Shadow Detection and Removal Using GAN

Takahiro Nagae , Ryo Abiko , Takuro Yamaguchi , Masaaki Ikehara *Dept., Keio Univ.*
Yokohama, Kanagawa 223-8522, Japan
{nagae, abiko, yamaguchi, ikehara}@tkhm.elec.keio.ac.jp

***Abstract*—To remove shadowed region in a single image, it is important to obtain high accuracy in both two processes, shadow detection and removal. In order to improve the results, recent methods perform these two processes simultaneously and use GAN for the training. However, since these methods do not try to maintain the luminance of non-shadowed regions, the output images tend to be faded. In this paper, to overcome fading problem, we proposed a new GAN structure based on shadow model. Since our GAN-based method focus on the variation of the illuminance, the illuminances of the shadowed regions, whose amount of change are large, are effectively estimated. In addition, non-shadowed regions remain slightly faded due to our new GAN structure and training method. Owing to our novel GAN structure and training method, our method outperforms state-of-the-art methods in PSNR and SSIM.**

***Index Terms*—shadow detection, shadow removal, GAN, illuminance**

I. Introduction

Shadowed regions often confound the process in computer vision such as segmentation, recognition or correction. The shadows removing is necessary to improve the accuracy of detecting and tracing objects. This process can be separated into two steps. The first step is the shadow detection process and the second step is the shadow removal process. In the shadow detection process, the shadowed regions are accurately recognized and generates a shadow mask that separates the shadowed area and non-shadowed area. In the shadow removal process, it focuses on the detected shadowed region process and reconstruct the shadow-free image. Both shadow detection and shadow removal has the same essential idea, and in shadow detection, the problem is solved by understanding the border of the shadowed and non-shadowed regions. Previously, shadow detection([4] [5] [6] [7]) and shadow removal([9] [10] [11]) have been performed separately. Wang's method [8] takes this into consideration and proposes a method that simultaneously learns the shadow detection and shadow removal processes. This method is able to remove shadows by sharing the information with the shadow detection. However, in terms of shadow removal, the non-shadowed regions gets closer to the shadowed regions while the shadow regions gets closer to the non-shadowed regions because of applying GAN [3]. Conventional methods have problems in the handling of non-shadowed regions. Unnecessary processing is done in the non-shadowed regions, resulting in the fading of these regions. In this paper we propose an end-to-end shadow removal method. Referencing the shadow model proposed in Yeal et al. [1], we improve over the problem of the degradation of quality on the non-shadow portions of the image in Wang et. al's [8] work. We compare the shadow region model with the non-shadow region model. As a result, light intensity is the only element that differentiates these two regions. Therefore, by aligning the lighting of the shadowed regions with the non-shadowed regions, shadow removal can be done. Instead of directly predicating and removing shadowed region, we ignore the pixel wise reciprocal intensity with respect to the non-shadowed region.

The details expressed are the following equation as:

$$I^{lit}(x,\lambda) = X(x,\lambda)Y(x,\lambda) \tag{1}$$

I^{lit}indicates light intensity of pixels.X and Y are input image and the estimated light intensity at position x respectively. If X is the image of non-shadowed regions, Y returns 1, otherwise returns the inverse of light intensity. As a result, only the pixels in the shadowed regions change, while keeping the pixels in the non-shadowed regions. Therefore, the proposed method is able to restrict the colors in non-shadowed regions from fading.

II. Background and Theory

A. Generative Adversarial Networks

The Generative Adversarial Network is comprised of two networks: the generator and the discriminator. The generator is a network that solves the given task, while the discriminator is a network that takes the output of the generator to determine if the image is real. The generator is trained so that it creates an image the tricks the discriminator to be real. In contrast, the discriminator is trained to be able to accurately differentiate between the ground-truth and generated images. This network is known as Generative Adversarial Network because two networks are trained in an adversarial manner to compete against each other, which forces the generator to accurately model the real data distribution.

