

## Shadow Detection and Removal From Photo-Realistic Synthetic Urban Image Using Deep Learning

Hee-Jin Yoon<sup>1</sup>, Kang-Jik Kim<sup>1</sup> and Jun-Chul Chun<sup>1,\*</sup>

**Abstract:** Recently, virtual reality technology that can interact with various data is used for urban design and analysis. Reality, one of the most important elements in virtual reality technology, means visual expression so that a person can experience three-dimensional space like reality. To obtain this realism, real-world data are used in the various fields. For example, in order to increase the realism of 3D modeled building textures real aerial images are utilized in 3D modelling. However, the aerial image captured during the day can be shadowed by the sun and it can cause the distortion or deterioration of image. To resolve this problem, researches on detecting and removing shadows have been conducted, but the detecting and removing shadow is still considered as a challenging problem. In this paper, we propose a novel method for detecting and removing shadows using deep learning. For this work, we first build a new dataset of photo-realistic synthetic urban data based on the virtual environment using 3D spatial information provided by VWORLD. For detecting and removing shadow from the dataset, firstly, the 1-channel shadow mask image is inferred from the 3-channel shadow image through the CNN. Then, to generate a shadow-free image, a 3-channel shadow image and a detected 1-channel shadow mask into the GAN is executed. From the experiments, we can prove that the proposed method outperforms the existing methods in detecting and removing shadow.

**Keywords:** Deep-learning, shadow detection, shadow removal, synthetic data.

### 1 Introduction

Virtual reality refers to a specific environment or technology that is similar to reality using artificial technology but is not real, and is used in various industries such as architecture, education, and games. In order to increase the realism, which is one of the most important elements in virtual reality, the use of real world image data is increasing. For example, to design a city and simulate it by taking into account the various interactions between the data, it is necessary to map the actual aerial image to a three-dimensional modeling building for realistic representation. However, aerial images taken during the day have the problem of distorting various information, such as the color of the image, the building texture, due to the shadows caused by urban buildings and various

---

<sup>1</sup> Department of Computer Science, Kyonggi University, Gyeonggi-do, 16227, Korea.

\* Corresponding Author: Jun-Chul Chun. Email: jchun@kyonggi.ac.kr.

structures. To solve this problem, shadow detection and removal has been known as an important preprocessing process and has been studied for a long time to improve performance. Traditional researches have been studied based on physical features such as color [Finlayson, Hordley and Drew (2002)], intensity [Murali and Govindan (2013)], chromaticity, and texture of shadows. However, there is a problem in that performance cannot be guaranteed when the surrounding environment such as the dark color of the object, the ambient lighting condition, and the complexity of the object is changed. Recently, as the gpu and computing performance have improved, it has been applied to various fields such as object detection, image segmentation, and medical image processing, and has shown high performance [Mu and Zeng (2019)]. However, various types of data are required for the trained model to maintain high performance. Currently published datasets are divided for the purpose of shadow detection and removal. The most representative SBU dataset [Yago, Hou, Yu et al. (2016)] in shadow detection is 4,089 images of training data, and 1,330 images of training data of ISTD [Wang, Li and Yang (2018)] in shadow removal. It is a single texture image composed of 135 scenes. Consequently, published datasets are hard to expect general performance because of their limited types and quantities. In this paper, we construct a virtual data set using the three-dimensional spatial information provided by VWORLD [Kim, Jang, Yoo et al. (2017)]. We also use the Convolution Neural Network to detect shadow areas from shadow images. The detected mask is removed through Generative Adversarial Network(GAN)-based shadow removal method. We adopted the CPD network [Wu, Su and Huang (2019)] for shadow detection and improved the performance by using the channel attention module that recalibrates the importance of channel-wise feature in the feature aggregation. The shadow mask detected through the convolutional neural network is used as a condition for shadow removal. The GAN-based shadow removal method consists of a residual based generator and a discriminator that takes into account global and local patch information. This discriminator helps the generator to make plausible images through adversarial learning.

In summary, our main contributions are:

- (1) To evaluate our methods, we construct various kinds of virtual data sets using aerial images and VWORLD information.
- (2) We designed a channel attention module in a partial decoder that aggregates features to consider the importance of channel-level features in existing shadow detection networks.
- (3) To remove the shadows, we developed a residual-based generator and a modified discriminator to utilize multiple contextual information.

