Research on Image Binary Classification Based on Fast Style Transfer Data Enhancement

Shuang Zheng, Junfeng Wu, Fugang Liu(✉), Jingyi Pan, and Zhuang Qiao

Heilongjiang University of Science and Technology, Harbin 150022, China
liufugang_36@163.com

Abstract. The essence of image classification task is to extract high-level semantic content features of images. The traditional data enhancement methods based on convolutional neural network (CNN) are translation, rotation, clipping, noise adding, etc. These methods have not changed the content and style of image data. This paper proposes a fast style migration data enhancement method, which can quickly apply the style art of one image to another image without changing the high-level semantic content characteristics of the image. Through the experimental comparison, it is found that the method of fast style migration data enhancement proposed here can further improve the accuracy of the model compared with the traditional data.

Keywords: Convolutional neural network · Data enhancement · Style transfer · Image classification

1 Introduction

With the rapid development of the Internet, images have become important data information. For these massive images, it is difficult for people to quickly find the information they expect. This makes people urgently need a way to quickly and effectively obtain the required image content information from a large number of images. With the rapid development of image classification technology [1], convolution neural network has incomparable advantages in image processing based on convolution neural network has become the mainstream of research [2–4].

Data enhancement refers to expanding the number of data sets to increase diversity in the training of convolutional neural networks, especially in the case of insufficient samples, data enhancement becomes more important. On the one hand, popular deep architectures such as AlexNet [7] or VGGNet [8] have millions of parameters, so it is necessary to train a fairly large data set for a specific task, and lack of sufficient data will lead to overfitting [5, 6]. On the other hand, collecting raw data is very time-consuming and expensive in many computer vision tasks. For example, in key tasks such as medical image analysis, industry, and agricultural production, researchers are often limited by the lack of reliable data. Style transfer is a neural network-based image style transfer method proposed by Gatys et al. [7]. It solves the complex process of manual modeling

X. Jiang (Ed.): MLICOM 2021, LNICST 438, pp. 79–89, 2022.
https://doi.org/10.1007/978-3-031-04409-0_8

in traditional methods. The deep learning technique can be used to model the texture features and apply the artistic style of one image to the task of another. However, at the same time, it also has certain drawbacks. It requires constant iteration to generate pictures, which takes too long [8]. This paper proposes a data enhancement method based on rapid style transfer to expand data. Firstly, a fast style transfer network is designed for efficient style transfer, and secondly, the traditional data enhancement method and the style transfer data enhancement method are used to expand the data set. Finally, the data sets expanded by the two methods are divided into data sets, and the image classification network is used for training, and the advantages and disadvantages of the methods proposed in this paper are compared and analyzed through experiments.

2 Image Fast Style Transfer Model

Image conversion network *fw* and loss network Φ are the parts of the fast style migration network. As shown in Fig. 1, the main structures of *fw* and Φ are the depth residuals network [9] and the VGG-19 network, respectively. Loss network is defined as feature loss "lfeat" and style loss "lstyle" to measure the gap between content and style.

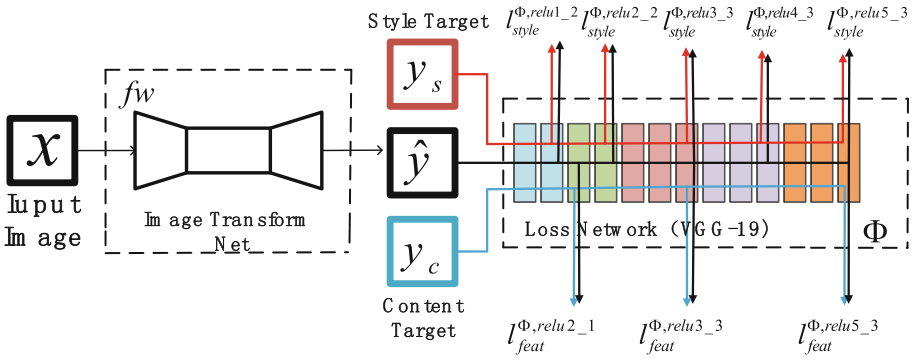


Fig. 1. Image quick style transfer system

2.1 Image Conversion Network

The theory shows that the depth of the network is an important factor in achieving efficient image recognition. This is because the deepening of the network will cause gradient explosion [10, 11] and gradient disappearance [12–14]. In order to solve the above problems and get a good image conversion network, this paper adopts the deep residual network as the backbone of the image conversion network. The deep residual network adds the residual module shown in Fig. 2 to the traditional convolutional neural network, so that the output of the previous layer is directly input to the input of the next layer without convolution operation. Assuming the input of a certain section of convolutional neural network is x and the expected output is $H(x)$. The identity mapping is passed to the next layer, so the learning residual function $F(x) = H(x) - x$.

