A Comparison of Applying Image Processing and Deep Learning in Acne Region Extraction

Chengrui Zhang, Guangyao Huang, Kai Yao, Mark Leach, and Jie Sun
School of Advanced Technology, Xi'an Jiaotong-Liverpool University, Suzhou, China
Email: {Chengrui.Zhang18, guangyao.huang21, kai.yao19}@student.xjtlu.edu.cn, {Mark.Leach, Jie.Sun}@xjtlu.edu.cn

Kaizhu Huang
Institute of Applied Physical Sciences and Engineering, Duke Kunshan University, Kunshan, China
Email: kaizhu.huang@dukekunshan.edu.cn

Xiaoyun Zhou and Liqiong Yuan
Suzhou Hospital of Traditional Chinese Medicine, Suzhou, China
Email: {445279274, 23422364}@qq.com

***Abstract*—Quantifying acne on face images is considered as a challenging task due to the complex skin surfaces, irregular edges and diverse appearances of acnes. A key in this campaign relies upon segmenting lesion areas precisely in the captured images against varying Imaging situation, e.g., illumination, skin condition, imaging angles and etc. To processing these acne data, either theory-driven image processing methods or data-driven deep learning (DL) based methods are commonly utilized in practice. In order to investigate the advantage and shortcoming of the abovementioned two technology roadmaps in quantifying acne task, we develop an image processing pipeline, and make comparison with the state-of-the-art DL methods such as SegFormer, UNETR, Swin-UNet and TransUNet using small data set. The quantitative comparison results reveal that TransUNet performs better in terms of precision, recall, F1-Score and accuracy, whilst image processing methods still have potential in practice due to its annotation-free.**

***Index Terms*—deep learning, image processing, neural network, image segmentation method**

I. Introduction

Acne is one of the 10 most common diseases worldwide [1], about one out of five teenagers get acne and 85% people from age 12 to 25 have recurrent bouts of acne [2]. It is commonly caused by hair follicle and sebaceous [3], the early manifestations are papules, pustules, cysts, and nodules, and it can cause atrophic or hypertrophic scars in the later stage. Moreover, getting acne can cause mental damage to the youth, especially to female and teen with dark skin [4]. Acne is not a killer, but it can scar people literally and psychologically [5], such as self-confidence and social appearing.

Clinicians and doctors are searching for effective treatment for acne and consequent scars. In the past ten years, Traditional Chinese Medicine (TCM) herbal mask is introduced in Suzhou Hospital of Traditional Chinese Medicine for acne treatment. The process of this treatment is demonstrated in Fig. 1. First, some TCM herbs are selected and grounded into very fine particles, and then they are mixed with distilled water to form mud. Eventually this mud is applied on the patient's face around 20 minutes for regular treatment.

(a) TCM Herb (b) TCM Powder (c) TCM Mud Mask

Figure 1. TCM mask treatment process.

The composition and the proportion of the masks powder are determined by the doctors' experience and their clinical observation. This treatment is becoming very popular since it can effectively reduce melanin synthesis (skin's pigment) thus minimizing chloasma, post-acne dark spots and evens-out imperfections. Different recipes are developed for the treatment and patients may also need customized recipes according to the severity and level of acne status. Thus, developing a quantitative evaluation method using face images for acne status is very necessary. The evaluation results can also be utilized to measure the effectiveness of the mask recipes on individual patients, and regularly track the herbal masks' treatment efficacy. This study can also facilitate the research on mask ingredients, and lead to deeper understandings each ingredients' influence on skin recovery processes.

In the current skin disease research using captured images, two types of methods are commonly used to identify colored spots like acne for quantitative analysis: threshold segmentation, and edge detection. The threshold segmentation method is proposed based on the results of

Manuscript received May 12, 2022; revised September 28, 2022.

doi: 10.18178/joig.10.4.166-171

wavelet transformation homographic filtering [6]. It tends to be difficult, because most of dermatologic diseases has irregular lesion edges and some parts of the unhealthy skin may disappear compared to the bared-eyes [7]. Detecting the dermatologic diseases can be done by choosing a deviation in color to differentiate between the sick skin area and the healthy skin. However, some pixels inside the spot have the same color as the healthy skin, which may be wrongly identified [8]. The edge detection is developed based on the image intensity transformation, and applied to detection of dermatologic diseases by the magnitude of the gradient in the region of interest [9]. This method is sensitive to noise and boundary, which may cause discontinuous contour when extracting unhealthy skin area.

