

Facial Wrinkle Detection Based on DeepLabV3+ and Semi-Automatic Labelling Strategy

Jiaxuan ZHONG^a, Xun LANG^{a1}, Xieyang ZHANG^a, Bingbing HE^a, Zhao ZHANG^b, Yufeng ZHANG^a

^a*Department of Electronic Engineering, Information School, Yunnan University, Kunming, 650091, China;*

^b*Department of Dermatology, Kunming Children's Hospital, Kunming Medical University, Kunming, 650000, China*

Abstract. Facial skin wrinkles are positively correlated with physiological age and are an important feature of aging. The traditional wrinkle detection algorithms are influenced by facial features and image backgrounds, thus cannot distinguish hair, eyes and eyebrows well, and thus need to cut the facial region into multiple blocks. Motivated by the above challenges, this work presents a facial wrinkle detection algorithm based on DeepLabV3+ and a semi-automatic labelling strategy, which is featured by following procedures: (i) The algorithm first combines the facial texture features and rough annotation of wrinkles by dermatologists to generate the ground truth required for deep learning. (ii) A lightweight network, MobieNetV2, is employed as the backbone model to reduce the amount of network parameters and calculations. The constructed deep learning model is then trained using the original images and ground truth labels. (iii) The accuracy of various algorithms is evaluated using the Jaccard Similarity Index (JSI). The results demonstrate that the proposed method exhibits superior performance in wrinkle detection.

Keywords. wrinkle detection; deep learning; deeplabv3+; semi-automatic annotation

1. Introduction

Senescence is an irreversible and natural process that occurs with advancing age, where wrinkles caused by noticeable structural changes in the skin are prominent features of aging [1]. Skincare and cosmetic treatments are becoming a focus of personal attention. Dermatologists have proposed various treatment strategies for facial wrinkle reduction [2-3], while cosmetic companies have launched a range of products. During the above process, wrinkle detection information can serve as an effective means of providing feedback on treatment efficacy.

Currently, various methods have been proposed for facial wrinkle detection. Loden et al. [4] were pioneers in the field of wrinkle detection, by using silicone models to create replicas of wrinkles to measure the characteristics of wrinkles. However, the efficiency of this method is limited owing to the difficulty of obtaining wrinkle replicas

¹ Corresponding author: Xun LANG, Department of Electronic Engineering, Information School, Yunnan University, e-mail: langxun@ynu.edu.cn

that exactly match the actual skin morphology. Batool et al. [5] proposed a Gabor filter-based wrinkle curve object detection method to quickly locate facial wrinkles. However, under high-resolution images, there are more noises, resulting in a lower wrinkle detection rate. Ng et al. [6] introduced an automatic approach for automatic detection and quantification of facial wrinkles, known as the Hybrid Hessian Filter (HHF), which is limited to detecting coarse wrinkles in local facial regions. Recently, to address the limitations of traditional wrinkle detection methods, Kim et al. [7] proposed a method that utilizes semi-automatic labeling to train U-Net for facial wrinkle detection. We highlight that the accuracy of this approach still requires improvement. In order to enhance the accuracy of wrinkle detection, further optimization of the network model is required. To this end, this paper presents a novel approach for facial wrinkle detection method based on DeepLabV3 Plus and a semi-automatic labelling strategy. The method introduces a semi-automatic labelling strategy developed by us to construct a dedicated dataset for wrinkle. Subsequently, the dataset is used as input data and trained using the DeepLabV3+ model [8] to enable the detection of facial wrinkles effectively. The current experiment conducted a performance comparison between the proposed algorithm in this paper, traditional methods, and the U-Net network on a self-created dataset for wrinkle detection, using the JSI metric. The experimental results demonstrate the effectiveness of the proposed algorithm for wrinkle detection.

The remaining sections of the paper cover the following: Section 2 elaborates on the process of generating the dataset with the aid of a semi-automatic labelling strategy. Section 3 introduces the theory relevant to this study. The experimental setup as well as the results are presented in Section 4, followed by a comprehensive summary in Section 5.

