Shadow Detection and Removal from Hand Imagesusing Synthetic Dataset

Shi-Jinn Horng
Asia University
Taichung, Taiwan
horngsj@yahoo.com.tw

Minh-Tuong Le
National Taiwan Univ. of Sci. & Tech.
Taipei, Taiwan
leminhtuong271@gmail.com

Dinh-Trung Vu
National Taiwan Univ. of Sci. & Tech.
Taipei, Taiwan
trungvd@vimaru.edu.vn

Thi-Van Nguyen
National Taiwan Univ. of Sci. & Tech.
Taipei, Taiwan
vannt.cnt@vimaru.edu.vn

***Abstract*—There is a significant demand for biometric accessin terms of authentication recently. In addition to the fingerprint recognition system, contactless hand-based biometric systems,such as palm vein recognition, have emerged as notable approaches. However, palm vein recognition systems face numerous challenges in real-world environments, one of which is shadows. Shadows can disrupt the palm vein structure and lead to incorrect recognition. In this project, our goal is to enhance palm vein images by addressing the issue of shadows on hand.To achieve this, we carried out experiments involving the ST- CGAN and SP+M Net in addition to a hand extraction modulefor shadow detection and shadow removal specifically on thehand. In addition, rather than using the unchanged shadow mask for model training, we introduce a random shape generator to produce the random shadow mask. Furthermore, we also presenta method for generating a new synthetic dataset during each training iteration, which aims to enrich the dataset and improve the training process.**

***Keywords*—shadow detection, shadow removal, synthetic dataset, biometric access, contactless hand based system, deep learning**

I. Introduction

The palm vein recognition system is a notable approach in biometric authentication that finds good applications. However, it often faces challenges in real-world environments which leads to variations in hand portrait and decreased recognition accuracy. One particular challenge is the presence of shadows on hand. Shadows can introduce noise into the information and disrupt the color distribution on the hand, making it difficult to extract and recognize palm veins accurately. Figure 1 describes the influence of shadows on vein structure, which contributes to the complexity of palm vein recognition; however, removing the shadows can recover vein structure, thereby enhancing the palm vein recognition system.

Compared to shadow detection on other objects, shadow detection on the hand can take advantage of two distinct characteristics, which are hand color and hand shape. Firstly, the color of the hand, typically human skin color, is unique and differs from the background. Secondly, the hand's geometry is also specific, although it may be distorted due to camera perspective. To leverage these characteristics for shadow removal, we should develop a component to extract this feature.

Based on the aforementioned two features, we aim to construct a hand extraction module that leverage these distinctive hand characteristics. Additionally, we conducted a research on notable shadow detection and shadow removal models [2]–[4] for this specific task. Furthermore, we introduce a method to regenerate a synthetic dataset during each training iteration,

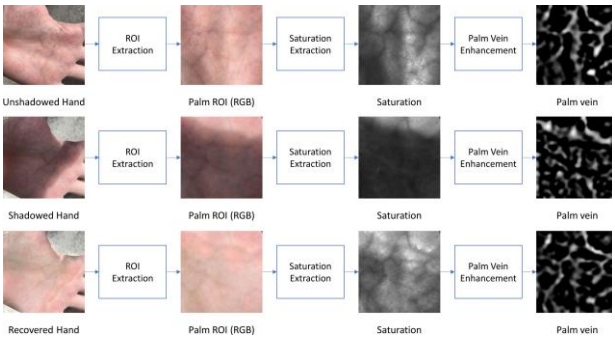

Fig. 1. The influence of shadows and the effectiveness of palm vein recovery through shadow removal on palm vein recognition.

and instead of relying on the unchanged shadow mask, we implement a random shape generator to generate the shadow mask. This innovative approach allows us to generate a fresh dataset and create a new shadow mask during each training iteration.