B. Shadow Model

The shadow model in this paper is based on the shadow formulation model of Yeal et al [1]. This model is expressed is the following equation as:

$$I(x,\lambda) = L(x,\lambda)R(x,\lambda) \tag{2}$$

where I is the intensity reflected from position x with wavelength λ, while L and R are the illumination and reflection at position $x$respectively. In non-shadowed regions, the illumination can be expressed as a sum of the direct illumination, i.e., the sun for outdoor scenes and lights in indoor scenes, and the ambient illumination. Consequently, the intensities in non-shadowed regions are expressed as:

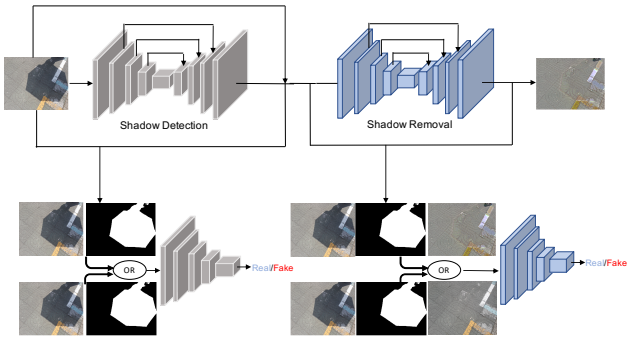


Fig. 1: Wang et. al's method [8]

$$I^{lit}(\boldsymbol{x}, \boldsymbol{\lambda}) = L^d(\boldsymbol{x}, \boldsymbol{\lambda})R(\boldsymbol{x}, \boldsymbol{\lambda}) + L^a(\boldsymbol{x}, \boldsymbol{\lambda})R(\boldsymbol{x}, \boldsymbol{\lambda}) \quad (3)$$

where L_d and L_a are the direct and ambient illumination respectively. On the contrary, in shadowed regions some objects occlude the direct illumination from the light source. The same object in the image can also occlude the ambient illumination. Therefore, the intensities in the shadowed regions are expressed as:

$$I^{shadow}(\boldsymbol{x}, \boldsymbol{\lambda}) = a(\boldsymbol{x})L^a(\boldsymbol{x}, \boldsymbol{\lambda})R(\boldsymbol{x}, \boldsymbol{\lambda}) \quad (4)$$

where $a(x)$ is the attenuation factor of the ambient illumination from the occluder in the image and takes a value between 0 and 1. When comparing the intensities of the non-shadowed and shadowed regions with equation Eq. (3) and Eq. (4), the reflectance $R(x, \lambda)$ does not change while the illumination changes. For this reason, shadows can be removed by estimating the illumination in shadowed and non-shadowed regions and then converting the illumination of the shadowed regions to the illumination of the non-shadowed regions.

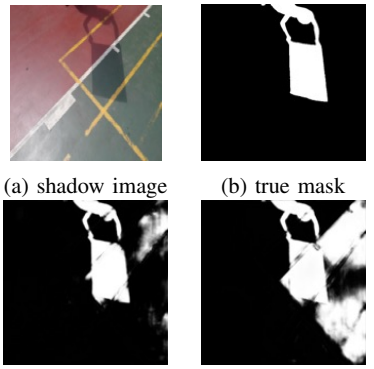
III. Related Work

In previous methods, the model process the whole image and alter shadowed and non-shadowed regions.

Since only the shadowed regions should be changed, this paper proposes a novel network that only changes the regions with shadows. The proposed method is based on the STCGAN (Wang et. al' s [8]) model and its algorithm is explained here. In terms of the shadow removal, the shadow model that was used is also explained.

A. STCGAN (Stacked Conditional Generative Adversarial Network)

The flow of Wang et. al ' s [8] is shown in Fig.1. In this diagram, D_1 and D_2 are the discriminators for the shadow detection and shadow removal respectively, and G_1 and G_2 are the generators for the shadow detection and shadow removal respectively.