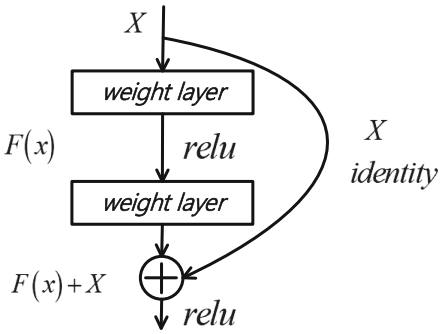


Fig. 2. Residual error module

Formula (1) is the parameter weight parameter W of image conversion network. It completes the conversion of input image x and output image by using $y = fw(x)$, and obtains $l_i(y, y_i)$ by loss function, to measure the difference between the generated image y of the output and the target image y_i.

$$W^* = \arg\min_W E_{x,\{y_i\}}\left[\sum_{i=1}\lambda l_i(f_W(x), y_i)\right] \tag{1}$$

2.2 Loss of the Network

For the loss network, this article defines two perceptual loss functions, content loss function and style loss function, to measure the difference in content and style between two pictures. Although the loss network here is also a convolutional neural network (CNN), the parameters are not updated, and are only used for the calculation of content loss and style loss. In order to ensure that this loss network has excellent extraction capabilities in terms of content and style, this paper uses the VGG-19 network model pretrained in ImageNet [15]. The VGG-19 network structure is shown in Fig. 3.

2.2.1 Content Loss Function

The formula (2) is the content loss function defined in this paper, $\phi_j(x)$ represents the jth layer of the network ϕ, the input is y, and $C_j \times H_j \times W_j$ is the shape of the feature map. This function penalizes the content deviation between the generated image and the target image, and requires the generated image to be very similar to the input target image in content details. The smaller the value of the loss function, the more similar the content of the images before and after processing, and vice versa.

$$l_{feat}^{\phi,j}(\hat{y}, y) = \frac{1}{C_j H_j W_j}\left\|\phi_j(y) - \phi(y)\right\|_2^2 \tag{2}$$

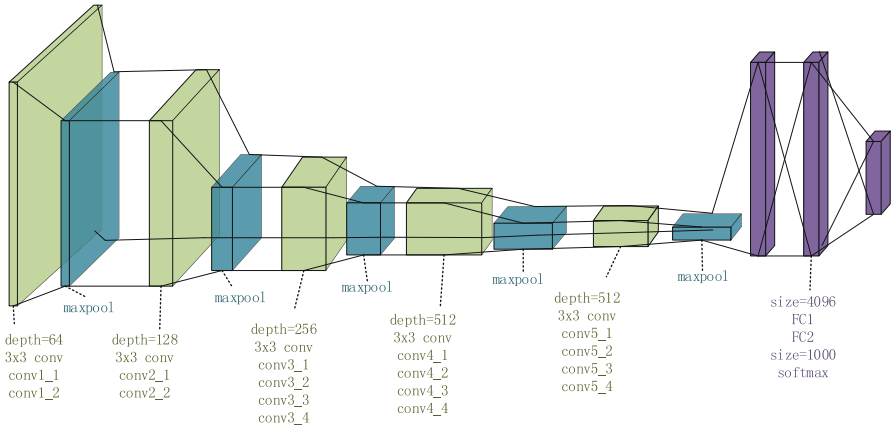


Fig. 3. VGG-19 network structure

2.2.2 Style Loss Function

It is also hoped that the style loss will penalize the deviation in style, such as color, texture, and common mode. In order to achieve this effect, this paper proposes the following style reconstruction loss function as formula (3). $\phi_j(x)$ represents the jth layer of ϕ. The input is x. The shape of the feature map is $C_j \times H_j \times W_j$.