2. The Dataset Production Process

Deep learning algorithms rely on a large amount of annotated image datasets. Currently, no publicly available dataset specifically focused on wrinkles has been found. To solve this problem, we propose a semi-automatic labelling strategy that combines the texture mapping of the original images with rough annotations provided by dermatologists to generate ground truth images, as illustrated in Figure 1. These generated ground truth images serve as the training dataset required for this study. The flow of the proposed strategy is shown in Fig. 1 and consists of the following steps:

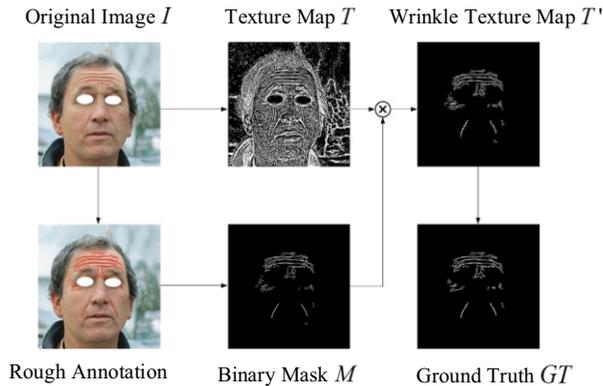


Figure 1. An example of generating the ground truth GT for wrinkle detection.

- (i) Firstly, regions of the original facial image I that exhibit wrinkle features are roughly labelled by dermatologists and converted into binary masks M .
- (ii) The feature texture maps T are extracted from the original facial images I using a Wiener filter [9] through Eq. (1).

$$T(x, y) = \left(1 - \frac{I(x, y)}{1 + I_{w(m,n)}(x, y)}\right) \times 255 \quad (1)$$

Where $I_{w(m,n)}$ is a Wiener filtered image, (m, n) represents that the filter size is $m * n$, and (x, y) is the coordinates of the pixel point.

- (iii) By multiplying, the binary mask M is multiplied with the feature texture map T , the unwrinkled texture is removed from T , resulting in a new texture feature map T' , as given by Eq. (2):

$$T'(x, y) = \begin{cases} T(x, y), M(x, y) > 0 \\ 0, otherwise \end{cases} \quad (2)$$

- (iv) Finally, adaptive thresholding [10] is used to generate a binarized image of the ground truth GT from the T' .

3. Theory of Related Work

3.1 DeepLabV3+ Network Model

DeepLabV3+ is a deep convolutional neural network model [11] proposed by Google, which is an improved version of the original DeepLabV3 network [8]. As shown in Figure 2, it consists of an encoder and a decoder.

In the encoding phase, DeepLabv3+ uses Xception as the backbone network to deepen the network. In addition, this network draws inspiration from the PSPNet network and improves it into the Atrous Spatial Pyramid Pooling (ASPP) module, which enables the segmentation of multi-scale segmentation.

In the decoding phase, the feature maps obtained from the encoding phase are first upsampled by 4-fold, and then these upsampled feature maps are then concatenated with the corresponding feature maps from the feature extraction backbone network along the channel dimension. Subsequently, another 4-fold upsampling is performed to restore the target boundary information and obtain the final prediction results.

3.2 The backbone structure of MobileNetV2

To facilitate the deployment of deep learning on mobile or embedded devices for various applications, we employ the lightweight network MobileNetV2 as the backbone network. Its lightweight design is primarily manifested through the use of depthwise separable convolutions [12]. Depthwise separable convolutions consist of depthwise convolutions and pointwise convolutions, whose specific process is illustrated in Figure 3.

Deep convolutions operates on a per-channel basis, where each channel is convolved with a dedicated kernel. The number of convolutional kernels is equal to the number of

channels, and the output channels remain unchanged. Pointwise convolution is similar to traditional convolution, but employs a kernel size of 1×1 . Its purpose is to fuse feature information across channels and generate feature maps for specific channels.