(a) shadow image (b) true mask

(c) using GAN (d) not using GAN

Fig. 2: usefulness of GAN

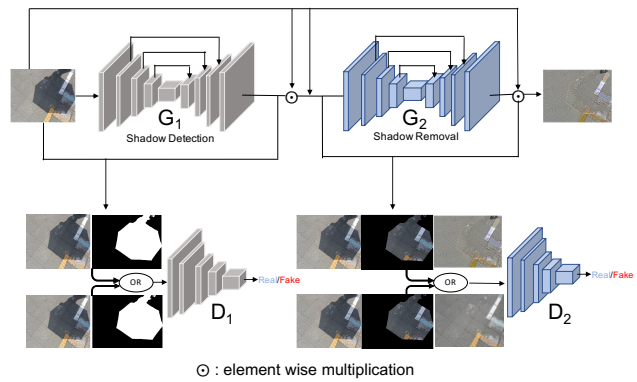


Fig. 3: structure of the proposed method

1) Generator: The network structure of the generator is based on a U-Net structure which is first used for segmentation of biomedical images. A characteristic of the U-Net structure is the skip connections that directly connects the feature maps of the encoder to the upsampled feature maps of the decoder. The advantage of this is that the skip connection propagates context information to the future layers that is lost during the down sampling process. For shadow detection, the U-Net can be effectively used for the segmentation of shadowed and non-shadowed regions, and for shadow removal, the U-Net is able to effectively reconstruct the details of shadowed regions. In the shadow detection generator, an RGB image with shadows is used as input to generate a binary mask which labels shadowed pixel as 1 and non-shadowed pixels as 0. On the other hand, the shadow removal network takes 4-channel input with the RGB image and the shadow mask and generates a clean RGB image without shadows.

2) Discriminator: During the shadow detection process, it is difficult to distinguish the shadowed regions with the dark regions that look like shadows, but by using a discriminator it is able to separate these regions by learning the features of shadows. The necessity of the discriminator for shadow detection is shown in Fig. 2. For the shadow removal process, the discriminator helps to generate more realistic images. Specifically during the shadow removal process, the generator creates a sort of overexposed region on the boundary of

shadowed and non-shadowed regions and the discriminator helps to mitigate this effect. The discriminator for shadow detection takes a 4-channel input with the input image and the binary mask from the shadow detection generator. The discriminator for the shadow removal takes a 7-channel input with the input RGB image, binary mask from the shadow detection generator, and the shadow-free RGB image from the shadow removal generator.

3) Training: The Adam optimizer is used to update the model weights of D_1 and D_2, and G_1 and G_2 back and forth. D_1 and D_2 is trained first and the results are fed into G_1 and G_2 to improve the accuracy of D_1 and D_2. Then, D_1 and D_2 are used to update the weights of G_1 and G_2 to improve their accuracy. By repeating this training process, the model is able to learn the optimum weights to effectively remove shadows.

IV. Proposed Method

The structure of the proposed method is shown in Fig.3. The proposed method is also divided into two processes: shadow detection and shadow removal. In the first process, the model of STCGAN is used to detect the shadows. In the second process, the shadow removal process, a new model is proposed based on the shadow model explained in Sec. II-B.

A. Shadow Removal

Instead of directly removing the shadow from images, the proposed generator estimates the ratio of the illumination. Accordingly, instead of directly using the output of the generator to calculate the loss with the ground-truth image, the output is obtained by element wise multiplication of the output of G_2 and the input RGB image and then compared to the ground-truth image. This forces G_2 to output 1 in non-shadowed regions and output the inverse ratio of the illuminance between the shadowed and non-shadowed region in the shadowed region. By taking element wise multiplication between the output of G_2 and the input image, the non-shadowed regions would be the same as the input. Furthermore, lighting would be the same for both shadow and non-shadow regions of the output. As a result, G_2 does not directly remove the shadow, but estimates the illuminance of the shadow. By using the shadow model to remove shadows from images, the proposed method is able to restrict the colors in non-shadowed regions from fading. In addition, through this process, pixels in the non-shadowed regions remain unchanged, while only the pixels in the shadow regions are altered.

B. Loss

The generator for the shadow detection and shadow removal are labeled G_1 and G_2 respectively and the discriminator for the shadow detection and shadow removal are labeled D_1 and D_2 respectively. The input image is labeled x, the ground-truth mask for shadow detection is labeled y, and the ground-truth image for shadow removal is labeled r. The loss functions used to train the network is explained below.

$$L_{CGAN_1}(G_1, D_1) = \boldsymbol{E}_{\boldsymbol{x},\boldsymbol{y}\sim pdata(\boldsymbol{x},\boldsymbol{y})}[logD_1(\boldsymbol{x},\boldsymbol{y})] + \boldsymbol{E}_{\boldsymbol{x}\sim pdata(\boldsymbol{x})}[log(1 - D_1(\boldsymbol{x}, G_1(\boldsymbol{x})))] \quad (5)$$

The loss function for D_1 is expressed in Eq. (5). In this equation, the ground-truth mask is labeled real and the output of G_1 is labeled fake to calculate the loss.