Other than the image processing method as mentioned before, the tremendous growth in the availability and acceptance of image database have made it possible for computer scientists to apply Machine Learning (ML) in medical image processing [10]. Most commonly used technics are Convolutional Neural Network (CNN) [11], [12]. CNN can be trained with imagines directly for end-to-end semantic segmentation [13]. It has been used for skin cancer classification in Stanford university, where the data set includes 129450 clinical images with 2032 different type of disease [14]. The diseases were put into three different categories, which are the benign lesions, malignant lesions and non-neoplastic lesions. The accuracy of CNN is 72.1% which is better compared with the judgement of the two dermatologists. However, such supervised ML method requires a huge amount of labeled image data, which requires tremendous time and effort.

In 2015, a CNN based network called U-Net was introduced for medical image segmentation and has demonstrated good performance when segmenting neuronal structures recorded by electron microscopic in data sets: “PhC-U373” and “DIC-HeLa” [15]. Other than U-Net, another network called transformer is also popular for medical image segmentation. Transformer was originally designed for Natural Language processing, and Dosovitskiy *et al.*, has created a transformer so that this method can be used for image processing tasks [16]. For example, SegFormer uses the transformer as its encoder to create a lightweight network for image applications [17].

Besides, a few DL networks are developed by the combination of U-Net and transformer for medical image segmentation, such as UNet Transformers (UNETR), Swin-UNet and TransUNet [18]-[20].

UNETR has the structure of a U-Net, but acts more like a Transformer. It uses transformers as encoder, but also followed the U-shape design from the U-Net. Then the performance was tested with a data set combined with Multi Atlas Labeling Beyond The Cranial Vault (BTCV) dataset for multi-organ segmentation and the Medical Segmentation Decathlon (MSD) dataset for brain tumor and spleen segmentation tasks with better performance than that of U-Net [18].

The Swin-Unet constructed a U-shape network with Swin Transformer, and performs well in the synapse multi-organ segmentation data set and cardiac diagnosis challenge data set [19].

As a supervised ML algorithm, TransUNet combines the advantage of both Transformers and U-Net. This network was demonstrated excellent performance when testing on 30 abdominal CT scan images in the MICCAI 2015 Multi-Atlas data set [20].

As mentioned above, diverse segmentation methods have achieved excellent performance in different tasks. This study would like to explore their performance in the task of skin acne segmentation. To be more specific, five semantic segmentation methods will be compared in acne detection tasks: a classic image processing method, SegFormer, UNETR, Swin-UNet and TransUNet. Their performance would be quantitatively evaluated and discussed based on the segmentation result’s precision, recall, F1-score and accuracy.

Due to the fact that a mass of people is suffering or have suffered from acne, it is a meaningful task to find a computational method for segmentation of acne region to assist dermatologists in their diagnosis. Even though, the four DL methods are proposed by other people before. Yet, the performance of those methods where never compared by training with a small data set for acne detection. Meanwhile, although the image segmentation of acne has been addressed in several studies and successful applications, there is the potential to develop new methodologies and improve the performance. This paper will compare the result of DL methods and an image processing method to discuss the influence brought by pretrained model in a data set with small size. Moreover, the influence brought by the size of input images and the data set size used to train the pretrain models will also be discussed.

II. Method

In this section, the data set will be introduced and the structure of an image processing method and four DL networks that have been mentioned above will be presented. Their accomplished tasks and the reason that they were chosen for this task will be discussed as well. Each technique that has been applied in the networks will be identified. The strengths and weaknesses of a network will be evaluated based on the techniques they used and the overall performance.

A. Datasets Development

The image database consists of 109 photos, which are taken by the Suzhou Hospital of Traditional Chinese Medicine with a DSC-W180 digital camera. These images are taken from 46 patients with different periods of treatment from one to eight weeks. Before using any algorithms, all the images have been cropped into a size of 512×512 pixels, so that the following methods can focus on the facial regions of patients.