## **2 Related works**

### ***2.1 Physical based shadow detection and removal***

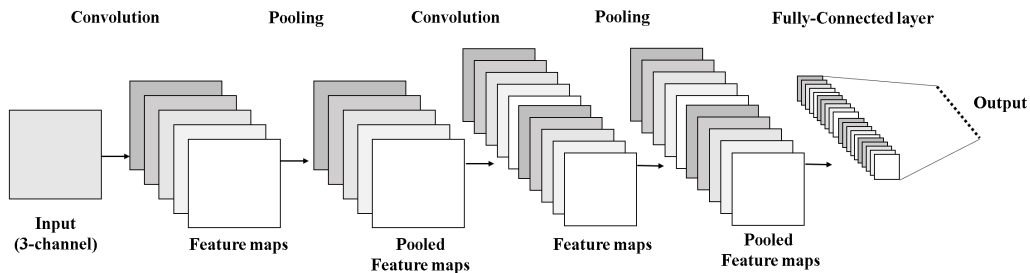
Many methods typically detect shadows and use the detected masks to remove the shadows. Traditional methods use hand-crafted features such as color, area, and user interaction to detect and remove shadows. The color-based model uses the difference between the original RGB image and the image conversion color space, such as CIELAB, HSV, and LUV. Finlayson et al. [Finlayson, Hordley and Drew (2002)] uses L2-norm to

create color-invariant images and sets the shadow edges to zero by comparing modified image with the original image. Guo et al. segmented areas with similar features, such as brightness and texture, based on shadow areas, and performed shadow area labeling using SVM-based area classifiers and Graph Cut. And then, shadow removal image is obtained by reconstructing the brightness value of each pixel using the designed illumination model [Guo, Dai and Hoeim (2013)]. Gong et al. detects the shadow boundary as the user strokes the shadow area, expanding the area with similar brightness values to the color model image changed with HSV, LUV, and YCbCr. After detection, the shadow model is efficiently removed by using the lighting model designed by referring to the lighting value change of the shadow boundary [Gong and Cosker (2014)]. However, the common limitations of these traditional methods are computationally expensive and cannot guarantee performance depending on conditions such as lighting values, texture colors, and light sources.

## 2.2 Deep learning-based shadow detection and removal

### 2.2.1 Convolution neural network-based shadow detection

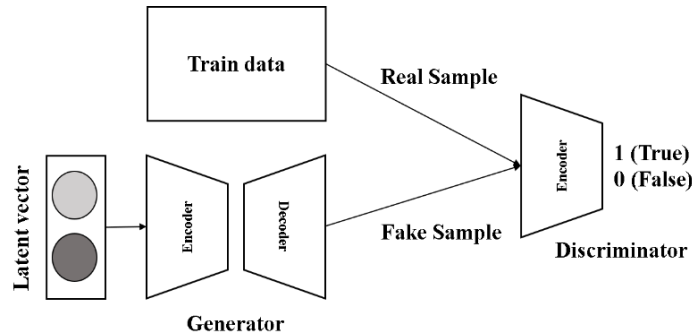
Inspired by the human visual perception system, the Convolution Neural Network (CNN) consists mainly of the convolution layer, the pooling layer, and the fully connected layer, based on local connectivity between neurons and neurons. A type of neural network used to analyze images such as object detection and image segmentation (see Fig. 1). Recently, due to the improvement of computing power, many researchers have studied the shadow detection method using deep learning and show higher performance than the existing hand-crafted feature methods. Qu et al. [Qu, Tian, He et al. (2017)] consists of three flow sub-networks that extract shape, semantic and global contextual features. Finally, the features extracted from each flow are finally combined to estimate the shadow mask. Hu et al. use the Recurrent Neural Network (RNN)-based direction aware module to learn two-dimensional spatial context information about features extracted from each layer of CNN. In addition, the contextual features of the surrounding shadows are aggregated in consideration of the weight along the direction [Hu, Zhu, Fu et al. (2018)]. Zhu et al. [Zhu, Deng, Hu et al. (2018)] make the shadow mask by refining and combining features (e.g., high-level and low-level context features) obtained through the bidirectional feature pyramid CNN structure using recurrent attention residual modules.