Our contributions can be summarized as follows:

- We conducted experiments on shadow detection and shadow removal methods for a new research field which is removing shadows specifically on the hand. Subse- quently, we evaluated trained models using the NTUST- HS test dataset, which was collected by us and specifi- cally designed to assess the influence of shadow effects on the hand.
- Rather than using a stable traing dataset consisting of shadow full images, shadow-free images (ground truth), and shadow masks, we presented a training approach that generates a synthetic training dataset within each training loop
- We introduced a random shape generator based on the Bezier curve to create a dynamically changing shadow mask in each training iteration instead of utilizing an unchanged shadow mask as in previous research.

II. Related Work

Regarding traditional approaches, there were illumination invariant methods proposed by Finlayson [6], [7] for shadow removal. Following that, Gong et al. [8] proposed a method to remove shadow using two input images. Then, Gryka et al. [9] introduced a learning-based approach to shadow removal, which utilized strokes provided by the user.

Deep learning techniques have rapidly advanced and demonstrated superior performance compared to other methods. One of the early papers applying deep learning to shadow removal was DeshadowNet by Qu et al. [10]. This method

employed a neural network to calculate a shadow free image from an input shadow image. Hu et al. [11] developed direction-aware spatial context (DSC) module, utilizing a spatial Recurrent Neural Network (RNN) model [12], which is an early application of RNN to object detection within a context.

Since the introduction of Generative Adversarial Network (GAN) [13], it has gained significant attention in the deep learning field. Wang et al. proposed ST-CGAN [4], which detects and removes shadow simultaneously. Following that, Mask-ShadowGAN [14] and ARGAN [15] applied generative models to the task of shadow removal in natural images. Later on, SP+M+I Net [3] is an improvement of SP+M Net [2], utilizes linear illumination transformation to model the shadow effects in the image and remove the shadows by combining the shadow-free image, the shadow parameters, and a matte layer.

III. Shadow Detection And Shadow Removal Framework

The shadow detection and removal framework for eliminating shadows on hand images consists of three main components: hand extraction, shadow detection, and shadow removal. The input to the framework is a hand image with shadows, and the output is a shadow-free image.

A. Hand Extraction

This stage takes the shadow-full image as input and extract the hand shape. The output is the hand extraction, which is used for both shadow detection and shadow removal.

Hand extraction is a part of skin detection, which has been extensively studied in the past. For the specific task of hand detection, we prefer to use HSV-YCbCr thresholding strategy since we can prioritize computational efficiency and cost-effectiveness. To fulfill these requirements, we have chosen a skin detection method based on the HSV and YCbCr color spaces [5]:

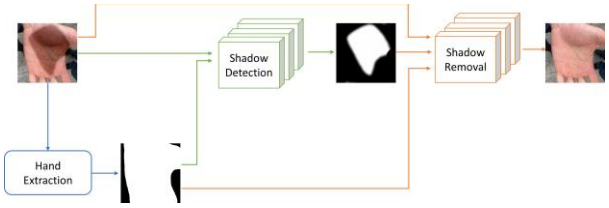

Fig. 2. Framework of removing shadow on hand method

$$(0 \leq H \leq 5) \wedge (0.23 \leq S \leq 0.68) \wedge \\ \wedge\ (85 \leq Cb \leq 135) \wedge (135 \leq Cr \leq 180) \quad (1)$$

To adapt the method to the specific task at hand, we modified certain parameters of this HSV-YCbCr thresholding strategy. Specifically, we utilized different kernel sizes, changing from (5,5) to (9,9), and we replaced the morphological opening function at the end of the output with a dilation function. With these modifications, we aimed to capture more spatial information in order to improve the performance of the hand extraction.

B. Shadow Detection

At this stage, the input are the image containing shadows and the hand mask. The output of shadow detection is a binary shadow mask, representing the regions of the image affected by shadows. We consider the shadow detection part of ST-CGAN [4] and incorporate an additional hand mask into the architecture for this task. Figure 3 describes ST-CGAN architecture incorporating an additional hand mask for the specific task of removing shadows on the hand.

