final segmented mask for the image.

Model DeepLabv3+ by Pinto Model Zoo

This model is realised by Encoder-Decoder neural network with Atrous separable convolution for semantic image segmentation. The module encodes multi-scale context information by applying parallel convolution at multiple scales [6], while a simple but effective decoder module refines segmentation results along object boundaries.

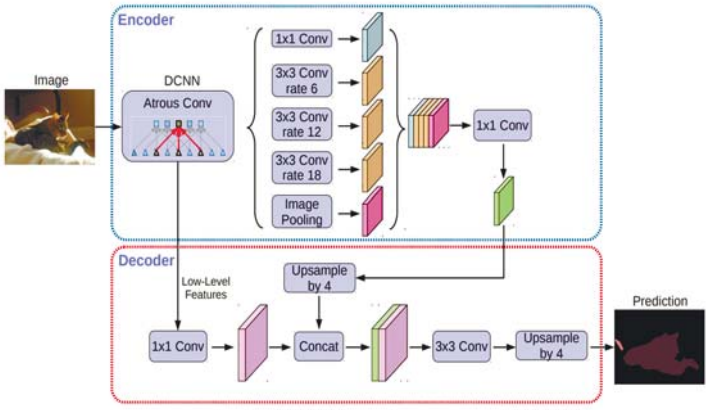

Fig. 3. DeepLabv3+ model architecture

Training dataset: DeepLabv3+ is pre-trained on the MS-COCO dataset, which includes over 330K images. Pretraining is also done on the JFT, which consists of one billion images: Similarly, the Xception model is also used, which was pretrained on ImageNet-1k containing 1.2M images and JFT-300M containing 300M images.

Evaluation dataset: This model is evaluated on the PASCAL VOC 2012 semantic segmentation test [13], which contains 20 foreground object classes and one background object class. The original dataset contains 1,456 images. This data set was further expanded, resulting in 10,582 training images [14].

Evaluation metric: To evaluate the effect of 1×1 convolution with 48 layers in the decoder module, a 3×3 convolution with 256 filters and Conv2 functions from the ResNet-101 network backbone, i.e., the last feature map in the res2x residual block, are used. After comparing Conv2 feature maps with DeepLabv3 feature maps, it turned out that it is more efficient to use two 3x3 convolutions with 256 filters than to use one or three convolutions. Changing the number of filters from 256 to 128 and the kernel size from 3x3 to 1x1 degrades performance [15, 16].

Neural network training: DeepLabv3+ uses the ResNet-101 [17] neural network trained on ImageNet-1k [18]. The model is trained directly without prior training of each component.

Conclusion on the model: The "DeepLabv3+" model uses an encoder-decoder structure where DeepLabv3 is used to encode rich contextual information, and a simple but effective decoder module is used to reconstruct object boundaries.

Background Removal by Xuebin Qin

This model uses a U-Net-like structure, specifically a two-level nested structure (U2 -Net) shown below:

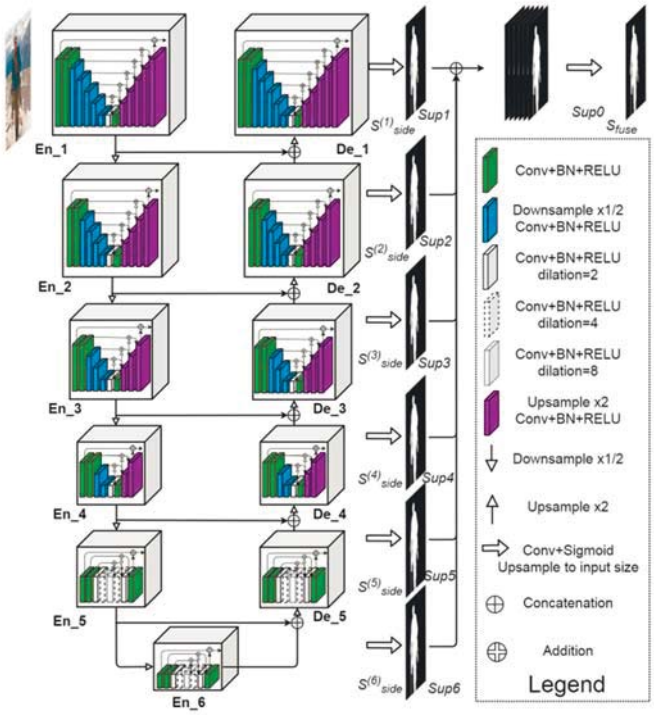

Fig. 4. Background Removal model architecture

U2-Net consists of three parts:

1. six stages encoder
2. five stages decoder
3. Module for saliency maps fusion maps with encoder and decoder steps

The neural network work description: An image is sent to the neural network. Further, in most cases, each layer of the neural networks passes data to the next layer. In this model, in a network with residual blocks, each layer transmits information to the next layer and directly to layers that are about 2-3 steps apart. This way, the design of such a neural network has a deeper architecture with rich multiscale features and relatively low computational and memory costs. In addition, since the U2-Net architecture is built only in ReSidual U-units (RSU), residual blocks, without using any pre-trained backbones adapted from image classification, it is flexible and easily adaptable to different work environments with little performance loss.

Training dataset: training of the neural network was carried out on DUTS-TR, which is part of the DUTS dataset [19]. DUTS-TR consists of 10553 images. At the moment, this is the largest and most commonly used dataset for training an object segmentation neural network. This dataset was enlarged by horizontal rotation to collect 21106 offline training images.

Evaluation dataset: evaluation of the neural network was carried out on the six most popular analyzing datasets: DUTOMRON [20], DUTS-TE [19], HKU-IS [21], ECSSD [22], PASCAL-S [23], SOD [24]. DUTOMRON includes 5168 images, most of which contain one or two structurally complex foreground objects. The DUTS dataset consists of two: DUTS-TR and DUTS-TE. As mentioned earlier, DUTS-TR was used for training.

Therefore, DUTS-TE, which contains 5019 images, is selected as one of the evaluation datasets. HKU-IS contains 4447 images with multiple foreground objects. ECSSD contains 1000 structurally complex images, many of which contain large foreground objects.

Evaluation metric: The output of the feature extraction method is a feature map that has the same spatial resolution as the input images. Each pixel of the selected object's predicted maps has a value ranging from 0 to 1 (or (0, 255)). The mask is a binary value map where each pixel is 0 or 1 (or 0 or 255), where 0 indicates background pixels and 1 indicates foreground pixels. In order to comprehensively assess the quality of these feature maps compared to the mask, six methods are used: (1) Precision-Recall curves, (2) maximal F-measure [7], (3) Mean Absolute Error [25, 28, 26], (4) weighted F-measure [25], (5) structure measure [29] and (6) relaxed F-measure of boundary [28].

Neural network training: During neural network training, each image is first resized to 320x320, randomly flipped vertically, and cropped to 288x288. The network is trained from scratch. The loss weights w (m) side and w fuse are set to 1. The Adam (adaptive learning rate) optimizer [30] is used to train the network. The network is trained until the losses converge without using the verification methods described earlier [25, 26, 31]. After 600K iterations, the training loss converges and the entire training process takes about 120 hours. During testing, the input images (H × W) are resized to 320 × 320 and transmitted to the network to obtain feature maps. The predicted 320×320 feature maps are resized to the original input image size (H×W). Both resizing processes use bilinear interpolation.

Conclusion on the model: The advantage of this design is that it is able to capture more contextual information from different scales by mixing receptive fields of different sizes in ReSidual U-boxes (RSUs).

Experiment

In order to figure out the best way to remove the background, we need to compare the outputs of the methods under the same conditions, so we create our own metric for evaluation. To do this, we make our own test data set (a photo with a person on the background and a photo of this background under the same conditions.)

Thus, we get a mask - a person without background by subtracting two images.

Then we process our test images through the chosen neural networks and thereby get the images, from which we again make masks to rate the quality, but using images with a removed background, on which we form masks, similar to the way of creating test images.

Further, we compare the masks received from photos processed using neural networks with a reference mask and select the best background removal model by getting the difference between the reference and the resulting mask. The model with the smallest error is taken as the best one.

While processing our test images through the proposed networks we had to compare only two models as the big mistake in Model DeepLabv3+ by Pinto Model Zoo was easily visually detected.

Fig. 5. Test images

Fig. 6. Reference masks

According to the Weber-Fechner law, the intensity of the sensation depends linearly on the logarithm of the intensity, so we will calculate the error in decibels, for a clear representation.

The results of comparing methods are shown in Figure 7.