**Figure 1:** Convolution neural network architecture

### 2.2.2 Generative adversarial network-based shadow removal

The GAN consists of two modules (see Fig. 2). It is an adversarial learning framework between a generator that makes similar fake data and a discriminator that distinguishes real and fake data. The generator learns how to create an image with an encoder that compresses the function of the input and a decoder that restores the data. The discriminator learns to distinguish the fake image created by the generator with the real image, and the purpose of the two modules is to influence each other during the learning process, and to deceive the discriminator from distinguishing the real data and the fake data. Wang et al. use two conditional gans of stacked structure, shadow detection and removal are performed by end-to-end method [Wang, Li and Yang (2018)]. They make shadow mask from RGB images in shadow detection. That shadow mask, which is used as a condition in next shadow removal, is combined with the original image to remove the shadow. Oleksii Sidorov makes a shadow-free image using a new loss function based on pix2pix [Sidorov (2019)]. They remove shadows using a synthetic image similar to the real world. However, because the traditional method uses an end-to-end approach, it is difficult to expect good performance in both shadow detection and removal processes. In this paper, we propose to remove the shadow using the detected mask to remove the shadow.



**Figure 2:** Generative adversarial network architecture

## 3 Proposed method

In this part, we explain in detail in three sections. First, we show how to construct synthetic dataset in a 3D rendered virtual environment. Then, we propose a new shadow detection method using the CNN-based channel attention module. Finally, we describe a GAN-based shadow removal method with new discriminator using the detected shadow mask as a condition.

### 3.1 Synthetic data

In order to obtain urban datasets including buildings in various environments, we implemented a virtual reality program that can generate shadow dataset considering the actual location information of the sun and VWORLD [T Kim, Jang, Yoo et al. (2017)], an open platform for spatial information. The shadow changes shape depending on the position of the sun, the light source. In order to make realistic shadows in the virtual

environment, light sources are implemented using the AESL algorithm [Soe and Song (1996)] and the azimuth angle using the KASI algorithm [Bretagnon (1982)] by referring to the movement path of the real sun over time. Based on this, the user can freely transform the scene to the desired position. Also, by adjusting the position of the sun, users can get shadows, shadow masks and shadowless data without annotating the data.(see Fig. 3) To obtain various situations of data, we calculated the pixel height value of the image and organized 4 types of data according to the scenario.

$$H(x, y) = \sqrt{\frac{(d_{(x,y)} - d_{(x',y')})^2}{H_{Max}}} * k \quad (1)$$

where  $H_{(x,y)}$  represents the height value for the two-dimensional image coordinate,  $d_{(x,y)}$  is the distance from the earth, and  $d_{(x',y')}$  is distance from the earth.  $H_{Max}$  on the surface of the earth represents the highest building value, and the height of the building was normalized by Eq. (1).  $k$  was obtained by setting it to 240, which is an optimized value to change the height value to [0,255]. Tab. 1 shows the amount of data according to the height value and is organized according to four scenarios.



**Figure 3:** Sample images in data sets: shadow image, mask, shadow-less

**Table 1:** The amount of data over a range of height values for a building

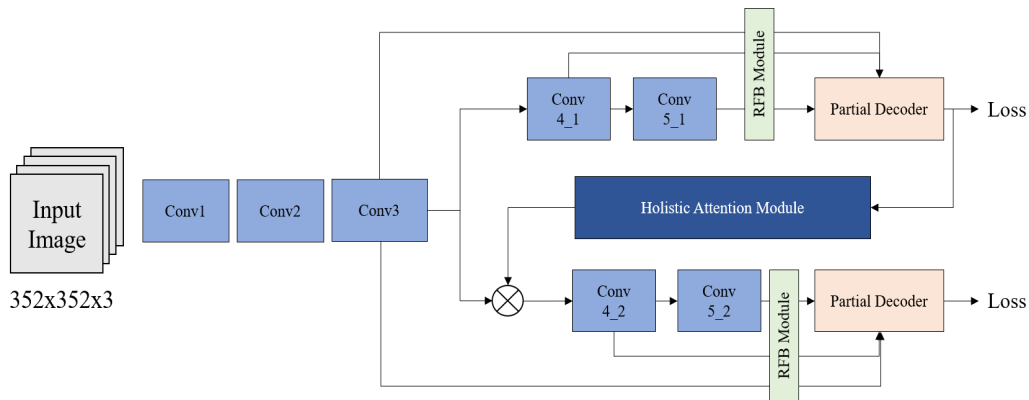
Type	Training	Test	Height range
Single building	1,230	122	$H(x,y) < 255$
Low building	1,500	130	$H(x,y) \leq 180$
High building	1,281	110	$180 \leq H(x,y) < 255$
Multiple	1,142	138	$H(x,y) < 255$