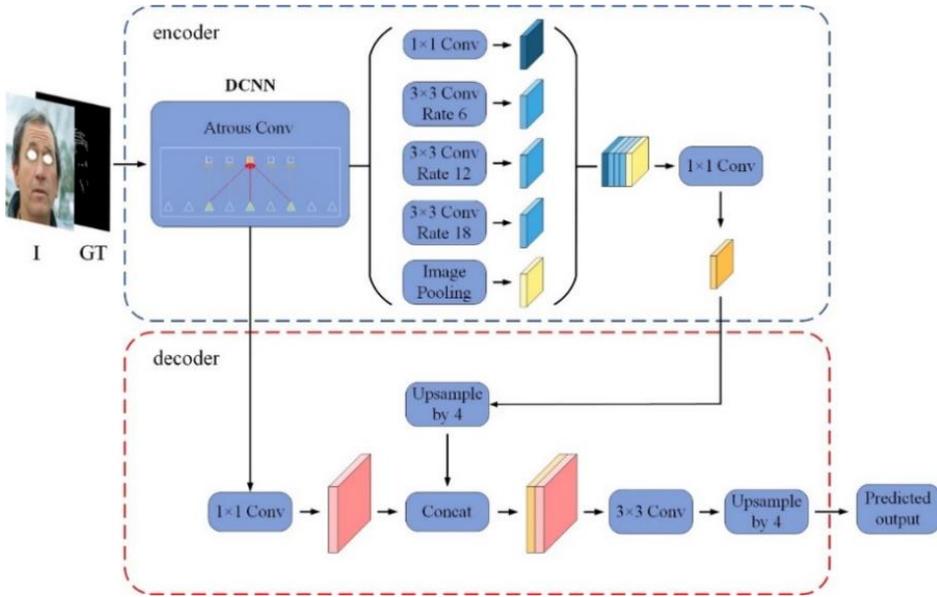


Figure 2. DeepLabV3+ network structure

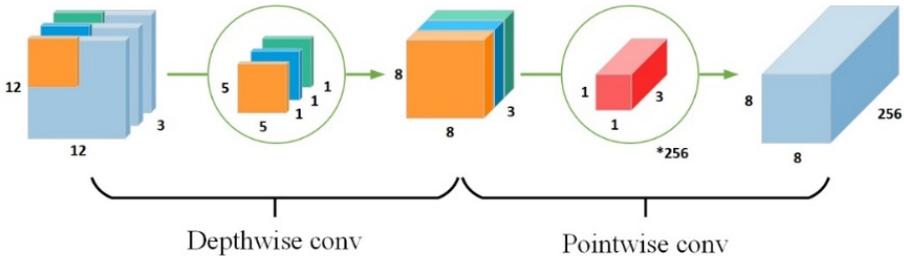


Figure 3. The process of depthwise separable convolution

3.3 Evaluation indicators

To analyze and compare the overall detection performance of various wrinkle detection algorithms and to evaluate their strengths and weaknesses, this study adopts the Jaccard similarity index (JSI) [6] as an evaluation metric, which is analogous similar to the definition of Intersection over Union (IOU). JSI is dedicated to assess the similarity and dissimilarity between a finite set of samples. As shown in Eq. (3), the higher the JSI, the higher the sample similarity.

$$J(A,B) = \frac{|A \cap B|}{|A \cup B|} \tag{3}$$

where A and B are annotation encodings, which in this paper refer to wrinkle detection results and the aforementioned GT , and $J(A, B) \in [0, 1]$.

4. Experiment

4.1 Data and environment

To verify the effectiveness of our proposed method for wrinkle detection, 300 facial images with 1024×1024 pixel containing visible wrinkle features were selected from the Flickr-Face-HQ dataset [13] in this study. Following the process described in Section 2, 300 ground truth images GT were generated for training and analysis. More specifically, the target dataset GT was divided into three parts: a training set of 225 images, a validation set of 25 images, and a test set of 50 images. The computer hardware configuration of the computer used in the experiment included an Intel Core i9-12900K CPU, an RTX 3070 graphics card, and 32GB of RAM. The deep learning framework used was PyTorch.

4.2 Result

The proposed method is compared with the U-Net network and the Hessian Filter method based on image processing on the test set. Due to the limitations of traditional methods, they only show good detection of horizontal wrinkles in localized regions such as the forehead, while they cannot differentiate regions such as glasses, hair, and eyebrows. Therefore, to better compare the detection performance of different methods, only the forehead and periocular regions were selected as the experimental objects.

The JSI values between the detection results set of each algorithm and the ground truth GT set are shown in Table 1. The results above indicate that the proposed algorithm detects fewer false wrinkles compared to other algorithms, resulting in a higher accuracy in wrinkle detection.

To demonstrate the effect of this study more intuitively, we selected three typical images of forehead wrinkles and three images of eye wrinkles from the test set, as shown in Figures 4 and 5, respectively. Figures 4 and 5 display the superimposed images of each algorithm's detection results and the original images. It can be observed that the traditional method exhibits the poorest detection performance, with issues such as failing to detect large areas of facial wrinkles and inaccurate edge segmentation. The use of the U-Net network compensates for some of the drawbacks of the traditional methods, but the detection rate for fine wrinkles remains limited. Our proposed algorithm in this study further improves upon the aforementioned issues found in the previous methods. Compared to other algorithms, it detects fewer false wrinkles, produces clearer edge segmentation, and achieves a higher accuracy in wrinkle detection.