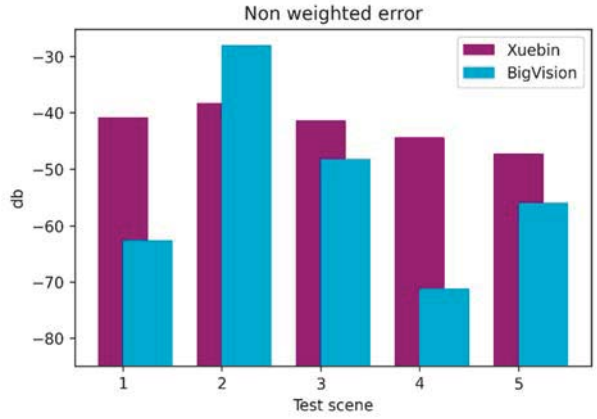

Fig. 7. Non weighted error comparison of BigVision and Xuebin Qin Model

Based on the results of the experiment, it is impossible to say unequivocally which of the models is the best, the figure 7 shows that although the Background Removal for Video Conferences by BigVision model has a minimal error on two test images, the Background Removal by Xuebin Qin model shows more stable result.

It should also be taken into account that in this experiment, studies were carried out only from a technical point of view and the aspects of human visual perception were not considered. That is why for a better understanding of the experiment results it was decided to digitally simulate the way of human reaction. People's eyes do not detect the error near the boundaries of the object, but if it is positioned far from the edge, it is more noticeable. Knowing this the boundaries of the reference masks were expanded Figure 8, by applying the Gaussian filter.

The same experiment is being held, but this time the results are weighted according to the significance of the error.

The error is calculated according these formulas:

Edge of the mask, calculated by differentiating mask image (1):

$$E(x,y) = Im(x,y)\frac{dY}{dx}\frac{dY}{dy} \quad (1)$$

Gaussian filter impulse response (2):

$$G(x,y) = \frac{1}{2\pi\sigma^2}\exp\left(-\frac{x^2+y^2}{2\sigma^2}\right) \quad (2)$$

Error weighting function, calculated by Edge of the mask and Gaussian filter impulse response convolution (3):

$$W(x,y) = E(x,y) * G(x,y) \quad (3)$$

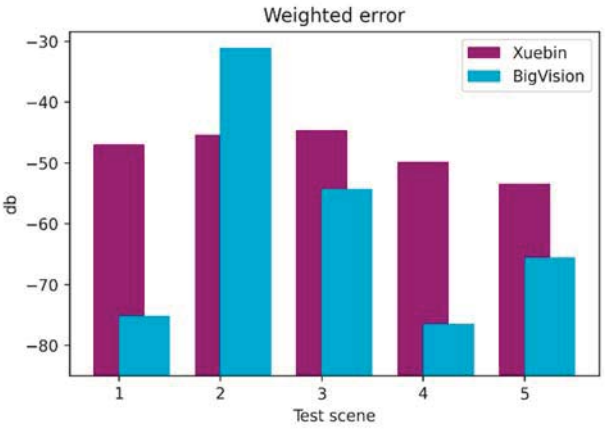

Fig. 9. Weighted error comparison of BigVision and Xuebin Qin Model

Now according to the weighted error results, the BigVision model can be considered the best.

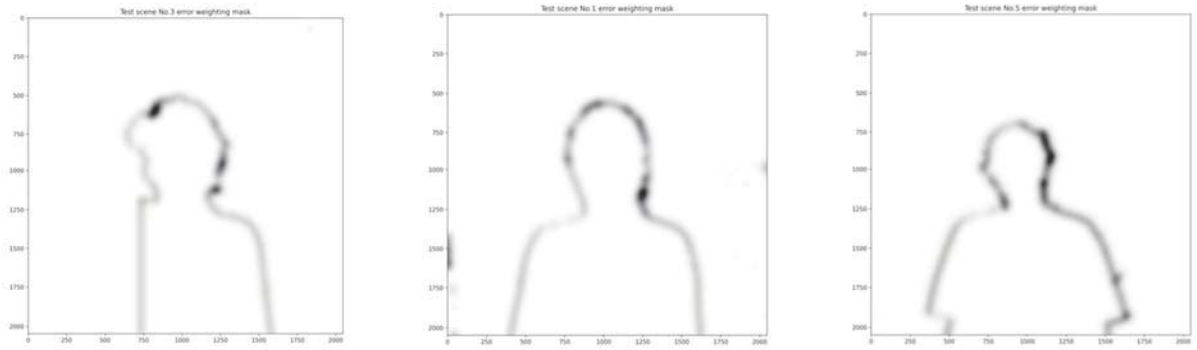

Fig. 8. Expanded reference masks

Conclusion

It was decided to continue the study further, including the use of a larger data set and a more effective metric for evaluating the quality of background removal, as well as conducting a subjective-evaluative examination.