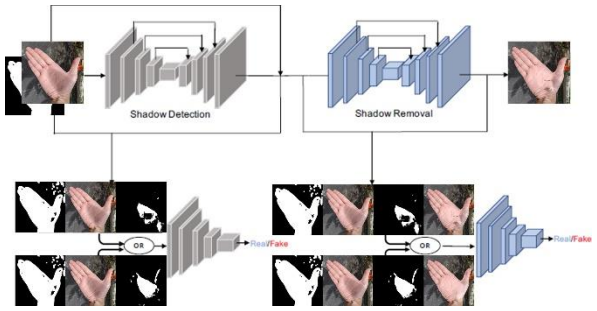

Fig. 3. Architecture of Stacked Conditional Generative Adversarial Network (STCGAN) incorporating an additional hand mask.

ST-CGAN consists of two parts: a shadow detector and a shadow removal module. Each part is implemented using a Generative Adversarial Network (GAN) architecture, which includes a generator and a discriminator. The generator in both parts is constructed based on a U-Net structure.

For the task of shadow detection in ST-CGAN, the output image is a binary shadow mask. The U-Net architecture is employed for segmenting the shadowed and non-shadowed regions. Compared to the originally proposed [4], we made a few changes in the architecture. We increased the depth of the U-Net within the generator, ensuring that the convolutions reach a minimum size of 1x1. Additionally, we also changed the discriminator type to PatchGAN, which has shown better performance as detailed in [3].

C. Shadow Removal

The next step in the process involves shadow removal. Typically, the input to this model consists of shadow-full image, the predicted shadow mask and the predicted hand mask, while the output is a shadow-free image. We consider SP+M Net [2] for shadow removal task and enhance the architectures by including an additional hand mask for this task.

The SP+M Net [2] is a shadow removal method that utilizes a physical illumination model and image decomposition. The main idea behind this approach is to illuminate the shadow in a shadow image. The illumination model, referred to as SP Net, is established using a physical formula [16]. The parameters calculated by SP-Net are then used to perform a linear transformation on the image, effectively converting the shadow into a lit version of the shadow image. On the other hand, the M-Net model computes the shadow matte based on the relit image, input image, and shadow mask. The shadow matte essentially consists of a shadow mask with the color depth. Finally, the shadow-free image is generated by combining the shadow matte, shadow mask, and the input shadow image.

Figure 4 describes the architecture SP+M Net including an additional hand mask as our proposal.

IV. Training With The Synthetic Dataset

These experiments were conducted on two NVIDIA GeForce RTX 2080 Ti 11GB GPUs.

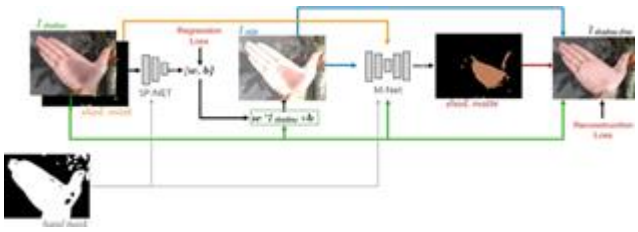

Fig. 4. Architectures of SP+M Net incorporating an additional hand mask.

A. *Method For Creating Synthetic Dataset*

The training flow in this scenario involves generating a new dataset in each training loop. To generate this new dataset, the training flow requires a group of raw datasets. In this case, the group consists of the background images provided by the GTEA dataset [17] (36,500 background images), hand images provided by The 11K Hands database [18] (11,076 images) and NTUST-IP gathered by us (11,220 images), and shadow images generated by the random shape generator. In each training loop, we generate a new training dataset using the flow as shown in Figure 5.

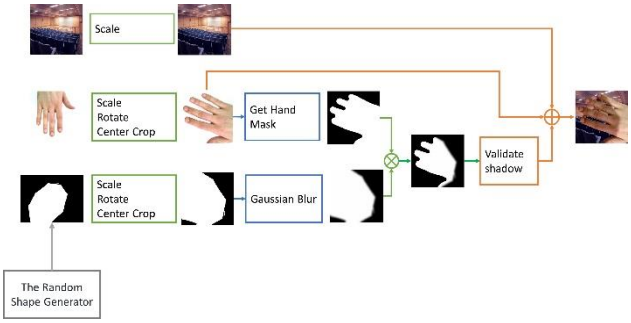

Fig. 5. Approach for generating the synthetic training dataset every training loop.

