DOI: 10.23977/jaip.2020.030104
ISSN 2371-8412

Tongue Localization Method Based on Cascade Classifier

Chao Song

School of Information Engineering, Nanjing University of Finance & Economics, Nanjing, 210046, China

853957104@qq.com

Keywords: Tongue localization, Cascade classifier, Local Binary Pattern

Abstract: Traditional Chinese Medicine (TCM) verifies that tongue images are closely related to the health of the human organs and tongues' visual features can provide valuable clues for disease diagnosis. How to locate the tongue region is an important step in the intelligent development of the TCM tongue diagnosis, because the effective removal of interference information outside the tongue can effectively enhance the extraction of tongue features. This paper proposes a cascade classifier based on Local Binary Pattern (LBP) feature to locate and segment the tongue body, which effectively improves the classification accuracy of the tongue feature.

1. Introduction

The localization and segmentation of the tongue region has always been a key step in the tongue image analysis process. Whether the localization of the tongue region can be accurately and quickly obtained is the basis for establishing a classification model. As shown in Figure 1, the original image of a person's face acquired by the image acquisition device contains other information that has nothing to do with the tongue target in addition to the tongue target. Therefore, in the tongue image collection process, it is often impossible to directly obtain a pure tongue image. However, other regions of the face other than the tongue body will cause a lot of interference in the extraction of tongue features. How to accurately detect the tongue area is the premise to improve the classification of tongue features.

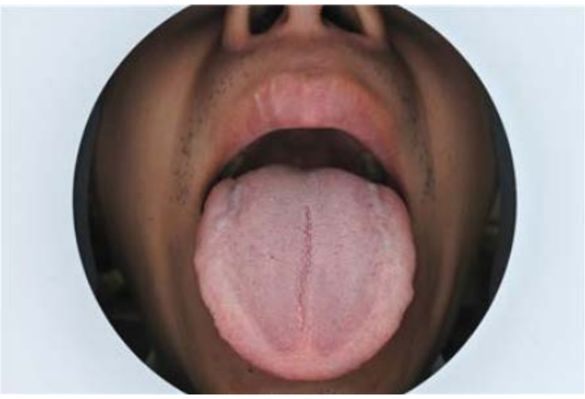

Figure 1. Original tongue image

As can be seen from the figure above, in addition to the tongue body, the original tongue image collected from professional equipment also contains information that has nothing to do with the tongue, such as lips, cheeks, and nostrils. The information will seriously interfere with the classification of tongue images. Therefore, in the process of tongue image analysis, tongue positioning is essential.

By observing the original image of the tongue, many factors interfere with the positioning and segmentation of the tongue region. For example, the color of the lips is close to the color of the tongue, which will make it easy to judge the lips as the tongue when judging the color characteristics. When collecting tongue data, we are unable to predict the position of the patient's tongue in the entire image and the size of the tongue, which leads to our inability to make good use of the position information of the tongue area. The constraints of these objective conditions make the tongue segmentation method have certain limitations in the degree of automation and segmentation progress. At the same time, the tongue image data we collected in a closed environment has a stable lighting environment, which can ensure a uniform light source, contrast, and exposure rate, which is beneficial to the segmentation of the tongue. The open environment and different lighting environments will have a certain impact on the existing tongue segmentation methods.

In this paper, a cascaded classifier based on LBP features is used for the localization and segmentation of the tongue region, which effectively removes the interference information in the original image and greatly improves the accuracy of the tongue image classification task.

2. Related work

As digital images, the tongue image contains a wealth of image feature information, such as color, shape, texture, etc. The traditional image processing technology captures these characteristics of the tongue image to realize the segmentation of the tongue image. The traditional tongue segmentation method tries different computer technologies to segment the tongue body from the background of the human face, mouth, and teeth, and extracts the features of the segmented tongue region, which can effectively improve the classification accuracy. Traditional tongue positioning and segmentation methods include GrabCut method [1], Otsu threshold method [2], and graph theory method [3].