### 3.2 Shadow detection

We describe the existing CPD [Wu, Su and Huang (2019)] to detect shadows in the virtual environment data set and how to improve the performance by adding channel attention module [Hu, Shen and Sun (2017)] in the aggregation of features.

### 3.2.1 Network architecture

The proposed network consists of a layer that extracts five features from the Resnet-50 as a backbone network. The network efficiently generates shadow masks using a two-step cascading method using partial layers information to reduce the time and computational cost. In the first step, the extracted feature shows an approximate shadow area, but in the second step, the HAM (Holistic Attention Map) is used to obtain a feature map with extended coverage, and refined through convolution operation to improve detection performance. The overall structure of the network is shown in Fig. 4.



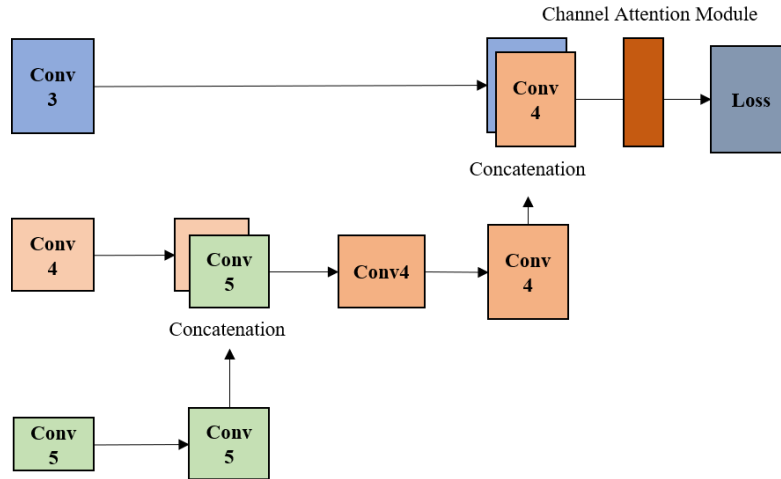
**Figure 4:** The architecture of CNN-based shadow detection network

First of all, each feature of each layer is obtained through the convolution, pooling, and activation function calculation with the input of  $352 \times 352$  size as the input 3-channel RGB image. The results generated in Convolution 3, 4, and 5 layers among the 5 generated features were applied to RFB (Receptive Field Block) module composed of dilation convolutions with different ratios to capture global context information. After that, the three feature informations are different sized by the convolution operations, upsampling and  $3 \times 3$  convolution operations are used to increase the size, and combined with the previous layer to calculate the loss function. Next step, we use the HAM (Holistic Attention Module) to optimize the convolution operation by increasing the scope of the convolution operation on the feature map in the initial stage. As a result, important features were highlighted. Then, using the RFB module and the partial decoder, the feature information generated in the initial stage is refined in the second stage.

### 3.2.2 Proposed partial decoder

The overall structure of the proposed decoder module is shown in the Fig. 5, and the loss function is calculated by combining the high context features of the partial layer. Since the feature map of the previous layer is twice as spatially different than the next layer, the upsampling-convolution  $3 \times 3$  operation strategy is used to refine and combine the features of various sizes. However, since the distinction of other feature information is not considered in the combination of various features of the existing partial decoder structure, some essential information is overlap or lost. This causes performance degradation and