Table 1. Comparison of JSI performance

ROI	JSI		
	Hessian	U-Net	DeepLabV3+
Forehead	0.20	0.55	0.62
Eye area	0.18	0.58	0.64

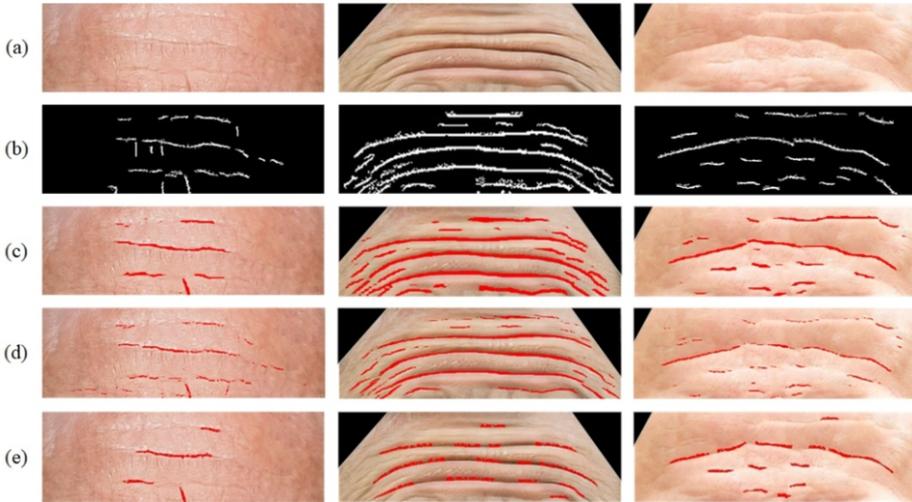


Figure 4. The comparison for detected wrinkle results on foreheads. (a) is an original image, (b) is a ground truth, (c) is a result of DeepLabV3+, (d) is a result of U-Net and (e) is a result of Hessian Filter.

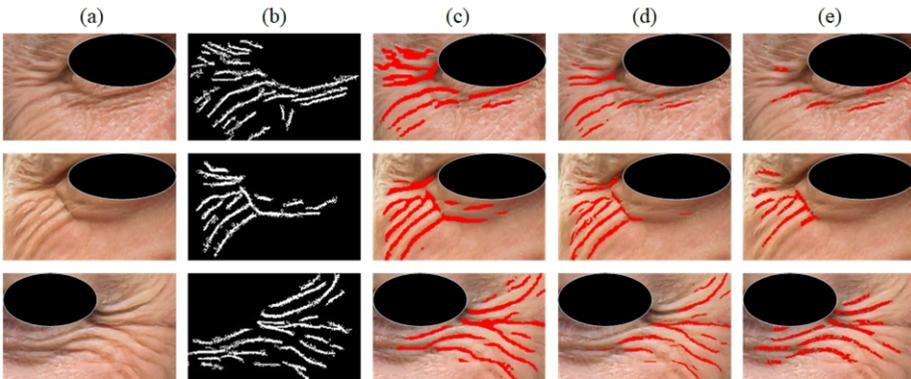


Figure 5. The comparison for detected wrinkle results on eye area. (a) is an original image, (b) is a ground truth, (c) is a result of DeepLabV3+, (d) is a result of U-Net and (e) is a result of Hessian Filter.

5. Conclusion

In this study, a facial wrinkle detection model based on semi-automatic labelling strategy and a deep learning has been proposed. The semi-automatic labelling strategy is utilized to generate ground truth GT from the original images I , which are then used in conjunction with the original images I for model training. The results demonstrate that the proposed method outperforms existing approaches in terms of wrinkle detection as characterized by the JSI metric. However, we highlight that the proposed method still has limitations in detecting fine wrinkles. For example, it occasionally misidentifies a small number of hair strands as wrinkles. Therefore, future work will focus on further enhancing the detection capability of fine wrinkles to improve the overall detection accuracy.

Acknowledgements

This work was supported in part by the National Natural Science Foundation of China under Grant NO. 62003298, 62201495 and 62261057, in part by the Yunnan Fundamental Research Projects under Grant NO. 202201AT070577, and the Second Professional Degree Graduate Practice Innovation Project of Yunnan University under grant No. ZC-22221926.