$$G_j^{\phi}(x)_{c,c'} = \frac{1}{C_j H_j W_j} \sum_{h=1}^{H_j} \sum_{w=1}^{W_j} \phi_j(x)_{h,w,c} \phi_j(x)_{h,w,c'} \tag{3}$$

3 Analysis of the Fast Style Transfer Model

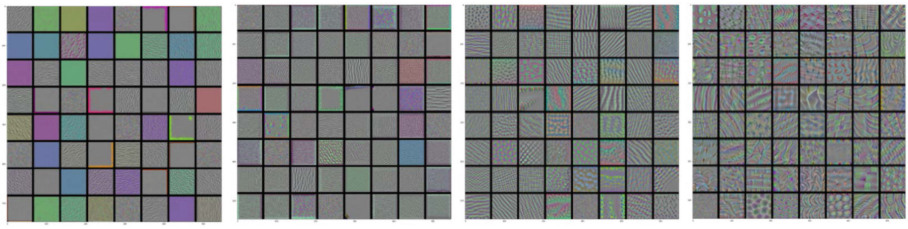
All style transfer models in this article are based on the VGG19 pre-training network, using Tensorflow deep learning framework to train coco data set (80000 pieces). The image size is normalized to 256 × 256, the number of iterations for all models is 20000, batch_size is set to 4, and epoch is set to 1.

3.1 Image Style Transfer

Visualizing the filters of the pre-trained VGG-19 network model will be of great help to the selection of the content and style extraction layer. The visualization of some filters from the low-level to the high-level is shown in Fig. 4.

From the visualization of these filters, it can be seen that the low-level filters extract simple directional edges and color features. As the number of layers deepens, the filter in the convolutional neural network becomes more and more complex, and the feature information that the filter can extract is also richer, such as feathers, leaves, eyes, etc.

In this paper, the higher layer conv5_3 is selected as the content extraction layer, and the lower layer conv1_2 and conv2_2 are selected as the style extraction layer to train



（a）block1 （b）block2 （c）block3 （d）block4

Fig. 4. Visual filter

(a) Target image (b) Style image (c) Generated image

Fig. 5. Image migration results

the style transfer model and combine the style transfer network to realize the image style transfer, as shown in Fig. 5.

From the results, when the high-level is only selected as the content extraction layer, and the low-level is used as the style extraction layer, the migration result is completely pixelated. This is because too high a layer contains too much content information, and a too low layer filter almost only extracts some unimportant texture information. So the trained model performs very poorly in style extraction, and it migrates the target image. When the target image is transferred, the single style information is over fitted to make the generated image completely pixelated. This article continues to improve the experiment, adjust the content and style extraction layer, select conv2_1, conv3_3, conv5_3 as the content extraction layer, select conv1_2, conv2_2, conv3_3, conv4_3, conv5_3 as the style extraction layer, train 3 migration models and combine the image conversion network to realize the image Style transfer (from left to right in Fig. 6); select conv3_3 as the content extraction layer, conv1_2, conv1_2, conv2_2, conv1_2, conv2_2, conv3_3, conv1_2, conv2_2, conv3_3, conv4_3, conv1_2, conv2_2, conv4_3, conv5_3 is the style extraction layer, which trains 5 style transfer models and combines the image conversion network to realize image style transfer (from left to right in Fig. 7).

From the results, it can be seen that multiple consecutive layers from the low-level to the high-level are selected as the style extraction layer, and the layer closer to the middle is selected as the content extraction layer can get the best migration effect.

(a) (b) (c)

Fig. 6. Migration results of different content extraction layer

(a) (b) (c) (d) (e)

Fig. 7. Migration results of different style layers

3.2 Image Generation Time

Through the experiment, it is found that when using the style transfer model to transfer the style of the target image, the transfer time of different size of the target image is often different. Use different style images to train a new style transfer model to continue the transfer, and make statistics on the transfer time as shown in Table 1.

Table 1. Timetable for different models.

	67 × 67	224 × 224	512 × 512	1024 × 1024
Style model 1	1.3742 s	2.5156 s	4.1475 s	9.3756 s
Style model 2	1.3786 s	2.5489 s	4.1549 s	9.3845 s
Style model 3	1.4025 s	2.5132 s	4.1135 s	9.3766 s
Style model 4	1.3812 s	2.6732 s	4.1303 s	9.2524 s
Style model 5	1.3744 s	2.5973 s	4.1387 s	9.3688 s

It can be seen from the statistical data in the above table that under the premise of the algorithm used in this article, the different style transfer models obtained by training are very close to the transfer time of the target image of the same size. With the increase of the size of the target image, the migration time will increase, that is, the migration time of the target image has nothing to do with the style migration model, but only with the size of the target image.