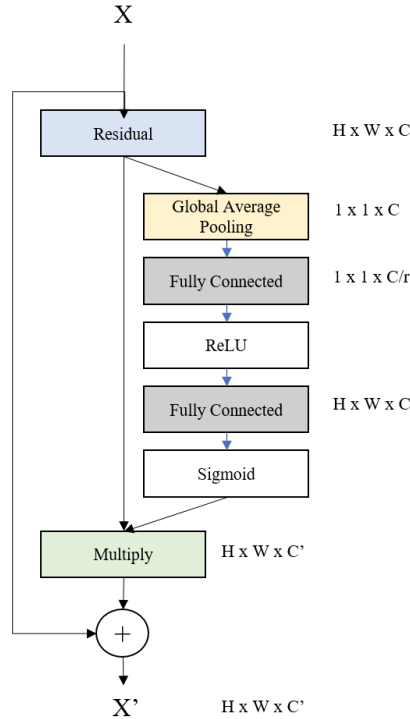
results in unfaithful estimation [Zhao and Wu (2019)]. To solve this problem, we design the channel attention module at the end of the three features to consider the importance of the aggregated context features, and obtain the final feature map through the Convolution  $3 \times 3$  layer and the Convolution  $1 \times 1$  layer.



**Figure 5:** The architecture of proposed partial decoder

### 3.2.3 Channel attention module

In the existing CNN, the convolution operation learns the local characteristics of an image or feature map, and the learned features are expressed differently in units of channels in the image and compresses the information in the bottleneck section to create a feature map with a global receptive field [Hu, Shen and Sun (2017)]. Since our network consists of two decoders, we need to simply recalibrate the compressed feature map. We used the squeeze and excitation method first proposed by Hu et al. The channel attention module is shown in Fig. 6. The size of the first channel feature vector is  $H \times W \times C$ , and the channel feature information is obtained through a global average pooling operation. It is a method used to extract important features by a channel-wise average of two-dimensional spatial features. Therefore, the number of channels is maintained from the size of the original feature map, and a feature map of size  $1 \times 1 \times C$  is reduced, which is reduced in two-dimensional space. The obtained feature map compresses the number of channels with the optimized ratio  $r=16$  through the Fully Connected layer. It is then activated by a ReLU nonlinear function that can simplify gradient vanishing or derivatives. The activated features are extended back to the original channel number through the Fully Connected layer, and the features are readjusted according to their importance and normalized by the Sigmoid function. Lastly, multiplication with previous layer emphasized important features and applied residual concept to add to existing feature map.



**Figure 6:** The illustration of flow in channel attention module

### 3.2.4 Training strategies

We trained the shadow image and the mask in pairs to learn the binary image, and used the BCE (Binary Cross Entropy) loss function as shown in the Eq. (2).

$$BCE = -\frac{1}{N} \sum_{i=0}^N y_i \cdot \log(\hat{y}_i) + (1 - y) \cdot \log(1 - \hat{y}_i) \quad (2)$$

where  $N$  denotes the total number of data,  $y_i$  is represent the binary of the target label which is 0 and 1 (if  $y=1$ , it is shadow region, and  $y=0$ , otherwise).  $\hat{y}_i$  learns to generate shadow shadow masks with classification probability scores from the network. (where  $\hat{y}_i \in [0,1]$ ). We used Adam Solver to optimize weights and set the learning rate  $1e^{-4}$ .

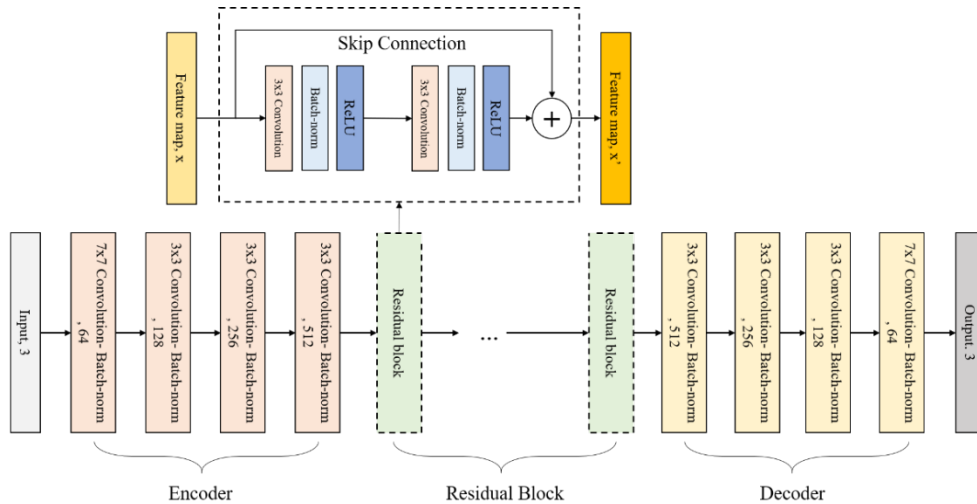
### 3.3 Shadow removal

We propose a new shadow removal method based on Conditional GAN by using the mask detected by CNN as a condition. It consists of a generator that generates fake images and a discriminator that distinguish real and fake, and the two modules generate plausible fake images through adversarial learning. We added global context information to the existing patch method discriminator so that the fake data created by the producer can take into account two contextual information.