2.1 GrabCut method

In recent years, researchers have proposed many methods for image segmentation. Among them, the segmentation method based on graph-cut has attracted the attention of researchers. This method transforms the image segmentation process into the process of solving the energy function minimization, and this energy function mainly contains the position information and edge information of the picture. Based on this theory, researchers have proposed a variety of image segmentation algorithms [4-6]. In 2004, the GrabCut algorithm was proposed on basis of the GraphCut algorithm and gradually became the mainstream method for segmenting images. However, due to the high cost of determining the parameters of the Gaussian mixture model, if the GrabCut algorithm is used directly, a good segmentation effect cannot be achieved. In order to solve this problem, researchers have proposed many improvement methods [7-10]. Through the improvement of GrabCut algorithm, its application in image segmentation is more extensive.

As shown in Figure 2, the basic idea of the GrabCut algorithm is to map an image into an s-t network graph. Among them, the source point s represents the foreground endpoint, and the sink point t represents the background endpoint. Model the segmented target and its background through the full covariance mixture Gaussian model (GMM) and generate a special vector $k = \{k_1, ..., k_n, ..., k_N\}$, k_n represents the Gaussian component corresponding to the n-th pixel, Where $k_n \in \{1, ..., K\}$. Each pixel comes from a Gaussian component of the target GMM or the background GMM.

As can be seen from Figure 2, the edge set E consists of two parts, one is the connecting edge between the source point and the sink and other nodes, and the other is the connecting edge between adjacent nodes. The weights on these edges reflect the similarity between the pixel and the foreground and background, and the color difference between adjacent pixels. This information can further reflect the regional attributes and boundary attributes of the image. Usually, these two kinds of information are called area energy and boundary energy. Here, the total energy, regional energy, and boundary energy of image segmentation are denoted as E(A), R(A) and B(A) respectively, then:

$$E(\underline{\alpha}, k, \underline{\theta}, z) = R(\underline{\alpha}, k, \underline{\theta}, z) + B(\underline{\alpha}, z) \tag{1}$$

The regional energy term is

$$R(\underline{\alpha}, k, \underline{\theta}, z) = \sum_{n} R_n(\alpha_n, k_n, \underline{\theta}, z_n) \tag{2}$$

$$\begin{aligned} & R_n(\underline{\alpha}, k, \underline{\theta}, z_n) = -\lg \pi(\alpha_n, k_n) + \\ & \frac{1}{2}\lg\det \sum(\alpha_n, k_n) + \frac{1}{2}[z_n - \mu(\alpha_n, k_n)]^T \times \\ & \sum(\alpha_n, k_n)^{-1}[Z_n - \mu(\alpha_n, k_n)] \end{aligned} \tag{3}$$

The boundary energy term is

$$B(\underline{\alpha}, z) = \gamma \sum_{(m,n)\in C} [\alpha_n \neq \alpha_m] \exp(-\beta) \left\| z_m - z_n \right\|^2 \tag{4}$$

Among them, z is image data, α_n is the attribute of pixel n, α_n=0 indicates the foreground, and α_n=1 indicates the background. k_n is the GMM label of pixel $\underline{\theta}$= {π(α, k), μ(α, k)},∑(α, k), α=0,k=1,...,K} It is the parametric model of GMM.

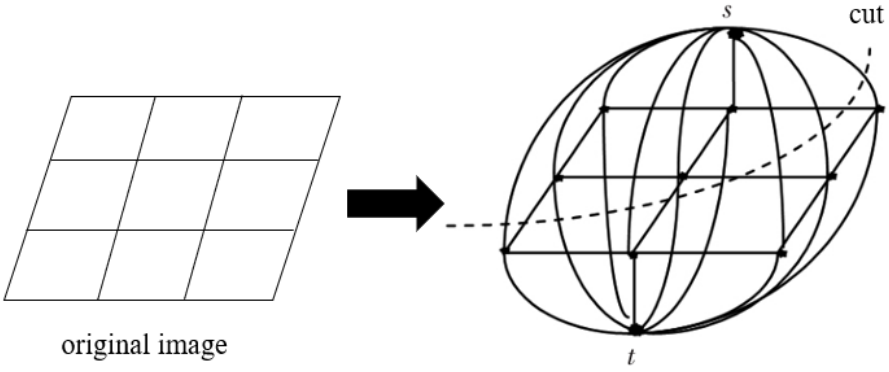

Figure 2. s-t network diagram

When the GrabCut algorithm performs image segmentation, it continuously updates the iteration and GMM parameters, and finally makes the algorithm tend to converge. In the iterative process, the group of parameters k, $\underline{\theta}$, α are continuously optimized, so that the segmentation energy E(A) is gradually reduced, and finally, E(A) can be guaranteed to converge to the minimum value, and finally, the tongue image segmentation is realized.