References

- [1] Quatresooz, P., Thirion, L., Piérard-Franchimont, C., & Piérard, G. E. (2006). The riddle of genuine skin microrelief and wrinkles. *International journal of cosmetic science*, 28(6), 389-395. <https://doi.org/10.1111/j.1467-2494.2006.00342.x>.
- [2] Gao, L., Kang, H., Li, Y., Lu, M., Song, W., Wang, Y., ... & Wang, G. (2020). Clinical efficacy and safety of 3DEEP multisource radiofrequency therapy combined with fractional skin resurfacing for periocular skin aging. *The Journal of Clinical and Aesthetic Dermatology*, 13(3), 41. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7159313/>.
- [3] Wu, X., Cen, Q., Zhu, J., Shang, Y., & Lin, X. (2022). Effectiveness and safety of nonablative fractional laser and infrared bipolar radiofrequency for treating periorbital wrinkles. *Journal of Cosmetic and Laser Therapy*, 24(6-8), 91-97. <https://doi.org/10.1080/14764172.2022.2120618>.
- [4] Lodén, M., Buraczewska, I., & Halvarsson, K. (2007). Facial anti-wrinkle cream: influence of product presentation on effectiveness: a randomized and controlled study. *Skin research and technology*, 13(2), 189-194. <https://doi.org/10.1111/j.1600-0846.2007.00220.x>.
- [5] Batool, N., & Chellappa, R. (2015). Fast detection of facial wrinkles based on Gabor features using image morphology and geometric constraints. *Pattern Recognition*, 48(3), 642-658. <https://doi.org/10.1016/j.patcog.2014.08.003>.
- [6] Ng, C. C., Yap, M. H., Costen, N., & Li, B. (2015). Automatic wrinkle detection using hybrid hessian filter. In *Computer Vision--ACCV 2014: 12th Asian Conference on Computer Vision*, Singapore, Singapore, November 1-5, 2014, Revised Selected Papers, Part III 12 (pp. 609-622). Springer International Publishing. https://doi.org/10.1007/978-3-319-16811-1_40.
- [7] Kim, S., Yoon, H., Lee, J., & Yoo, S. (2022, July). Semi-automatic Labeling and Training Strategy for Deep Learning-based Facial Wrinkle Detection. In *2022 IEEE 35th International Symposium on Computer-Based Medical Systems (CBMS)* (pp. 383-388). IEEE. 10.1109/CBMS55023.2022.00075.
- [8] Chen, L. C., Zhu, Y., Papandreou, G., Schroff, F., & Adam, H. (2018). Encoder-decoder with atrous separable convolution for semantic image segmentation. In *Proceedings of the European conference on computer vision (ECCV)* (pp. 801-818). https://openaccess.thecvf.com/content_ECCV_2018/html/Liang-Chieh_Chen_Encoder-Decoder_with_Atrous_ECCV_2018_paper.html.
- [9] Chen, J., Benesty, J., Huang, Y., & Doclo, S. (2006). New insights into the noise reduction Wiener filter. *IEEE Transactions on audio, speech, and language processing*, 14(4), 1218-1234. 10.1109/TSA.2005.860851.
- [10] Lang, X., Zhang, Y., Xie, L., Li, P., Horch, A., & Su, H. (2022). Detrending and Denoising of Industrial Oscillation Data. *IEEE Transactions on Industrial Informatics*, 19(4), 5809-5820. 10.1109/TII.2022.3188844.
- [11] Lei, J., Lang, X., He, B., Liu, S., Tan, H., & Zhang, Y. (2022, July). Ultrasonic Carotid Blood Flow Velocimetry Based on Deep Complex Neural Network. In *2022 IEEE 35th International Symposium on Computer-Based Medical Systems (CBMS)* (pp. 143-148). IEEE. 10.1109/CBMS55023.2022.00032.
- [12] Khan, Z. Y., & Niu, Z. (2021). CNN with depthwise separable convolutions and combined kernels for rating prediction. *Expert Systems with Applications*, 170, 114528. <https://doi.org/10.1016/j.eswa.2020.114528>.
- [13] Pasmينو, D., Aravena, C., Tapia, J. E., & Busch, C. (2023, April). Flickr-PAD: New Face High-Resolution Presentation Attack Detection Database. In *2023 11th International Workshop on Biometrics and Forensics (IWBF)* (pp. 1-6). IEEE. 10.1109/IWBF57495.2023.10157771.