4 Image Binary Classification Based on Fast Style Transfer Feature Enhancement

For image classification tasks, the previous data enhancement only performed simple processing such as translation, cropping, rotation, and noise addition on the image. In this paper, the original image data set will be processed, and the main content features will be retained by using fast style migration technology, and some environment style features will be changed. Then, the training set is added to the image which changes some features to improve the number of samples and help image classification. In order to improve the classification accuracy of the classification model, this paper also verifies the effectiveness of the fast style migration data enhancement compared with the traditional data enhancement.

4.1 Choice of Style Model

Based on the previous work of this article, five style models (waves, feathers and leaves, abstract, animation, Van Gogh starry sky) have been trained. To save the time of image style transfer and ensure conversion effect, we first process the image size to 224×224, and then carry out style transfer. In order to select a suitable style transfer model for data enhancement, this paper uses the class activation algorithm (grad-CAM) to measure the effectiveness of the transfer model for data enhancement. Grad-CAM performs a reverse operation according to the output vector to obtain the gradient of the feature map, that is, the gradient map corresponding to the feature map, and then averages each gradient map. This average value corresponds to the weight of each feature map, and then the weight and feature map are weighted and summed, and the final class activation map is obtained through the activation function. Figure 8 is a diagram showing the class activation performance on the image generated by the style transfer with the target image as the output vector.

It can be seen from the thermal map of class activation after the target image is migrated by five kinds of fast style migration models that for the image whose target image is cat, the effect of wave style activation is the most obvious, and the thermal map of class activation almost covers the whole object; For the image whose target image is dog, the activation effect of abstract style is the most obvious. Therefore, this paper selects ocean wave style and abstract style from 5 style models to enhance the data of cat and dog in the binary image data set.

4.2 Image Classification Network

If the data set used by the pre-training network is large and general enough, the spatial level of the features can effectively serve as a general visual world model. Thus, these features can be applied to various computer vision problems, even if the categories involved in these new problems are completely different from the original tasks. VGG-16 pre-training network is a convolutional neural network with 1000 classes. Therefore, it is necessary to improve the VGG-16 pre-training network to perform two class prediction output. In this paper, the improvement of VGG-16 pre-training network is shown in