$$L_{data1}(G_1) = \boldsymbol{E}_{\boldsymbol{x},\boldsymbol{y}\sim pdata(\boldsymbol{x},\boldsymbol{y})} \|\boldsymbol{y} - G_1(\boldsymbol{x})\|_2 \quad (6)$$

In Eq. (6), the L_2 loss between the ground-truth mask and the output mask from G_1 is calculated to make a mask that is close to the ground-truth.

$$L_{data2}(G_2|G_1) = \boldsymbol{E}_{\boldsymbol{x},\boldsymbol{r}\sim pdata(\boldsymbol{x},\boldsymbol{r})} \|\boldsymbol{r} - G_2(\boldsymbol{x}, G_1(\boldsymbol{x}))\|_2 \quad (7)$$

where R denotes for $\boldsymbol{x}$ corresponding non-shadow region. In Eq. (7), the L_2 loss between the ground-truth image and the output image from G_2 is calculated to generate an image that is close to the ground-truth.

$$L_{CGAN_2}(G_2, D_2|G_1) = \boldsymbol{E}_{\boldsymbol{x},\boldsymbol{r}\sim pdata(\boldsymbol{x},\boldsymbol{y},\boldsymbol{r})}[logD_2(\boldsymbol{x},\boldsymbol{y},\boldsymbol{r})] + \boldsymbol{E}_{\boldsymbol{x}\sim pdata(\boldsymbol{x})}[log(1 - D_2(\boldsymbol{x}, G_1(\boldsymbol{x}), G_2(\boldsymbol{x}, G_1(\boldsymbol{x}))))] \quad (8)$$

Eq. (8) is the loss function for D_2. In this equation, the ground-truth image conditioned on the input image and ground-truth mask is labeled as real and the output image from G_2 conditioned on the input image and output mask from G_1 is labeled as fake to calculate the loss. Based on the above loss functions, the weights of the various networks are updated. In the experiment, Eq. (9) expressed below was used.

$$L = L_{data1}(G_1) + 5 \times L_{data2}(G_2|G_1) + 0.1 \times L_{CGAN_1}(G_1, D_1) + 0.1 \times L_{CGAN_2}(G_2, D_2|G_1) \quad (9)$$

C. Input Image of G_2

In the previous method, the generated mask from G_1 was used as input for G_2. However, in this proposed method, the element wise product between the generated mask from G_1 and the input image is used as input (Fig.4) for G_2.

D. Input Image

Input image used in our method are shown in Fig.5 and Fig.6. Fig.5 shows shadow detection generator (G_1) input and shadow removal generator (G_2) input. Fig.5a is the RGB image for G_1 input. For G_2 input, the RGB image (same as Fig.5a) and generated image for Fig.4. Fig.6 shows shadow detection discriminator (D_1) input and shadow removal discriminator (D_2) input. D_1 includes mask generated by G_1 with G_1 input. D_2 contains RGB removal image by G_2 with G_2 input.

V. Experiment

We used ISTD dataset of STCGAN(Wang et. al ' s [8] also adopts this as experiment dataset.). In the experiment, 1330 images are used for training and 540 images are used for testing. The images are selected randomly from the test set for comparison between the methods. First, we show the qualitative results in Fig.7. Compared to the two step process by Guo

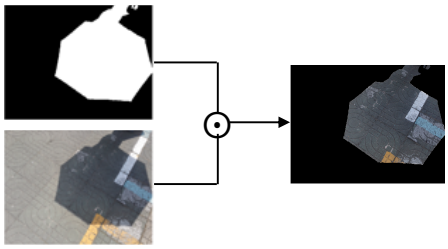
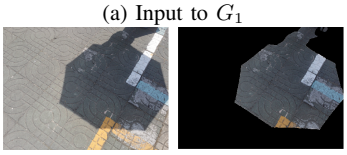


Fig. 4: Our input mask G_2

(a) Input to G_1



(b) Input to G_2

Fig. 5: Generator input of Our method

et. al [2], Wang et. al's [8] and the proposed method does not seem a sharp change in the boundaries of the shadowed and non-shadowed regions. Second the method of Wang et. al [8] had a problem with the fading of colors in the entire image including non-shadowed regions, but in the proposed method, it is evident that the color does not fade when compared to the ground-truth image. This attributes to the effect of the network structure that estimates the illumination and restricts making changes to non-shadowed regions.