B. Image Processing

For the image processing method, it requires some pre-processing before starting to detect acne spots. First, the images need to be transformed into YCrCb form. Then to identify acne position, other four image processing steps are involved as Fig. 2: homomorphic filtering, face segmentation, facial smoothing and acne segmentation.

Homomorphic Filtering is applied to eliminate the influence brought by light source, which can compress the illumination and enhance the difference between sick skin and healthy skin [21]. In Face Segmentation, face-parsing is used to prepare skin area [22], by the well-trained BiSeNet model [23]. This model can extract the eyes, nose, mouth and eyebrows, and then remove them to extract facial skin for future processing. After that a guided filter is applied to smooth the skin to preserve edge, which is similar to the functions of bilateral filter with better edge contour [24]. Eventually, the acne region is detected by subtracting the facial images with the smoothed skin, and the edge of skin lesion is obtained using a threshold. Of course, this acne segmentation results are obviously sensitive by the selection of threshold.

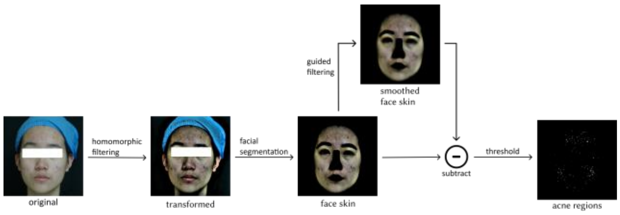

Figure 2. Image processing pipeline.

C. UNETR

UNETR is a network that combines transformer with the "U-Shaped" network. The UNETR is built with a transformer encoder and a CNN decoder equipped with skip connection. The input images of the UNETR will be patch-embedded with linear layers and then passed to a stack of transformer blocks. The decoder part of the UNETR works similar to U-Net [18]. The parameters in UNETR are not pretrained, therefore they were initialized randomly. This method is applied in this paper to prove the assumption that a pretraining model can improve the performance of DL network when there are limited data.

D. SegFormer

SegFormer is a lightweight framework within the transformer family. It is designed for handling semantic segmentation task to deal with large size image, which takes advantages of efficient self-attention module, efficient multi-head self-attention module, mixed feed-forawrd network and overlapped path merging layers.

This method is chosen because the image with size of 512×512 can be directly input to SegFormer without being divided into small patches with size of 224×224 like other transformer-based methods [17]. This network significantly saves the computation workload when processing large size image. Additionally, by comparing the performance of SegFormer with the other networks the benefit of intake a larger sized image set will be shown.

E. Swin-UNet

The Swin-UNet is a "U-shaped" network built with Swin Transformers. Similar to the UNETR, this method is built by a Swin Transformer encoder and a CNN decoder [19]. The Swin Transformer has achieved great success in general computer vision tasks. Thus, it has been introduced into UNet for medical image tasks. The method was pretrained using the imagenet-1k data set to improve the performance of transfer learning on relatively small data set. Due to the small number of images available in our data set, Swin-UNet has been chosen to be applied in this task.

F. TransUNet

TransUNet is one of the state-of-the-art semantic segmentation methods for biomedical images in UNet-like family. Taking advantages of inductive bias of CNN, a hybrid vision transformer pretrained on ImageNet-21k is engaged to extract local meaningful features, as shown in Fig. 3. Then, the features are embedded into tokens from spatial dimension, followed by a 12-layer Transformer to model the global dependencies via Multi-head Self-Attention (MSA). Finally, the processed high-level features are then up-sampled and fused with multi-level low-level features extracted from CNN backbone through skip connection to further boost the performance [20].

As mentioned before, TransUNet used pretraining model for the network. In addition, it was trained on a dataset 14 million data which is larger than the models used for SegFormer and Swin-UNet. Thus, the pretrained model in TransUNet should have better performance than the models in the other two networks. It is chosen to be the best existing network for acne segmentation.