### 3.3.1 Generator

The generator is responsible for making plausible fake images to deceive the discriminator. As shown in Fig. 7, the proposed generator structure consists of an encoder, residual block, and decoder. In the encoder, it is a part to compress the information of the image and is composed of four convolution blocks. The convolution block consists of Convolution-batchnorm-ReLU, and we set the kernel size of convolution of the first layer to  $7 \times 7$ . The larger the convolution kernel size, the more spatial information can be obtained from the original image at once, but the smaller the feature map size is. The other convolution layer kernel size is  $3 \times 3$  and makes a final feature map of 512 channels. Residual block can solve the problem caused by deep network structure such as vanishing gradient or exploding through residual learning, and can improve performance by deeply constructing network depth [He, Zhang, Ren et al. (2015)]. Residual block consists of  $3 \times 3$  convolution and skip connection. And it has the advantage that the gradient flows well in back propagation without the need for additional calculation amount or parameter due to skip connection. Decoder is the process of restoring the original image while increasing the compressed features. ConvTranposed2d-batchnorm-Relu is symmetrically constructed with the encoder part and creates a fake image that is similar in size to the input image.

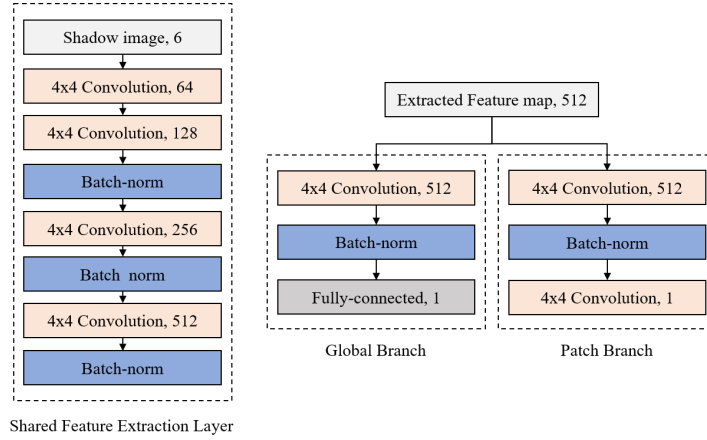


**Figure 7:** The structure of generator: encoder, residual block, decoder

### 3.3.2 Discriminator

The discriminator is responsible for classifying the fake image created by the generator and the real image of the training data. In this paper, global context information is added to the existing patch wise discriminator model, and two information can be optimized by joint loss. The proposed discriminator model is divided into two branches that take into account global context information and patch context information through a shared convolution layer (see Fig. 8). Composed of batchnorm, compresses the characteristics of the input image. Then, the patch branch extracts the patch feature vector through the  $4 \times 4$  convolution

layer and the other branch extracts the global feature vector using the Fully Connected Layer. Optimized through joint loss to combine and learn two context information.



**Figure 8:** The structure of the discriminator has two branch: global, patch

### 3.3.3 Training strategies

The joint loss equation used to train the proposed structure is shown in Eq. (3).

$$L_{joint\ loss} = \lambda_1 * L_{reconstruction} + \lambda_2 * L_{global} + \lambda_3 * L_{patch} \quad (3)$$

$$L_{reconstruction} = \sum_{i=1}^n |y_{GT} - Y_{Predicted}| \quad (4)$$

$$L_{adversarial}(G, D) = E_{x \sim P_{data}} \log D(x|y) + E_{x \sim P(x|y)} \log(1 - D(G(x|y), y)) \quad (5)$$

We set optimized parameters as  $\lambda_1 = 0.995$ ,  $\lambda_2 = 0.005$  and  $\lambda_3 = 0.005$ . In Eq. (4),  $L_{reconstruction}$  has the purpose of generate image similar to ground truth by applying the L1-norm distance loss, and  $L_{global}$  and  $L_{patch}$  are adversarial loss. This loss function helps match true color distributions. [Isola, Zhu, Zhou et al. (2017)] We used the Adam optimizer to make learning fast and stable, with a learning rate of  $1e^{-4}$  and an epoch of 400.