After loading the data from the database and getting random shadow mask from random shape generator, we perform several pre-processing steps, including scaling and rotating the images. Additionally, we extract the hand mask from the loaded data. To prepare the shadow mask, we apply a blur operation to ensure a smoother representation of the shadow. The hand mask and the blurred shadow mask are then combined together to obtain the final shadow mask. To validate the shadow mask quality, we examine the percentage of shadow pixels within the hand mask region. The criteria demand a range of 10% to 90% to ensure adequate coverage without being overly dominant or insufficient. Finally, we combine the validated shadow mask with the background image to create the shadow-full image. This process involves overlaying the shadow mask on the background, effectively applying the shadow effect to the hand region.

Fig. 6. Samples from a synthetic dataset.

As the synthetic dataset is generated based on hand images, its length corresponds to the combined number of images in The 11K Hands database [18] and The NTUST-IP database, resulting in a total of 22,296 images. With each training loop, a new synthetic dataset is created, presenting a fresh set of training samples for the model. Figure 6 displays some examples of the generated synthetic images.

B. *The Random Shape Generator*

The idea of generating a random shape is inspired from the principles of the Bezier curve [19]. A Bezier curve is formed by a collection of control points. The arrangement and quantity of these points will define the shape of the Bezier curve. Figure 7 illustrates the variety of Bezier curve shapes achieved by altering the number and position of control points.

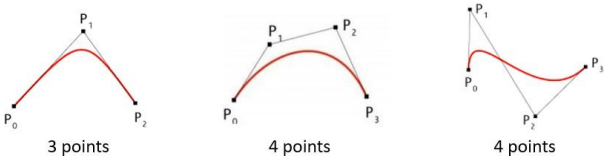

Fig. 7. Examples of the Bezier curve by varying location and quantity of control points

The approach to generating a random shape involves creating a set of random points and utilizing them as control points to construct the Bezier curve. In this particular case, we will maintain a fixed number of control points, specifically 7 points, and randomize positions of these 7 control points. By varying the positions of the control points, we can achieve a diverse range of shapes for the random curves. Figure 8 shows some random shapes generated by the random shape generator, which are consider as shadow masks.

Fig. 8. Samples from generating shadow mask.

V. Experiment Result

The experiments were conducted on the NVIDIA GeForce RTX 2080 Ti 11GB GPU.

A. *The result of trained models*

In this experiment, we intend to perform the experiments involving SP+M Net [2], SP+M+I Net [3] and the shadow removal component of ST-CGAN [4] for shadow removal, while the shadow detection component of ST-CGAN [4] is considered for shadow detection, as described in section III-B. Additionally, we want to compare the performance of models when applying hand extraction and when not applying hand extraction. Table I shows the results of trained models on a synthetic dataset. The synthetic dataset used for evaluation is generated randomly using the method described in Figure 5.

As observed from Table I, the results indicate that applying hand extraction did yield better outcomes. Among the eval- uated methods, SP+M Net with the hand extraction module achieved the most favorable result in term of PNSR and SSIM.

B. *Evaluation on NTUST-HS test dataset*

The NTUST-HS dataset is a collection of contactless RGB palm images from NTUST (National Taiwan University of Science and Technology). This database focuses on palm images that are affected by shadows. It comprises 1600 shadow images, with 800 images captured in an outdoor

TABLE I RESULT OF MODELS ON A GENERATED SYNTHETIC DATASET

		Synthetic Dataset	
		PNSR	**SSIM**
Input shadow image		12.6877	0.7768
ST-CGAN	Without Hand Extraction [4]	25.0984	0.9158
	With Hand Extraction	25.1892	0.9170
SP+M+I Net	Without Hand Extraction [3]	21.7376	0.8948
	With Hand Extraction	21.7464	0.8954
SP+M Net	Without Hand Extraction [2]	25.3956	0.9259
	With Hand Extraction	**25.3997**	**0.9261**

Fig. 9. Some samples from NTUST-HS database.

scenario and 800 images in an indoor scenario. Figure 9 shows some samples from the database.