2.2 Otsu threshold method

In addition to the tongue image segmentation method based on GrabCut, threshold segmentation is also a widely used image segmentation method. The threshold method first extracts the gray information of the tongue image and then selects the best threshold through specific criteria to finally achieve the tongue image segmentation. The more widely used threshold algorithms are P-quantile method, bimodal method, and maximum between-class variance method (Otsu threshold method). The P-quantile method has good anti-noise performance, but it needs to pre-calculate the proportion of the whole background of the tongue image target station, so it will consume a lot of calculation cost. The bimodal method requires a large contrast between the target and the background to show a better segmentation effect. However, the tongue area and the background regions such as lips have a large degree of similarity in color. Using the bimodal method is easy to misjudge the lip region as a tongue body area. The Otsu threshold method has a simple calculation process and has a certain degree of anti-interference in terms of contrast and brightness, so it is more suitable for the segmentation of the tongue image. However, in the process of tongue image segmentation, if the Otsu threshold method is directly applied to the gray image, the obtained tongue will have long and narrow protruding areas and the edges of the tongue will have many burrs.

In recent years, some researchers have proposed combining the Otsu algorithm with a gray-scale projection method [11], Snake active contour method [12], and other algorithms to improve the tongue image segmentation effect, but there are still certain shortcomings. The ability to refine the algorithm in graphics will make it overfit to the experimental data, making it lack a certain generalization ability. At the same time, in order to improve the accuracy of the algorithm, it will also bring great difficulties to the algorithm design. Therefore, how to establish a simple, effective, and universal tongue image segmentation method has become a problem for researchers to solve.

2.3 Graph theory method

After the researchers tried the GrabCut method and the threshold method, another tongue image segmentation method based on graph theory appeared. The mainstream methods include the minimum spanning tree method, the Normalized-Cut method, and the main set-based method [13]. Their unified method idea is to convert the image into a weighted undirected graph form. The nodes in the undirected graph are the pixels in the original image, and then these pixels are distinguished by clustering to distinguish the image where the tongue body is included and the tongue body is not included, and finally, the tongue body image segmentation is realized.

The graph theory method uses the characteristics of the data structure in the tongue image, divides it into different enclosed areas, and clusters these areas, thereby effectively segmenting the tongue image. Pedro and Daniel proposed a segmentation method based on the minimum spanning tree, which became the Graph-Based method [14]. The workflow of this method is to first use the region growing method to pre-divide the original image. The principle of division is that these regions have a high degree of similarity. Combined with the MST method, the divided regions are segmented. This method sets a threshold function to control the size of the segmentation area, and this threshold is called k. However, there is no clear standard for determining the value of k, so it is difficult to apply it to different segmentation tasks. Later, Huang Qian et al. also proposed that the Graph-Based method may lead to the incorrect merging of image partitions. To solve this problem, they proposed to merge the canny edge detection method with this method to improve the performance of the algorithm, and the problem of the excessive merging of images was successfully avoided.

When using the Graph-Based method in tongue image segmentation, because the tongue body and the background have high similarity in some cases, it is possible to segment the tongue body and the background. At the same time, when there are tongue features such as cracks on the tongue, the algorithm will divide the tongue into several parts, which will directly lead to incorrect segmentation results. These traditional tongue image segmentation methods have certain deficiencies in the segmentation effect and the degree of automation. Therefore, this paper proposes to segment the tongue image by combining the cascade classifier in the machine learning method.

3. Construction of cascade classifier

Because the original tongue image data we obtained contains a large number of non-tongue parts, and these parts will affect the subsequent neural network's extraction of tongue image features. Therefore, to reduce the cost of manually extracting tongue features for target detection, and realize automatic location detection of tongue parts, this paper proposes to combine LBP features with AdaBoost (adaptive boosting) algorithm to form a screening cascade classifier for tongue The volume image pixels are classified into two categories, and the tongue target is automatically located and coarse segmentation is realized. Compared with the traditional segmentation method, the cost is lower. After the positioning of the tongue body area, a large amount of interference information is removed, which lays a solid foundation for the establishment of a tongue feature classification model.