Reference

1. F. Rosenblatt (1958), "The Perceptron: A Probabilistic Model For Information Storage And Organization in the Brain", *Psychological Review*, no. 65 (6), pp. 386-408.
2. J. Weng, N. Ahuja and T. S. Huang (1997), "Learning recognition and segmentation using the Cresceptron," *International Journal of Computer Vision*, vol. 25, no. 2, pp. 105-139, Nov. 1997.
3. H.T. Siegelmann, E.D. Sontag (1991), "Turing computability with neural nets" (PDF). Appl. Math. Lett., no. 4 (6), pp. 77-80.
4. Nicola Secomandi (2000), "Comparing neuro-dynamic programming algorithms for the vehicle routing problem with stochastic demands". Computers & Operations Research. 27 (11–12), pp. 1201-1225.
5. Goodfellow, Ian; Pouget-Abadie, Jean; Mirza, Mehdi; Xu, Bing; Warde-Farley, David; Ozair, Sherjil; Courville, Aaron; Bengio, Yoshua, (2014). Generative Adversarial Networks. *Proceedings of the International Conference on Neural Information Processing Systems (NIPS 2014)*, pp. 2672-2680.
6. L.C. Chen, G. Papandreou, F. Schroff, H. Adam (2017). Rethinking atrous convolution for semantic image segmentation. arXiv:1706.05587.
7. R. Achanta, S. Hemami, F. Estrada, and S. Susstrunk (2009). Frequency-tuned salient region detection. *In 2009 IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1597-1604.
8. A. Brandt (1977). Multi-level adaptive solutions to boundary-value problems. *Mathematics of computation*, no. 31(138), pp. 333-390.
9. D. Terzopoulos (1986). Image analysis using multigrid relaxation methods. TPAMI, (2), pp. 129-139.
10. W. L. Briggs, V. E. Henson, and S. F. McCormick (2000). A multigrid tutorial. SIAM.
11. G. Papandreou and P. Maragos (2007). Multigrid geometric active contour models. *TIP*, no. 16(1), pp. 229-240.
12. L.-C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, and A. L. Yuille (2016). Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs. arXiv:1606.00915.
13. M. Everingham, S.M.A. Eslami, L.V. Gool, C.K.I. Williams, J. Winn, A. Zisserman (2014). The pascal visual object classes challenge a retrospective. IJCV.
14. B. Hariharan, P. Arbel´aez, L. Bourdev, S. Maji, J. Malik (2011). Semantic contours from inverse detectors. *In: ICCV*.
15. O. Ronneberger, P. Fischer, T. Brox (2015). U-net: Convolutional networks for biomedical image segmentation. *In: MICCAI*.
16. V. Badrinarayanan, A. Kendall, R. Cipolla (2017). Segnet: A deep convolutional encoder-decoder architecture for image segmentation. *PAMI*.
17. K. He, X. Zhang, S. Ren, J. Sun (2016). Deep residual learning for image recognition. *In: CVPR*.
18. O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, A.C. Berg, L. Fei-Fei (2015). ImageNet Large Scale Visual Recognition Challenge. *IJCV*.
19. Lijun Wang, Huchuan Lu, Yifan Wang, Mengyang Feng, Dong Wang, Baocai Yin, and Xiang Ruan (2017). Learning to detect salient objects with image-level supervision. *In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 136-145.
20. Chuan Yang, Lihe Zhang, Huchuan Lu, Xiang Ruan, and Ming-Hsuan Yang (2013). Saliency detection via graph-based manifold ranking. *In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 3166-3173.
21. Guanbin Li and Yizhou Yu (2016). Visual saliency detection based on multiscale deep cnn features. *IEEE Transactions on Image Processing*, no. 25(11), pp. 5012-5024.
22. Yu Zeng, Yunzhi Zhuge, Huchuan Lu, Lihe Zhang, Mingyang Qian, and Yizhou Yu (2019). Multi-source weak supervision for saliency detection. *In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 6074-6083.
23. Yin Li, Xiaodi Hou, Christof Koch, James M Rehg, and Alan L Yuille (2014). The secrets of salient object segmentation. *In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 280-287.
24. Vida Movahedi and James H Elder (2010). Design and perceptual validation of performance measures for salient object segmentation. *In 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition-Workshops*, pp. 49-56.
25. Jiang-Jiang Liu, Qibin Hou, Ming-Ming Cheng, Jiashi Feng, and Jianmin Jiang (2019). A simple pooling-based design for realtime salient object detection. *In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 3917-3926.
26. Nian Liu, Junwei Han, and Ming-Hsuan Yang (2018). Picanet: Learning pixel-wise contextual attention for saliency detection. *In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 3089-3098.
27. D.A. Egorov (2019). Detection of anomalous behavior of a group of objects in a video sequence. Proceedings of the xiii international scientific and technical conference. *In the collection: technologies of the information society. Materials of the XIII International Industry Scientific and Technical Conference*, pp. 167-169.
28. Xuebin Qin, Zichen Zhang, Chenyang Huang, Chao Gao, Masood Dehghan, and Martin Jagersand (2019). Basnet: Boundaryaware salient object detection. *In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 7479-7489.
29. Deng-Ping Fan, Ming-Ming Cheng, Yun Liu, Tao Li, and Ali Borji (2017). Structure-measure: A new way to evaluate foreground maps. *In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 4548-4557.
30. Diederik P Kingma and Jimmy Ba (2014). Adam: A method for stochastic optimization. arXiv preprint.
31. Lu Zhang, Ju Dai, Huchuan Lu, You He, and Gang Wang (2018). A bi-directional message passing model for salient object detection. *In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1741-1750.
32. I. V. Vlasyuk (2019), "A method for evaluating the effectiveness of video signal processing algorithms in television cameras when transmitting moving objects," *in the collection: technologies of the information society. materials of the xiii international industry scientific and technical conference*, pp. 158-160.