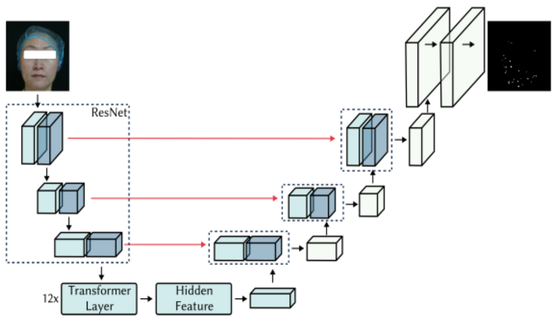

Figure 3. The overall of TransUNet.

III. Experimental Result

In the following sections, discussion will be based on the segmentation experimental results and applicability of these five methods. One of the purposes for this paper is to demonstrate how well DL methods are compared to image processing method. Thus, the performance of the image processing will be discussed first. Then the performance of each DL method will be compared. The results will be evaluated both quantitatively and qualitatively. The predicted regions obtained for the acnes are compared with ground truths annotated manually. The performance of the segmentation is determined by the training models and networks used to solve the problem. The quantitative evaluation will be done by using precision value, recall value, F1-score and accuracy rate in Table I. The qualitative performance evaluation will be based on the segmentation result in Fig. 4 and the result in highlighted boxes will be discussed in detail. Some assumptions are made before looking at the results: the image processing method should have poorest performance because it is not a DL network, UNETR might have the worst segmentation result due to the small data set, Segformer will outperform Swin-UNet and TransUNet might provide the best segmentation. The assumptions will be proved or disapproved by look at the quantitative result and qualitative result.

A. Results from Image Processing Method

The results of each step in the image processing method can be seen in the Fig. 2. The original image is the image from data base with background cropped. Then after applying homomorphic filtering to the images, the light source influence was reduced and the difference between healthy skin and unhealthy skin is enhanced. Normally the image and the skin area should become brighter after the filtering. While, the image after the filtering becomes darker as shown in the Fig. 2, due to the black background. Regardless of brighter or darker, the illumination are both compressed and the difference between healthy skin and acne region are enhanced. Only skin area is extracted after facial segmentation, and then the images will be smoothed using guided filtering. In the last step the acne region will be obtained by subtracting segmented image with the smoothed image.

B. Quantitative Performance Evaluation

Additionally, the detected regions results are measured using statistical metrics. To be specific, the performances are evaluated using precision, recall, F1-score and accuracy rate. Those values were calculated using the following formulars: $precision = \frac{n_{TP}}{n_{TP}+n_{FP}}$, $recall = \frac{n_{TP}}{n_{TP}+n_{FN}}$, $F1 = \frac{2\times precision\times recall}{precision+recall}$, $accuracy = \frac{n_{TP}+n_{TN}}{n_{TP}+n_{TN}+n_{FP}+n_{FN}}$, where n_{TP}, n_{TN}, n_{FP} and n_{FN} are defined to be the number of true-positive, true-negative, false-positive and false-negative segmentation result pixels in an image, respectively. These values are determined based on the label of each pixel. The pixel is called positive if the labeled or detected as unhealthy, else it would be called negative. If the pixel is labeled as unhealthy and detected as unhealthy then is it true-positive. When a pixel is detected as healthy and the label for that pixel is also healthy then it is noted as true-negative. The pixel is false-negative when it is labeled healthy but detected as unhealthy. False-positive refers to pixels labeled unhealthy but detected as healthy. Precision, recall and accuracy are calculated based on n_{TP}, n_{TN}, n_{FP} and n_{FN}. A higher F1 score indicates a better intersection between the ground truth and the predicted segmentation masks.