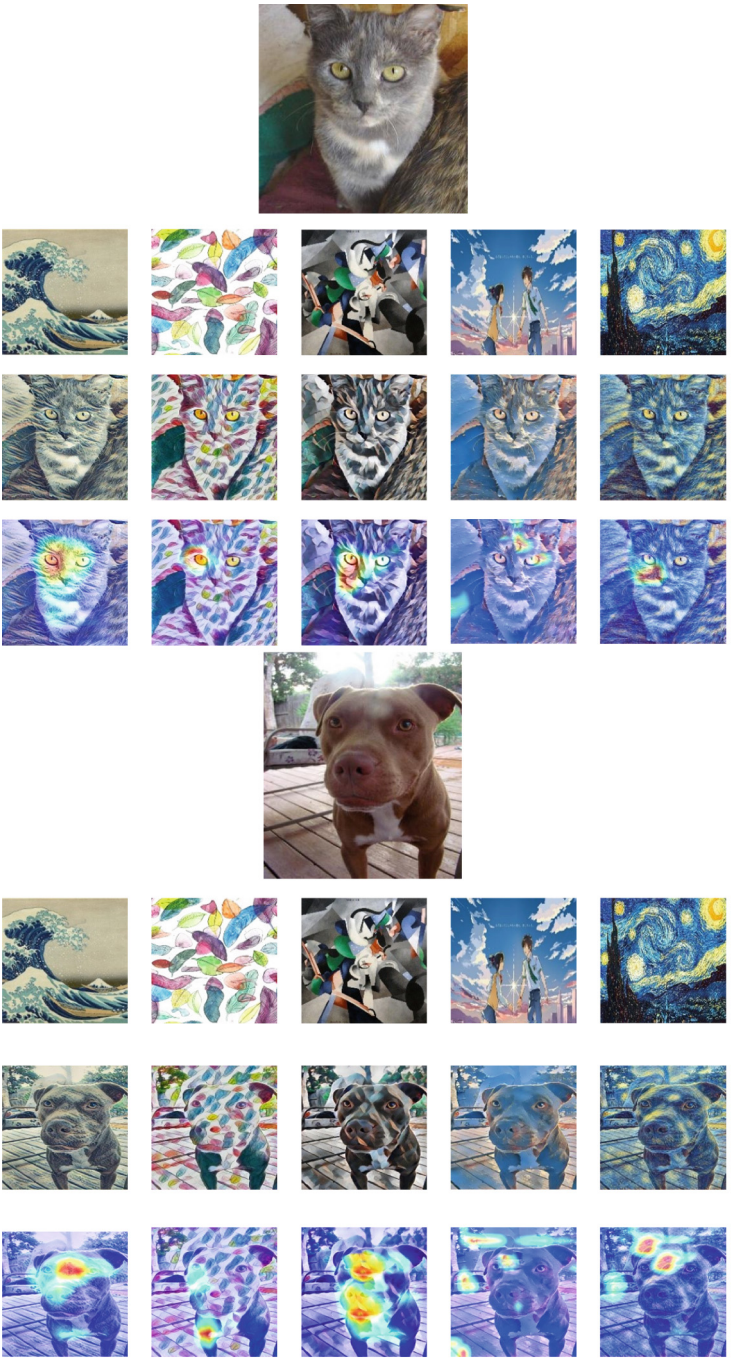


Fig. 8. Style transfer and class activation heat map

Fig. 9. The convolution base trained in the network is retained, all dense connection layers are deleted, and then a two-layer dense connection layer and sigmoid activation function are added after the convolution base.

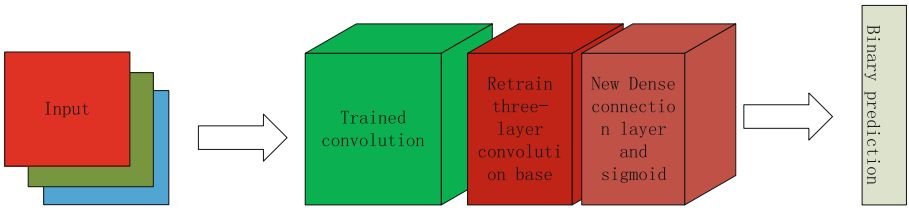


Fig. 9. Improved VGG-16 pre-training network model

4.3 Experiment and Analysis

The data set used in this article is partial data of two categories of cats and dogs in the "kaggle" data set, with 5000 images of cat and dog respectively. The image data of cats and dogs are respectively enhanced by the traditional data enhancement method selected at random and the fast style transfer method proposed in this paper. Two expanded data sets are obtained, and the number of images is 20,000 respectively. Among them, the styles used in the rapid style migration data enhancement methods for cats and dogs are ocean wave style and abstract style respectively. The two expanded data sets are divided into training set, verification set and test set, and the ratio of 6:2:2 is adopted. At the same time, in order to ensure the effectiveness of data enhancement, the enhanced data is also based on the ratio of 6:2:2.

Two expanded data sets are used to train on the improved VGG-16 pre-training network. The parameters are set as: learning rate 1e−5, epochs = 30, batch_size = 20, and loss function as binary_crossentropy. The final training result data is shown in Table 2.

Table 2. VGG-16 binary training data table

Data enhancement method	Training result
No	85.32%
Pan + crop	87.13%
Crop + rotate	86.57%
Rotate + noise	88.35%
Pan + noise	88.59%
Crop + rotate	85.43%
Wave style + abstract style	90.84%

From the statistical data in the above table, we can see that whether it is the traditional data enhancement method or the fast style transfer data enhancement method proposed in this article, the accuracy of classification prediction is relatively good, and both are above 85%.This is because this paper adopts the idea of transfer learning and uses the optimized VGG-16 pre training to train the network. The fast style migration data enhancement method proposed in this paper is significantly better than the traditional data enhancement method, and the maximum increase is 5 percentage points.