ИССЛЕДОВАНИЕ ЭФФЕКТИВНОСТИ МОДЕЛЕЙ УДАЛЕНИЯ ФОНА

Давыдова Анастасия Анатольевна, *МТУСИ, Москва, Россия,* ***a.a.davydova@mtuci.ru***
Егоров Дмитрий Аркадьевич, *МТУСИ, Москва, Россия,* ***d.a.egorov@mtuci.ru***

Аннотация
В статье исследуются различные модели для "удаления" фона, сравнивается эффективность работы. Актуальность данной проблемы вызвана прежде всего необходимостью защиты конфиденциальности в связи с увеличением использования сервисов видеосвязи, а также для создания контента, визуального отслеживания и оперативного решения задач. В материале сравнивается работа наиболее популярных моделей, создается собственная метрика для оценки выходных данных, выявляется лучшая модель удаления заднего фона.

Ключевые слова: *удаление фона, изображение, нейросеть, обработка, датасет, энкодер, декодер, модель, метрика, объект.*

Литература

1. *F. Rosenblatt*. The Perceptron: A Probabilistic Model For Information Storage And Organization in the Brain // Psychological Review, №65 (6). С. 386-408, 1958.
2. *J. Weng, N. Ahuja and T. S. Huang*. Learning recognition and segmentation using the Cresceptron // International Journal of Computer Vision, vol. 25, no. 2, pp. 105-139, Nov. 1997.
3. *H.T. Siegelmann, E.D. Sontag*. Turing computability with neural nets (PDF) // Appl. Math. Lett. №4 (6). С. 77-80, 1991.
4. *Nicola Secomandi.* Comparing neuro-dynamic programming algorithms for the vehicle routing problem with stochastic demands // Computers & Operations Research. № 27 (11-12). С. 1201-1225. 2000)
5. *Goodfellow, Ian; Pouget-Abadie, Jean; Mirza, Mehdi; Xu, Bing; Warde-Farley, David; Ozair, Sherjil; Courville, Aaron; Bengio, Yoshua*. Generative Adversarial Networks // Proceedings of the International Conference on Neural Information Processing Systems (NIPS 2014). С. 2672-2680.
6. *Chen, L.C., Papandreou, G., Schroff, F., Adam, H*. Rethinking atrous convolution for semantic image segmentation. arXiv:1706.05587. 2017.
7. *R. Achanta, S. Hemami, F. Estrada, and S. Susstrunk*. Frequency-tuned salient region detection. *In 2009 IEEE Conference on Computer Vision and Pattern Recognition*. С. 1597-1604, 2009.
8. *A. Brandt*. Multi-level adaptive solutions to boundary-value problems. Mathematics of computation. №31(138). С. 333-390, 1977.
9. *D. Terzopoulos*. Image analysis using multigrid relaxation methods. TPAMI. № (2). С.129-139, 1986.
10. *W. L. Briggs, V. E. Henson, and S. F. McCormick*. A multigrid tutorial. SIAM, 2000.
11. *G. Papandreou and P. Maragos*. Multigrid geometric active contour models. TIP. № 16(1). С. 229-240, 2007.
12. *L.-C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, and A. L. Yuille*. Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs. arXiv:1606.00915, 2016.
13. *Everingham M., Eslami S.M.A., Gool L.V., Williams C.K.I., Winn J., Zisserman A*. The pascal visual object classes challenge a retrospective. IJCV, 2014.
14. *Hariharan B., Arbelaez P., Bourdev L., Maji, S. Malik, J*. Semantic contours from inverse detectors. In: ICCV, 2011.
15. *Ronneberger O., Fischer P., Brox T*. U-net: Convolutional networks for biomedical image segmentation. In: MICCAI, 2015.
16. *Badrinarayanan, V., Kendall, A., Cipolla, R*. Segnet: A deep convolutional encoder-decoder architecture for image segmentation. PAMI, 2017.