## 4 Experiment result

Our environment is Ubuntu 16.04 LTS 64-bit NVIDIA GeForce 1080Ti, using the pytorch framework. We experimented with 5,653 photo-realistic urban shadow image. Among them, 5,153 images are used for training and the rest of 500 images are used for testing. We experimented with shadow detection and removal respectively, and the results are as follows.

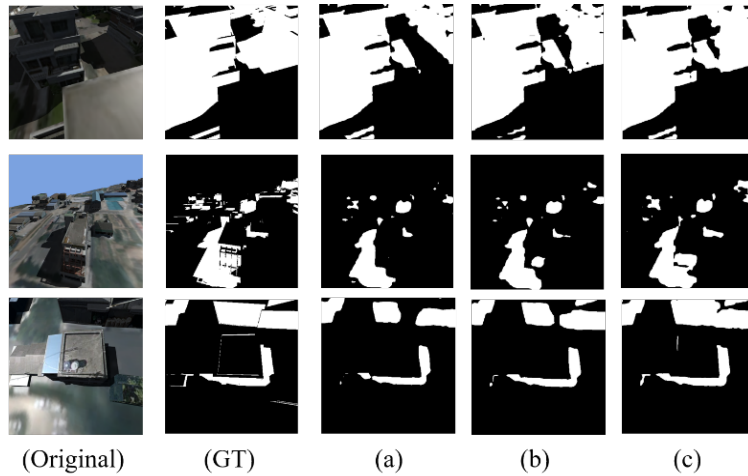
#### 4.1 Shadow detection result

We evaluated the shadows, non-shadows, and all regions separately using BER (Balanced Error Rate) method. In Tab. 2, the channel attention module is applied to the position where the features of the convolution operation are combined in the partial decoder. Experimental results show that when the channel attention is used, the performance is improved by 0.38% in the shadow area and 0.53% in the non-shadow area compared to the existing method.

$$BER(\text{Balanced Error Rate}) = 1 - \frac{1}{2} * \left( \frac{TP}{TP+FN} + \frac{TN}{TN+FP} \right) \quad (6)$$

**Table 2:** The quantitative comparison of channel attention module

Method	Shadow	Non-shadow	All
CPD [Wu, Su and Q.Huang (2019)]	7.51	2.60	5.05
Conv 3,4	7.21	2.30	4.75
Conv 4,5	7.13	2.37	4.62



**Figure 9:** Comparison of results with channel attention module, (a) basic CPD, (b) channel attention with conv3,4 (c) channel attention module with con4,5

#### 4.2 Shadow removal result

We used RMSE (Root Mean Square Error) in the LAB color space [Kahu and Bhurchandi (2018)] to compare performance. In addition, SSIM (Structural Similarity) is used to evaluate how similar the restored results are. In Tab. 3. The existing methods were tested and compared with Patch-GAN with different discriminator. Experimental results show better performance than other traditional methods, and we improved 0.3 in

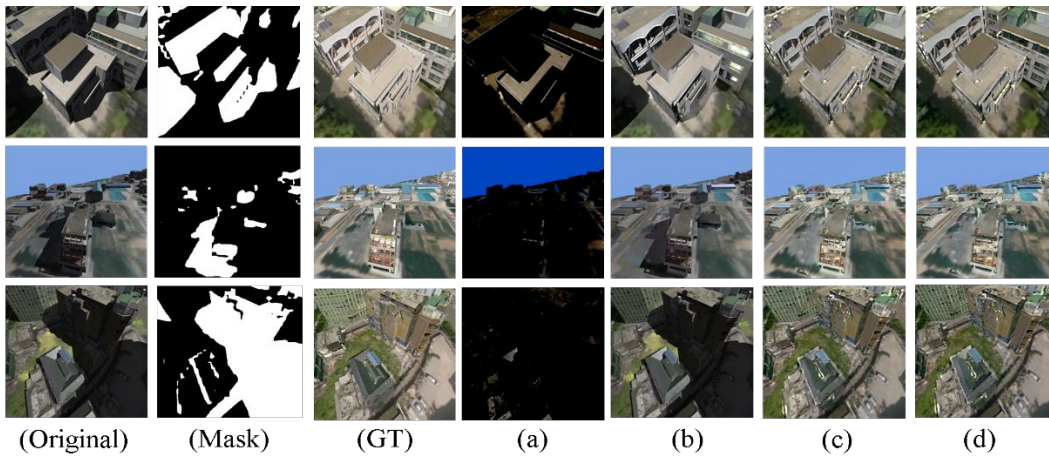
the shadow area and 1.18 in the all area when the discriminator was modified. It also shows a 0.0220 improvement in SSIM (see Eq. (7)).