The comparative results of the five methods are shown in Table I. As observed, all DL methods significantly outperform the image processing method which only got 0.3 w.r.t. F1-Score on this dataset. The intuition behind is that the features of acne are very difficult to be extracted with only intensity thresholding. On the other hand, compared with other methods pretrained on large-scale dataset, UNETR only attains 0.4540 and 0.5443 in terms of precision and recall, indicating the importance of pretraining on small dataset. Swin-Unet and SegFormer obtain F1-score of 0.5971 and 0.6075 respectively and the difference between their accuracy is only 0.0001. Thus, according to F1-score and accuracy it is hard to tell a difference between two. Yet, Swin-Unet have higher recall, and SegFormer have higher precision. It shows that the receptive field of methods may be a trade-off factor between precision and recall. According to the functions, a higher precision indicates lower false positive value and higher recall value means lower false negative value. Therefore, SegFormer is better at detecting acne regions and Swin-UNet is better at segmenting healthy regions in this task. With the largest datasets pretrained and most trainable parameters, TransUnet achieved the state-of-the-art performance on this dataset, demonstrating the superior of this method on acne segmentation.

TABLE I. QUANTITATIVE COMPARISON

Method	Precision	Recall	F1-Score	Accuracy
Image Processing	0.3456	0.3120	0.3008	0.9859
UNETR	0.4540	0.5443	0.4951	0.9897
Swin-UNet	0.5946	0.5996	0.5971	0.9911
SegFormer	0.6271	0.5891	0.6075	0.9910
TransUNet	**0.6442**	**0.6039**	**0.6121**	**0.9920**

C. Qualitative Performance Evaluation

In the four supervised learning methods i.e., SegFormer, UNETR, Swin-UNet and TransUNet, no image preprocessing is required. While, from the 109 labeled image data set, 79 images are used as the training set and 30 images are used for testing.

In Fig. 4, the results of five methods are shown by segmenting four patients' images. Patient 1 has the most acne regions centered at her chain, patient 2 and patient 4 have few acne regions and patient 3 has acne regions on his full head and cheeks. It is obvious that the image processing method has the worst result and TransUNet is the best method for this segmentation task. As observed, the image processing methods tend to detect smaller acne regions than the actual region and the detected regions appears to have irregular lesion edges.

As for the DL methods, by looking at the segmentation result showing in the boxed regions for patient 1 and patient 4, it can be said that the TransUNet is the closest to the ground truth, SegFormer and Swin-UNet provides similar result and UNETR is not sensitive to small object segmentation. Although many acne regions are identified by UNETR, it failed to detect small acne regions as shown in the result for patient 4.

By looking at the boxed results for patient 2 and patient 3, TransUNet has shown outstanding results. The image for patient 3 has a nevus on his neck and all the other networks has detected it as an acne region but not TransUNet. While, several acne regions are correctly detected by SegFormer, UNETR and Swin-UNet. It's difficult for them to segmenting regions with small contrast amount the edges with low contrast. This shows that TransUNet is good at avoiding noises. Then by looking at the boxed area for patient 2, TransUNet has detected an acne region that failed to be labeled in ground truth. Therefore, TransUNet can provide result exciting the performance of mankind.

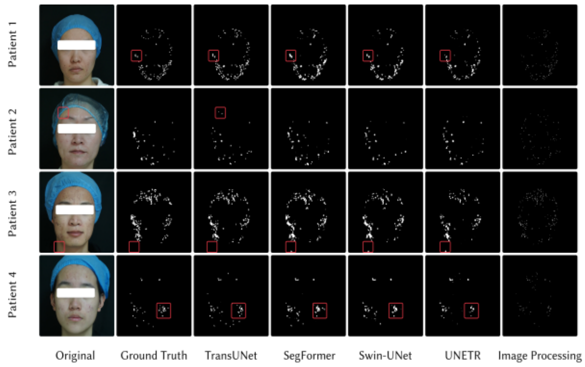

Figure 4. Qualitative results.

IV. Discussion and Future Work

There is an increasing interest on the prevention and early diagnosis of skin disease. Amount all skin disease relative tasks acne segmentation is an important step for the effective computational diagnosis. Although the image segmentation has been addressed in several studies and successful applications. There is a lot of potential to develop new methodologies and improve the performance.

Different techniques were introduced to extract acne regions, with a focus on their subsequent segmentation. From the current studies, one may conclude that acne segmentation is difficult due to irregular edges and small sized regions.