5 Conclusion

Based on the shortcomings of traditional data enhancement methods of diversified and single data after enhancement, this paper proposes a method of style transfer data enhancement to help convolutional neural networks in image classification, and addresses the shortcomings of traditional style transfer speed and poor extraction effect. Firstly, a fast style transfer model is designed, which can transfer target images of different sizes in a few seconds to complete a specific style transfer. Secondly, this article uses the grad-CAM algorithm to select appropriate style transfer styles for different data sets, and then selects different data sets Perform style transfer data enhancement. Finally, through experiments, it is found that the fast style transfer data enhancement method proposed here can help convolutional neural networks to classify images and improve the prediction accuracy of the model. The inability to process the data during the training process is one of the shortcomings of this article. The follow-up work of this article will focus on the study of more rapid and effective data enhancement methods, and is committed to improving the accuracy of the classification model.

Acknowledgements. This work has been partially supported by "Heilongjiang Science Foundation Project (LH2021F052)" and "2020 scientific research project of basic scientific research expenses of provincial colleges and universities in Heilongjiang Province (2020-KYYWF-0684)".

References

1. Guo, Y., Rothfus, T.A., Ashour, A.S., Si, L., Chunlai, D., Ting, T.-F.: Varied channels region proposal and classification network for wildlife image classification under complex environment. IET Image Process. **14**(4), 585–591 (2020). https://doi.org/10.1049/iet-ipr.2019.1042
2. Seo, S., Do, W.-J., Luu, H.M., Kim, K.H., Choi, S.H., Park, S.-H.: Artificial neural network for Slice Encoding for Metal Artifact Correction (SEMAC) MRI. Magn. Reson. Med. **84**(1), 263–276 (2020). https://doi.org/10.1002/mrm.28126
3. An, T., et al.: Black tea withering moisture detection method based on convolution neural network confidence. Food Process Eng. **43**(7),(2020). https://doi.org/10.1111/jfpe.13428
4. Qi, Y., Chen, J., Huo, Y., Li, F.: Hyperspectral image classification algorithm based on multi-scale convolutional neural network. Infrared Technol. **42**(9), 855–862 (2020). https://doi.org/10.3724/SP.J.7102910261
5. Krizhevsky, A., Sutskever, I., Hinton, G.E.: ImageNet classification with deep convolutional neural networks. Commun. ACM **60**(6), 84–90 (2017)

6. Helmrich, I.R.A.R., van Klaveren, D., Steyerberg, E.W.: Research Note: prognostic model research: overfitting, validation and application. J. Physiother. **65**(4), 243–245 (2019). https://doi.org/10.1016/j.jphys.2019.08.009
7. Gatys, L., Ecker, A., Bethge, M.: A neural algorithm of artistic style. J. Vis. **16**(12), 326 (2016). https://doi.org/10.1167/16.12.326
8. Zhao, X., Zhao, X.-M.: Deep learning of brain magnetic resonance images: a brief review. Methods **192**, 131–140 (2021). https://doi.org/10.1016/j.ymeth.2020.09.007
9. Zhang, Z., Tong Zhou, Y., Zhang, Y.P.: Attention-based deep residual learning network for entity relation extraction in Chinese EMRs. BMC Med. Inf. Decis. Making **19**(S2) (2019). https://doi.org/10.1186/s12911-019-0769-0
10. Chen, J., Xiang, Y.: Summarization of research on gradient instability in deep neural network training. **29**(07), 2071–2091 (2018)
11. Yuexiu, G., Wei, Y., Qi, L., Wang, Y.: Summary of residual network research. Comput. Appl. Res. **37**(05), 1292–1297 (2020)
12. Glorot, X., Bengio, Y.: Understanding the difficulty of training deep feedforward neural networks. J. Mach. Learn. Res. **9**, 249–256 (2010)
13. Tomoyuki, O., Kouyu, S., Naohiko, K.: Proposal and evaluation of pavement deterioration prediction method by recurrent neural network. Int. J. Adv. Res. Eng. **3**(4), 16 (2017)
14. Castillioni, K., Wilcox, K., Jiang, L., Luo, Y., Jung, C.G., Souza, L.: Drought mildly reduces plant dominance in a temperate prairie ecosystem across years. Ecol. Evol. **10**(13), 6702–6713 (2020). https://doi.org/10.1002/ece3.6400
15. Chunshui, C., Yongzhen, H., Yi, Y., et al.: Feedback convolutional neural network for visual localization and segmentation. IEEE Trans. Patt. Anal. Mach. Intell. **41**(7), 1627–1640 (2019)