We evaluate the models on the NTUST-HS test dataset in two scenarios. In the first scenario, we compare the palm region-of-interest (ROI) images which are color images before and after shadow removal using the PSNR and SSIM methods. In the second scenario, we compare the vein extraction images which are onen-channel images before and after shadow removal using the dice coefficient and SSIM methods. Result of these two scenarios are shown in Table II. The palm ROI and palm vein images will be obtained using the extraction method described in [1].

As can be seen from Table II, the models effectively enhance the palm ROI and palm vein images. The higher SSIM indicates a greater similarity of the recovered images and the ground truth in terms of structural information, luminance, and contrast; while the higher PSNR indicates better quality of the reconstructed images, and the higher dice coefficient denotes a greater similarity in terms of overlap between two images. In this case, the SP+M Net incorporating an additional hand mask showcases the better performance among the considered models. Throughout the results, we can observe that the influence of shadows on palm images in outdoor scenarios is more challenging than in indoor scenarios.

In Figures 10, we show examples of shadow removal on hand using methods in this project. These examples include images from different individuals in various scenarios, outdoors and indoors, and captured from left and right hands.

VI. CONCLUSION

We conducted experiments on shadow detection and removal methods tailored specifically for hands. Our training process involved generating a synthetic training dataset in each loop. Besides, instead of using the unchanged shadow mask to train models as in previous related works, we introduced a random shape generator based on the Bezier curve to create a dynamically changing shadow mask in each training iteration. Based on the experiment results, removing shadows on the hand proves to be useful for reducing noise and enhancing the clarity of hand portraits. This enhancement strengthens the capabilities of contactless hand-based biometric systems, which find applications in various fields such as biometric authentication for online payments, security systems, adminis- trative identification, and class attendance management... Fur thermore, applying the hand extraction module demonstrates better results compared to not applying the hand extraction module.

In addition, the network occasionally produces incorrect results, particularly when it fails to detect shadows in images containing complex shadow patterns or extremely dark shadows. To address this issue, further development of the model will involve increasing the diversity of the dataset, particularly with a variety of shadow masks in terms of color and shapes.

REFERENCES

[1] S. -J. Horng, D. -T. Vu, T. -V. Nguyen, W. Zhou, and C. -T. Lin, "Recognizing Palm Vein in Smartphones Using RGB Images," IEEE Transactions on Industrial Informatics, vol. 18, no. 9, pp. 5992-6002, Sept. 2022. doi: 10.1109/TII.2021.3134016.

[2] H. Le and D. Samaras, "Shadow removal via shadow image decomposition," in Proc. ICCV, 2019.

[3] H. Le and D. Samaras, "Physics-based shadow image decomposition for shadow removal," IEEE Trans. Pattern Anal. Mach. Intell. (2021).

[4] J. Wang, X. Li, and J. Yang, "Stacked conditional generative adversarial networks for jointly learning shadow detection and shadow removal," Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2018.

[5] D. Dahmani, M. Cheref, and S. Larabi, "Zero-sum game theory model for segmenting skin regions," Image and Vision Computing, Volume 99, 2020, 103925, ISSN 0262-8856, doi: 10.1016/j.imavis.2020.103925.

TABLE II. EVALUATE PALM ROI IMAGES AND PALM VEIN IMAGES OF METHODS BY PNSR AND SSIM.