TABLE I: Results of removal using ground-truth mask

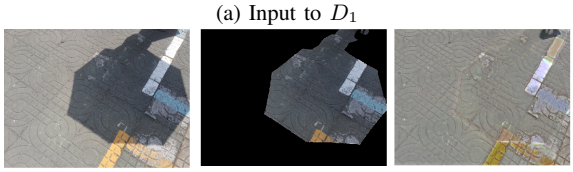
	PSNR(average of test set)	SSIM(average of test set)
Guo [2]	24.37	0.928
Wang [8]	26.47	0.948
Ours	**26.95**	**0.958**

TABLE II: Results of removal using STCGAN mask

	PSNR(average of test set)	SSIM(average of test set)
Wang [8]	25.42	0.939
Ours	**25.60**	**0.950**

For quantitative comparison with the ground-truth shadow-free image, PSNR and SSIM was used to compare Guo et. al [2], Wang et. al [8], and the proposed method. The results are shown in Table I. Wang et. al's [8] and the proposed method also compared in terms of the shadow detection and removal and the results in Table II. As seen in Table I, when we used the ground-truth mask, our proposed method is better than Guo et. al [2] which focuses on shadow removal. Likewise, Table II shows that our end-to-end model improves over Wang et. al [8].

(a) Input to D_1



(b) Input to D_2

Fig. 6: Discriminator input of Our method

VI. Conclusion

This method presents an end-to-end network that is able to effectively remove shadows from a single image. As a result, the proposed method is able to numerically outperform the previous methods . The presented method focuses on solving the problem with the previous method of processing the entire image when removing the shadows. Referencing the shadow model proposed in Yael et al. [1], we improve over the problem of the degradation of quality on the nonshadow regions of the image by estimating and manipulating only the shadows regions shadows in an image. As a result, quantitive measures show that our method's strengths over previous works. The visual quality of our output also shows little to no fading for the non-shadow regions compared to previous works.

References

[1] Shor, Yael, and Dani Lischinski. "The shadow meets the mask: Pyramid - based shadow removal." Computer Graphics Forum. Vol. 27. No. 2. Oxford, UK: Blackwell Publishing Ltd, 2008.

[2] Guo, Ruiqi, Qieyun Dai, and Derek Hoiem. "Single-image shadow detection and removal using paired regions." CVPR 2011. IEEE, 2011.

[3] Mirza, Mehdi, and Simon Osindero. "Conditional generative adversarial nets." arXiv preprint arXiv:1411.1784 (2014).

[4] Lalonde, Jean-Francois, Alexei A. Efros, and Srinivasa G. Narasimhan. "Detecting ground shadows in outdoor consumer photographs." European conference on computer vision. Springer, Berlin, Heidelberg, 2010.

[5] Khan, Salman Hameed, et al. "Automatic feature learning for robust shadow detection." 2014 IEEE Conference on Computer Vision and Pattern Recognition. IEEE, 2014.

[6] Vicente, Tomas F. Yago, Minh Hoai, and Dimitris Samaras. "Leave-one-out kernel optimization for shadow detection and removal." IEEE Transactions on Pattern Analysis and Machine Intelligence 40.3 (2017): 682-695.