17. *He, K., Zhang, X., Ren, S., Sun, J*. Deep residual learning for image recognition. In: CVPR, 2016.
18. *Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., Huang, Z., Karpathy, A., Khosla, A., Bernstein, M., Berg, A.C., Fei-Fei, L*. ImageNet Large Scale Visual Recognition Challenge. IJCV, 2015.
19. *Lijun Wang, Huchuan Lu, Yifan Wang, Mengyang Feng, Dong Wang, Baocai Yin, and Xiang Ruan*. Learning to detect salient objects with image-level supervision // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. С. 136-145, 2017.
20. *Chuan Yang, Lihe Zhang, Huchuan Lu, Xiang Ruan, and Ming-Hsuan Yang*. Saliency detection via graph-based manifold ranking // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. С. 3166-3173, 2013.
21. *Guanbin Li and Yizhou Yu*. Visual saliency detection based on multiscale deep cnn features // IEEE Transactions on Image Processing. № 25(11). С. 5012-5024, 2016.
22. *Yu Zeng, Yunzhi Zhuge, Huchuan Lu, Lihe Zhang, Mingyang Qian, and Yizhou Yu*. Multi-source weak supervision for saliency detection // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. С. 6074-6083, 2019.
23. *Yin Li, Xiaodi Hou, Christof Koch, James M Rehg, and Alan L Yuille*. The secrets of salient object segmentation // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. С. 280-287, 2014.
24. *Vida Movahedi and James H Elder*. Design and perceptual validation of performance measures for salient object segmentation // 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition-Workshops. С. 49-56, 2010.
25. *Jiang-Jiang Liu, Qibin Hou, Ming-Ming Cheng, Jiashi Feng, and Jianmin Jiang*. A simple pooling-based design for realtime salient object detection // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. С. 3917-3926, 2019.
26. *Nian Liu, Junwei Han, and Ming-Hsuan Yang*. Picanet: Learning pixel-wise contextual attention for saliency detection // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. С. 3089-3098, 2018.
27. *Егоров Д.А.* Выявление аномального поведения группы объектов в видеопоследовательности // Сборник трудов xiii международной научно-технической конференции. В сборнике: технологии информационного общества. Материалы XIII Международной отраслевой научно-технической конференции. 2019. С. 167-169.
28. *Xuebin Qin, Zichen Zhang, Chenyang Huang, Chao Gao, Masood Dehghan, and Martin Jagersand*. Basnet: Boundaryaware salient object detection // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. С. 7479-7489, 2019.
29. *Deng-Ping Fan, Ming-Ming Cheng, Yun Liu, Tao Li, and Ali Borji*. Structure-measure: A new way to evaluate foreground maps// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. С. 4548-4557, 2017.
30. *Diederik P Kingma and Jimmy Ba*. Adam: A method for stochastic optimization. arXiv preprint, 2014.
31. *Lu Zhang, Ju Dai, Huchuan Lu, You He, and Gang Wang*. A bi-directional message passing model for salient object detection // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. С. 1741-1750, 2018.
32. *Власюк И.В.* Метод оценки эффективности алгоритмов обработки видеосигналов в телевизионных камерах при передаче движущихся объектов // Технологии информационного общества. материалы XIII международной отраслевой научно-технической конференции. 2019. С. 158-160.