		Both		**Indoors**		**Outdoors**	
Scenario 1: Evaluate Palm ROI images (color image)							
		PNSR	**SSIM**	**PNSR**	**SSIM**	**PNSR**	**SSIM**
Input shadow image		15.949	0.7601	16.416	0.7848	15.483	0.7353
STCGAN	Without Hand Extraction [4]	19.440	0.7840	20.277	0.8136	18.603	0.7544
	With Hand Extraction	19.671	0.7915	20.262	0.8172	19.081	0.7657
SP+M+I Net	Without Hand Extraction [3]	19.560	0.7901	20.635	0.8224	18.485	0.7579
	With Hand Extraction	19.885	0.7952	**20.737**	0.8255	19.034	0.7650
SP+M Net	Without Hand Extraction [2]	19.649	0.7947	20.372	0.8235	18.926	0.7658
	With Hand Extraction	**19.932**	**0.7977**	20.710	**0.8273**	**19.155**	**0.7680**
Scenario 2: Evaluate Palm Vein images (one-channel image)							
		Dice	**SSIM**	**Dice**	**SSIM**	**Dice**	**SSIM**
Input shadow image		0.1809	0.2856	0.1825	0.2944	0.1792	0.2768
STCGAN	Without Hand Extraction [4]	0.1790	0.2737	0.1801	0.2797	0.1778	0.2678
	With Hand Extraction	0.1762	0.2737	0.1761	0.2769	0.1763	0.2705
SP+M+I Net	Without Hand Extraction [3]	0.1814	0.2817	0.1829	0.2907	0.1800	0.2727
	With Hand Extraction	0.1819	0.2826	0.1829	0.2889	0.1809	0.2762
SP+M Net	Without Hand Extraction [2]	0.1847	0.2887	0.1860	0.2956	0.1834	0.2817
	With Hand Extraction	**0.1866**	**0.2914**	**0.1882**	**0.2995**	**0.1850**	**0.2832**