Overall, DL methods are superior to the proposed image processing methods, whilst the methods with larger size of pretraining data perform better. In fact, image processing method is annotation-free, which significantly saves labor and time compared with DL methods. However, due to the poor performance it provides, even if this method can save massive labor and time, the image processing method is unlikely to be chosen for acne region segmentation tasks. Even though that pretrained models can improve the performance of networks, training the proposed networks still requires humongous amount of data. Since the data set used in this paper is small, it is not expected for the DL methods to reach performance saturation. Therefore, more labeled data are required to improve the performance. Meanwhile, DL method achieves relatively successful segmentation comparing to image processing methods.

In future works, a more accurate labeled ground truth is essential. Labeling by experienced doctors is required and methods to identify ground truth need to be discussed. On another hand, considering extremely expensive labeling costs and miss labeling, a semi-supervised DL method needs to be proposed thus take full use of unlabeled data. Meanwhile, harder tasks including multi-class semantic segmentation and object detection shall be involved and further investigated. Furthermore, the combination of different methods would be explored to improve the image segmentation process, such as finding the approximate location of acne, and automatically defining the initial regions.

V. Conclusion

To have the acne spots clearly and accurately outlined is a key for the dermatological application such as diagnosis of acnes. we have introduced five methods to extract acne regions, with a focus on comparison of their subsequent segmentation results. TransUNET is generally an optimal solution, with good edge-based segmentation capabilities, its ability to solve local minima or overlapping problems, its computational efficiency, and its excellent performance in detecting acne borders in facial images. Moreover, a pretrained model will improve the performance of a network on tasks with small size data set and such model with larger data set can provide even better result. This paper also proved that a network that can take larger size image with high resolution is expected to segment small object such as acnes.

As mentioned above, not only the networks for acne segmentation, but also the techniques employed in each method are explained, with their strengths and weaknesses. On top of that, better DL networks can be constructed to develop unsupervised or semi-supervised algorithms with accurate region detection, as well as to take into account computational performance, automaticity level, and image noise smoothing and removal.

Conflict of Interest

The authors declare no conflict of interest.

Author Contributions

Xiaoyun Zhou and iqiong Yuan collected and provided the dataset. Chengrui Zhang and Kai Yao designed the overall experimental plan and performed experiments. Guangyao Huang interpreted data and wrote the manuscript with support from Mark Leach and Kaizhu Huang. Jie Sun supervised the project and conceived the original idea. All authors read and approved the manuscript.

Acknowledgment

The work was partially supported by the following: Municipal Key Lab for Wireless Broadband Access Technologies in the Department of Electrical Engineering, Xi'an Jiaotong Liverpool University, for research facilities; Key Program Special Fund in Xi'an Jiaotong-Liverpool University under Grant KSF-E-37.

References

[1] S. M. Tuchayi, E. Makrantonaki, R. Ganceviciene, C. Dessinioti, S. R. Feldman, and C. C. Zouboulis, "Acne vulgaris," *Nature Reviews Disease Primers*, vol. 1, pp. 1-20, September 2015.
[2] J. Zhao, Y. Wang, L. Jiang, and Y. Mu, "The application of skin care product in acne treatment," *Dermatologic Therapy*, vol. 33, no. 6, September 2020.
[3] K. Bhate and H. Williams, "Epidemiology of acne vulgaris," *British Journal of Dermatology*, vol. 168, no. 3, pp. 474-485, December 2013.

[4] M. N. Natsuaki and T. M. Yates, "Adolescent acne and disparities in mental health," *Child Development Perspectives*, vol. 15, no. 1, 37-43, January 2021.

[5] J. E. Brody. (January 2019). Managing teenage acne. *The New York Times*. p. 5. [Online]. Available: https://www.nytimes.com/2019/01/07/well/live/manag-ing-teenage-acne.html

[6] C. Xu, F. Huang, and Z. Mao, "An improved two-dimensional Otsu threshold segmentation algorithm," *Electronic Technology Applications*, vol. 42, no. 12, pp. 108-111, 2016.

[7] M. Ebner, "A parallel algorithm for color constancy," *Journal of Parallel and Distributed Computing*, vol. 64, no. 1, pp. 79-88, 2004.

[8] D. F. Graham, S. Bernt, and L. C. James, "Comprehensive colour image normalization," in *Proc. European Conference on Computer Vision*, 1998, pp. 475-490.