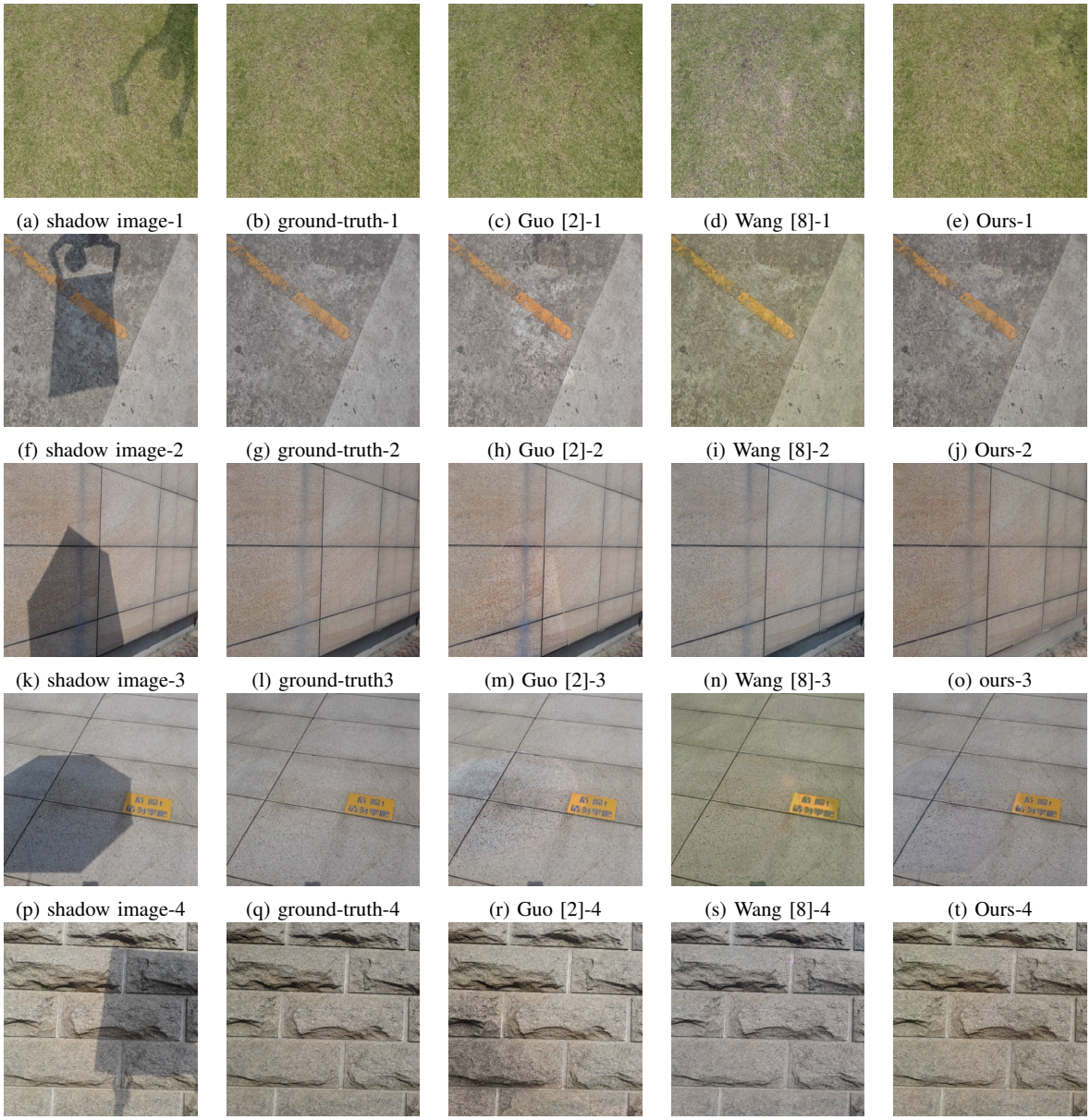
[7] Vicente, Yago, et al. "Noisy label recovery for shadow detection in unfamiliar domains." Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2016.

[8] Wang, Jifeng, Xiang Li, and Jian 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. 2018.

[9] Arbel, Eli, and Hagit Hel-Or. "Shadow removal using intensity surfaces and texture anchor points." IEEE transactions on pattern analysis and machine intelligence 33.6 (2010): 1202-1216.

[10] Finlayson, Graham D., Mark S. Drew, and Cheng Lu. "Entropy minimization for shadow removal." International Journal of Computer Vision 85.1 (2009): 35-57.

[11] Finlayson, Graham D., et al. "On the removal of shadows from images." IEEE transactions on pattern analysis and machine intelligence 28.1 (2005): 59-68.



(a) shadow image-1 (b) ground-truth-1 (c) Guo [2]-1 (d) Wang [8]-1 (e) Ours-1

(f) shadow image-2 (g) ground-truth-2 (h) Guo [2]-2 (i) Wang [8]-2 (j) Ours-2

(k) shadow image-3 (l) ground-truth3 (m) Guo [2]-3 (n) Wang [8]-3 (o) ours-3

(p) shadow image-4 (q) ground-truth-4 (r) Guo [2]-4 (s) Wang [8]-4 (t) Ours-4

(u) shadow image-5 (v) ground-truth-5 (w) Guo [2]-5 (x) Wang [8]-5 (y) Ours-5

Fig. 7: Final result image