3.1 LBP feature description operator

LBP (Local Binary Pattern) [15] is an operator used to describe the local texture features of an image. It has the characteristics of gray-level invariance and rotation invariance, and it is significant in extracting texture image features from different angles. Advantage. As shown in Figure 3, since the texture of the tongue is significantly different from other organs of the face, this paper extracts the features of the tongue image by extracting the LBP features of the original image.

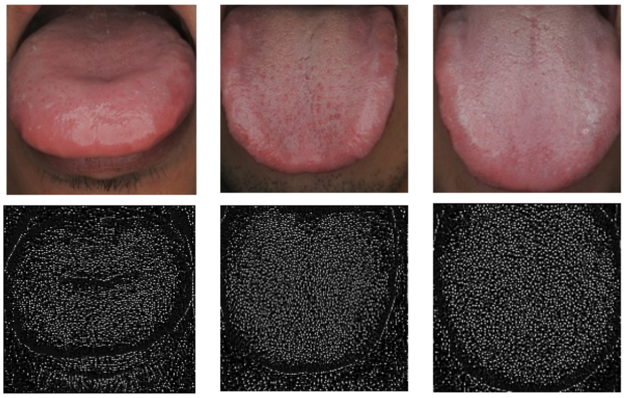

Figure 3. Tongue image (above) and its LBP feature extraction effect image (below)

In the 3*3 matrix, the gray value of the center pixel and the gray value of its neighboring pixels are compared one by one. If the neighboring pixels are greater than the center pixel value, the pixel will be marked as 1, otherwise, it is marked as 0. Finally, a pixel matrix containing only 0 and 1 is obtained. After arranging these 0 and 1 in a certain order, a binary number is obtained. The decimal number obtained by the value of this number is the LBP value of the center pixel and it reflects the texture information of the surrounding area. This feature calculation method is the original LBP feature description operator.

The formula of the LBP operator is as follows:

$$LBP(x_c, y_c) = \sum_{p=0}^{P-1} 2^p s(i_p - i_c) \quad (5)$$

Among them, (x_c, y_c) is the coordinate of the center pixel, p is the p-th pixel in the neighborhood, i_c is the gray value of the center pixel, i_p is the gray value of the neighborhood pixel, s(x) is the sign function, The specific definition is as follows:

$$s(x) = \begin{cases} 1 & x \geq 0 \\ 0 & \text{otherwise} \end{cases} \quad (6)$$

By extracting the LBP feature of the original image, the texture feature of the tongue image can be extracted, so that the subsequent training process of the classifier will focus on the texture difference of the sample. Through the extraction of the texture feature of the tongue image, effective classification can be trained the detector can have higher detection performance.

3.2 Training of cascade classifier

The texture of the tongue is an important feature for identifying the tongue region. By extracting the LBP features in the image, the tongue region can be well distinguished from other regions, to achieve an effective tongue segmentation effect. Therefore, this paper combines LBP features with AdaBoost (adaptive boosting) algorithm to form a screening cascade classifier, and then achieve the automatic positioning and segmentation of the tongue body saves the time cost and calculation cost of constructing the tongue image training data set.

Boosting algorithm is an iterative process, used to adaptively change the distribution of training samples so that the base classifier focuses on the samples that are difficult to distinguish. The iterative process is as follows: Firstly, the training samples are given, and then the weights of the training samples are initialized, the weak classifier is trained iteratively, the error rate of the weak classifier is calculated, the appropriate threshold is selected to minimize the error, and the sample weight is updated. After T cycles, T weak classifiers are obtained, and a strong classifier is finally obtained by weighting and superimposing the weight of the importance of each weak classifier. The formula of the weak classifier is as follows:

$$h_j(x) = \begin{cases} 1 & p_j f_j(x) < p_j \theta_j \\ 0 & \text{otherwise} \end{cases} \quad (7)$$

Among them, $h_j(x)$ is the judgment value of the weak classifier, and a value of 1 means that the picture is a tongue, otherwise it is a non-tongue; x represents the input picture sub-window, $f_j(x)$ is the j-th image on the x image The value of the feature; θ_j is the classifier threshold; p_j is the inequality sign direction. If the weak classification result is greater than the threshold, it is -1, otherwise, it is +1 to ensure that the inequality sign direction remains unchanged.