[9] Y. C. Liu, W. H. Chan, and Y. Q. Chen, "Automatic white balance for digital still camera," *IEEE Transactions on Consumer Electronics*, vol. 41, no. 3, pp. 460-466, 2004.

[10] S. Jain, V. Jagtap, and N. Pise, "Computer aided melanoma skin cancer detection using image processing," *Procedia Computer Science*, vol. 48, pp. 735-740, 2015.

[11] K. Yao, Z. Su, K. Huang, X. Yang, J. Sun, A. Hussain, and F. Coenen, "A novel 3D unsupervised domain adaptation framework for cross-modality medical image segmentation," *IEEE Journal of Biomedical and Health Informatics*, 2022.

[12] K. Yao, J. Sun, K. Huang, L. Jing, H. Liu, D. Huang, and C. Jude, "Analyzing cell-scaffold interaction through unsupervised 3d nuclei segmentation," *International Journal of Bioprinting*, vol. 8, no. 1, 2022.

[13] L. Alzubaidi, *et al.*, "Review of deep-learning: Concepts, CNN architectures, challenges, applications, future directions," *Journal of Big Data*, vol. 8, no. 1, 2021.

[14] E. Andre, *et al.*, "Dermatologist-level classification of skin cancer with deep neural networks," *Nature*, vol. 542, pp. 115-118, 2017.

[15] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *Proc. International Conference on Medical Image Computing and Computer-Assisted Intervention*, October 2015.

[16] A. Dosovitskiy, *et al.*, "An image is worth 16×16 words: Transformers for image recognition at scale," arXiv preprint arXiv:2010.11929, June 2020.

[17] E. Xie, W. Wang, Z. Yu, A. Anandkumar, J. M. Alvarez, and P. Luo, "SegFormer: Simple and efficient design for semantic segmentation with transformers," *Advances in Neural Information Processing Systems*, vol. 34, October 2021.

[18] A. Hatamizadeh, *et al.*, "Unetr: Transformers for 3d medical image segmentation," in *Proc. the IEEE/CVF Winter Conference on Applications of Computer Vision*, 2022, pp. 574-584.

[19] H. Cao, Y. Wang, J. Chen, D. Jiang, X. Zhang, Q. Tian, and M. Wang, "Swin-unet: Unet-like pure transformer for medical image segmentation," arXiv preprint arXiv:2105.05537, May 2021.

[20] J. Chen, Y. Lu, Q. Yu, X. Luo, E. Adeli, Y. Wang, and Y. Zhou, "Transunet: Transformers make strong encoders for medical image segmentation," arXiv preprint arXiv:2102.04306, 2021.

[21] C. G. Rafael and E. W. Richard, *Digital Image Processing*, New York, NY: Pearson, 2018.

[22] Zllrunning. (March 2022). GitHub - zllrunning/face-parsing. PyTorch: Using modified BiSeNet for face parsing in PyTorch. [Online]. Available: https://github.com/zllrunning/face-parsing.PyTorch

[23] C. Yu, *et al.*, "Bisenet v2: Bilateral network with guided aggregation for real-time semantic segmentation," *International Journal of Computer Vision*, vol. 129, no. 11, pp. 3051-3068, April 2021.

[24] K. M. He, J. Sun, and X. O. Tang, "Guided image filtering," in *Proc. European Conference on Computer Vision*, 2010.

Chengrui Zhang was born in Zhengzhou, Henan, China in 2000. He has a Bachelor of Engineer degree in Computer Science and Technology at Xi'an Jiao-Tong Liverpool University (XJTLU). He is a Ph.D. candidate at XJTLU now. His research interests include computer vision in medical image, 3D face reconstruction, and software development.

Guangyao Huang was born in China in 1996. He has a Bachelor of Science degree in Mathematics at University of New Hampshire (UNH). He is now earning his Master of research degree in computer since at Xi'an Jiaotong-Liverpool University (XJTLU). He worked as a research assistant in City College of New York (CCNY) at Bingmei Fu's lab. Currently, he is a research assistant in XJTLU at Jie Sun's lab. His research interests include computer vision, deep learning, and software development.