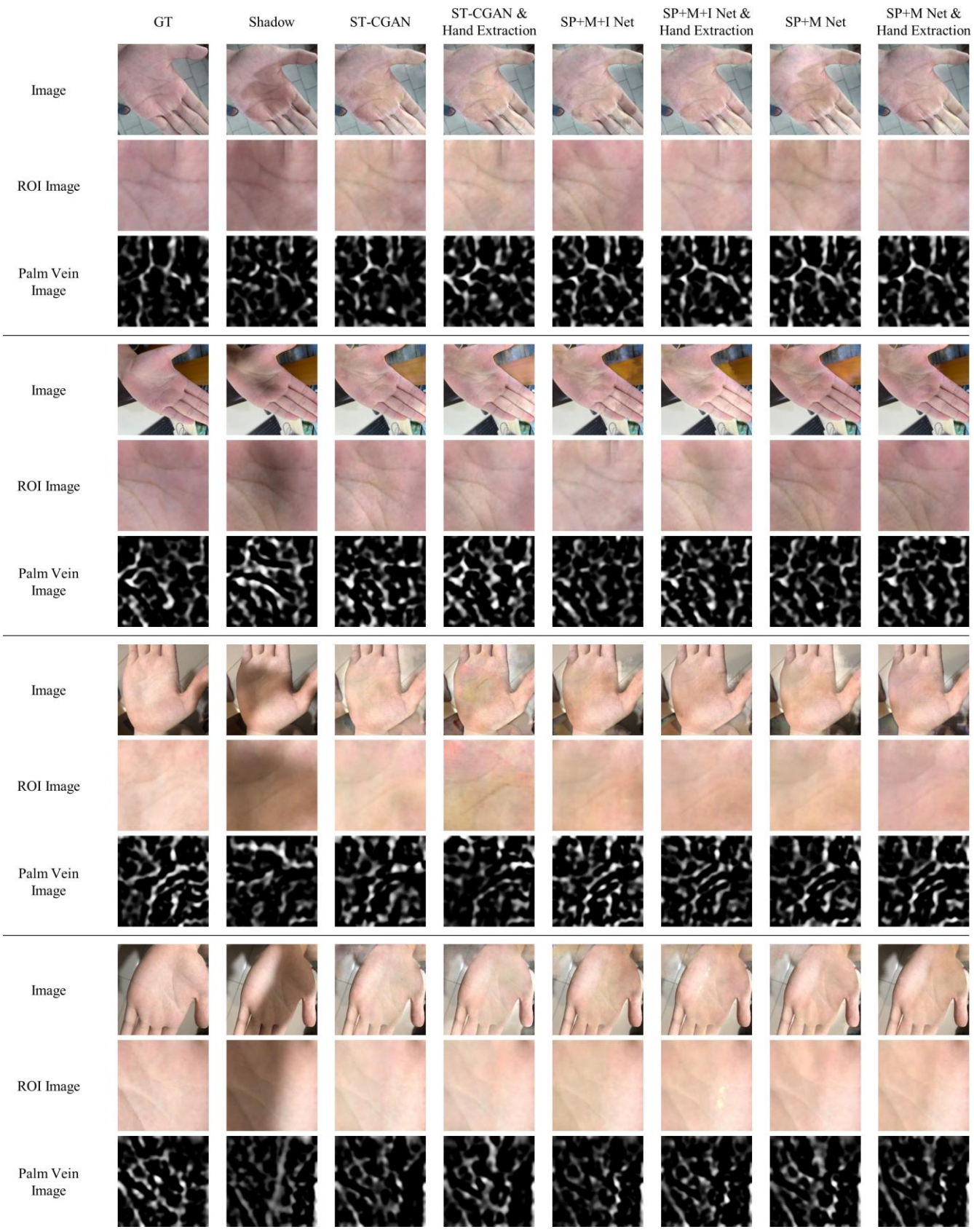

Fig. 10. Enhance the palm vein images by removing shadows on hands.

[6] G. D. Finlayson, S. D. Hordley, C. Lu, and M. S. Drew, "On the removal of shadows from images," IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI), vol. 28, no. 1, pp. 59-68, 2005

[7] G. D. Finlayson, M. S. Drew, and C. Lu, "Entropy minimization for shadow removal," International Journal of Computer Vision, vol. 85, no. 1, pp. 35-57, 2009.

[8] H. Gong and D. Cosker, "Interactive shadow removal and ground truth for variable scene categories," In Proceedings of British Machine Vision Conference (BMVC), 2014.

[9] M. Gryka, M. Terry, and G. J. Brostow, ”Learning to remove soft shadows,” ACM Transactions on Graphics (TOG), vol. 34, no. 5, 153, 2015.

[10] L. Qu, J. Tian, S. He, Y. Tang, and Rynson W.H. Lau, ”DeshadowNet: A multi-context embedding deep network for shadow removal,” In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017, pp. 4067-4075.

[11] X. Hu, L. Zhu, C.-W. Fu, J. Qin, and P. A. Heng, ”Direction-aware spatial context features for shadow detection,” In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2018, pp. 7454-7462.

[12] S. Bell, C. Lawrence Zitnick, Kavita Bala, and Ross Girshick, ”Inside-Outside Net: Detecting objects in context with skip pooling and recurrent neural networks,” In Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016, pp. 2874-2883.

[13] I. Goodfellow et al., ”Generative adversarial networks,” Communications of the ACM, vol. 63, no. 11, pp. 139-144, 2020.

[14] X. Hu, Y. Jiang, C.-W. Fu, and P.-A. Heng, ”Mask-ShadowGAN: Learning to remove shadows from unpaired data,” In Proceedings of IEEE International Conference on Computer Vision (ICCV), 2019, pp. 2472-2481.

[15] B. Ding, C. Long, L. Zhang, and C. Xiao, ”ARGAN: Attentive recurrent generative adversarial network for shadow detection and removal,” In Proceedings of IEEE International Conference on Computer Vision (ICCV), 2019, pp. 10213-10222.

[16] Y.-Y. Chuang, D. B. Goldman, B. Curless, D. H. Salesin, and R. Szeliski, ”Shadow matting and compositing,” ACM Transactions on Graphics, vol. 22, no. 3, pp. 494–500, July 2003.

[17] GTEA dataset, Available at: https://cbs.ic.gatech.edu/fpv/

[18] 11K Hands dataset, Available at: https://sites.google.com/view/11khands

[19] Bezier Curve, Available at: https://en.wikipedia.org/wiki/Bezier curve