After T iterations, T best weak classifiers $h_1(x)$, ..., $h_t(x)$ are obtained, which is combined by the following formula to construct a strong classifier.

$$C(x) = \begin{cases} 1 & \sum_{t=1}^{T} \alpha_t h_t(x) \geq \frac{1}{2} \sum_{t=1}^{T} \alpha_t \\ 0 & \text{otherwise} \end{cases} \quad (8)$$

Among them, T is the number of cascaded classifiers and the number of selected features, αt is the weight of a single classifier. In this paper, we calibrate the tongue body part as positive samples, and the non-tongue body parts as negative samples, and train through the cascade classifier to generate a strong classifier. Then slide on the target image continuously through the small window. Each time it slides to a position, feature extraction is performed on the area in the small window. If the extracted features pass the judgment of the trained strong classifier, the small window is judged the region contains the tongue body part, and finally, the detected tongue body region is segmented.

The workflow of the cascaded classifier proposed in this paper is specifically: calibrate the tongue body part as positive samples, and calibrate the non-tongue body part as negative samples, train through the cascade classifier to generate a strong classifier; then use a small window to display the target image Slide up continuously, each time it slides to a position, feature extraction is performed on the area in the small window. If the extracted features pass the judgment of the trained strong classifier, it is determined that the area where the small window is located contains tongue Part, finally segment the detected tongue area. Through the iterative training of the LBP features extracted from the tongue image by the cascaded classifier, it is possible to construct a classifier that distinguishes the texture of the tongue region from the texture of other regions of the face. From the point of view of detection effect, the method proposed in this paper has high accuracy and can accurately locate and segment the tongue region. Compared with traditional segmentation methods, it has higher segmentation performance and automation.

4. Experiments

4.1 Dataset

The experimental data in this paper are provided by Shanghai University of Traditional Chinese Medicine. These image data were collected from multiple hospital collection points. The dataset includes 516 images of tooth-marked tongue, 566 images of non-tooth-marked tongue, 391 images of cracked tongue, 250 images of non-cracked tongue, 392 images of thick tongue coating and 130 images of thin tongue coating. The size of the original images is fixed at 5568 * 3172 Pixels.

4.2 Evaluation metric

In the field of machine learning, the measurement and evaluation of models is crucial. Only by selecting an evaluation metric that matches the problem, can we quickly find problems that may occur during model selection and training. In classification problems, we usually use classification accuracy to evaluate the quality of the network model. According to the prediction or correct or incorrect prediction of the classifier on the test dataset, it can be divided into four cases: TP and FN refer to positive classes are predicted as positive classes and as negative classes respectively, TN and FP refer to negative classes are predicted as negative classes and as positive classes respectively. P and N indicate the number of all positive and negative classes. The accuracy of the model is calculated as follows:

$$Accuracy = \frac{TP+TN}{P+N} \tag{9}$$

According to the formula above, the accuracy of the model refers to the proportion of samples in the test set that are correctly predicted. By analyzing the accuracy, we can evaluate whether a network model has good recognition performance.

4.3 Analysis of experimental results

To further explore the influence of different tongue image data sets on the deep transfer learning feature extraction process, this paper selects three different training data: original tongue image, manually segmented tongue image and tongue image segmented based on cascade classifier (As shown in Figure 4), through deep transfer learning on the VGG16 network[16] to explore the feature extraction performance of data with different segmentation effects under the deep transfer neural network.

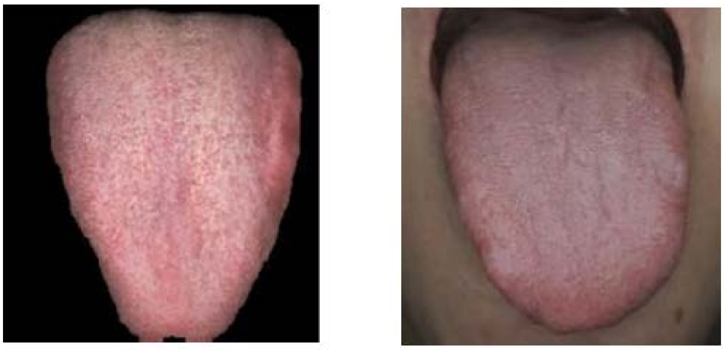

Figure 4. Tongue image segmented manually and by cascade classifier

Table 1 Classification accuracy of tongue features based on different segmentation methods (%)

Training data	toothmarked/nontoothmarked	cracked/noncracked	thick /thin coating
Original image	52.50	66.20	88.00
Manually segmented	97.80	95.90	94.00
Segmented by cascade classifier	**98.00**	**95.50**	**97.40**

The experimental results show that comparing the original image, the tongue image data after manual segmentation, and the tongue image after the cascade classifier localization and segmentation have significantly improved the classification accuracy after deep transfer learning. The classification accuracy of the cascade classifier segmentation method is slightly higher than that of manual segmentation, and it is far lower than manual segmentation in terms of time and labor costs. Therefore, this paper chooses the method of cascading classifiers to automatically locate the original data, which provides a good data foundation for the subsequent exploration of the influence of different network structures and network depths on the classification of tongue features.

5. Conclusion

This paper introduces different tongue image segmentation methods and elaborates the cascaded classifier based on LBP features proposed in this paper, which realizes the automatic localization of tongue targets and effectively removes the interference information in the original image. The classification effects of the original images, manually segmented images, and image segmented by a cascade classifier are compared. It laid a solid foundation for the subsequent extraction of tongue features and the establishment of classification models.

Acknowledgements

The author gratefully acknowledges the financial support from Postgraduate Research & Practice Innovation Program of Jiangsu Province (Grant No. KYCX18_1393).

References

[1] Wei Yuke, Fan Peng, Zeng Gui. Application of improved GrabCut method in tongue diagnosis system [J]. Sensors

and Microsystems, 2014, 33 (10): 157-160.

[2] Jiang Shuo, Hu Jie, Xia Chunming, et al. Tongue image segmentation method based on Otsu threshold method and morphological adaptive correction [J]. High Technology Letters, 2017 (2).

[3] Chen Shanchao, Fu Hongguang, Wang Ying. Application of an improved graph theory segmentation method in tongue image segmentation [J]. Computer Engineering and Applications, 2012, 48 (5): 201-203.

[4] Boykov Y Y. Interactive graph cuts for optimal boundary & region segmentation of objects in n-d images [C]// Proc Eighth IEEE International Conference on Comput Vis. IEEE Computer Society, 2001.

[5] Rother C. GrabCut: Interactive foreground extraction using iterated graph cuts [J]. Proceedings of Siggraph, 2004, 23.

[6] Agarwala A, Dontcheva M, Agrawala M, et al. Interactive Digital Photomontage [J]. ACM Transactions on Graphics, 2004, 23 (3).

[7] Zhou Liangfen, He Jiannong. Improved image segmentation algorithm based on GrabCut [J]. Journal of Computer Applications, 2013, 33 (01):49-52.

[8] Xu Qiuping, Guo Min, Wang Yarong. Fast image segmentation algorithm based on multi-scale analysis and graph cutting [J]. Application Research of Computers, 2009, 26 (10): 3989-3991.

[9] Han S, Tao W, Wang D, et al. Image Segmentation Based on GrabCut Framework Integrating Multiscale Nonlinear Structure Tensor [J]. Image Processing IEEE Transactions on, 2009, 18 (10): p.2289-2302.

[10] Shanmugavadivu P, Thenmozhi G. Detection of microcalcification in mammogram images using semi-automated texture based Grabcut Segmentation [C]// International Conference on Emerging Trends in Science, Engineering & Technology. IEEE, 2012.

[11] Zhang Ling, Qin Jian. Tongue image segmentation method based on gray projection and automatic threshold selection [J]. Chinese Tissue Engineering Research and Clinical Rehabilitation, 2010, 14 (09): 1638-1641.

[12] Zhang Zhishun, Liu Yong. Tongue extraction algorithm based on dynamic threshold and modified model[J]. Computer and Modernization, 2014 (11): 49-52.

[13] Huang Qian, Yang Wenliang, Gu Jiefeng. An improved image segmentation method based on graph theory[J]. Science Technology and Engineering, 2009, 9(13): 3652-3656+3671.

[14] Dongcai S. Efficient Graph based image Segmentation [J]. image processing.

[15] P. Viola, M. Jones, Rapid object detection using a boosted cascade of simple features, in: Proc. 2001 IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognition. CVPR 2001, IEEE Comput. Soc, 2001: pp. I-511-I–518.

[16] K. Simonyan, A. Zisserman, Very deep convolutional networks for large-scale image recognition, 3rd Int. Conf. Learn. Represent. ICLR 2015 - Conf. Track Proc. (2015) 1–14.