$$SSIM(x, y) = \frac{(2\mu_x\mu_y + c_1)(2\sigma_{xy} + c_2)}{(\mu_x^2 + \mu_y^2 + c_1)(\sigma_x^2 + \sigma_y^2 + c_2)} \quad (7)$$

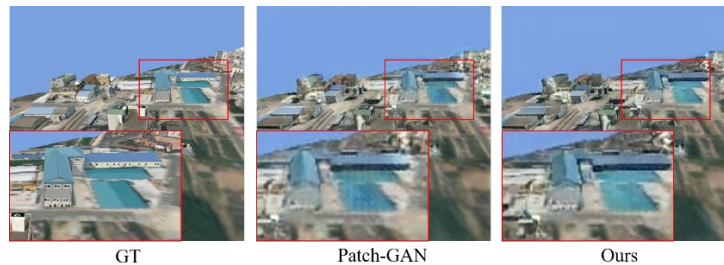
It means between two windows  $x, y$  of size  $13 \times 13$ , where  $\mu$  mean,  $\sigma$  variance,  $\sigma_{xy}$  covariance of  $x$  and  $y$ ,  $c_n = (k_n L)^2$ .

**Table 3:** Quantitative comparison of removal using RMSE and SSIM

Method	Shadow	Non-shadow	All	SSIM
Angular-GAN [Sidorov (2019)]	15.11	20.76	16.44	0.1230
Guo [Guo, Dai and Hoiem (2013)]	11.71	8.08	10.24	0.6078
Patch-GAN	4.55	4.18	6.11	0.8460
Ours	4.25	4.22	5.03	0.8680



**Figure 10:** Comparison of results with various methods, (a): Angular-GAN [Sidorov (2019)], (b): Guo et al. [Guo, Dai and Hoiem (2013)], (c): patch discriminator based GAN, (d): ours



**Figure 11:** Detailed comparison of generated images according to the discriminator

## 6 Conclusion

In this paper, we propose a new method for detecting and removing photo-realistic synthetic image shadows. According to the experimental results, the method trained in two stages showed higher performance than the end-to-end method for shadow removal. In shadow detection, we improved the performance by using a channel attention module that account for the channel-level importance. In shadow removal, joint loss is used to generate fake data by considering global and patch information at the same time. By considering both context information, we have confirmed that it generates a more plausible image. In future work, we will use the spatial attention model for CNN-based shadow detection network.

**Acknowledgement:** This research was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (NRF-2018R1D1A1B07048819). This work also was partially supported by Kyonggi University's Graduate Research Assistantship 2019.

## References

- Bretagnon, P.** (1982): Theory for the motion of all the planets-the VSOP82 solution. *Astronomy and Astrophysics*, vol. 114, no. 2, pp. 278-288.
- Finlayson, G. D.; Hordley, S. D.; Drew, M. S.** (2002): Removing shadows from images. *Proceedings of the 7th European Conference on Computer Vision*, pp. 823-836.
- Gong, H.; Cosker, D. P.** (2014): Interactive shadow removal and ground truth for variable scene categories. *Proceedings of the British Machine Vision Conference*.
- Guo, R.; Dai, Q.; Hoiem, D.** (2013): Paired regions for shadow detection and removal. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 35, pp. 2956-2967.
- He, K.; Zhang, X.; Ren, S.; Sun, J.** (2015): Deep residual learning for image recognition. *IEEE Conference on Computer Vision and Pattern Recognition*.
- Hu, J.; Shen, L.; Sun, G.** (2017): Squeeze-and-excitation networks. arXiv:1709.01507.
- Hu, X.; Zhu, L.; Fu, C. W.; Qin, J.; Heng, P. A.** (2018): Direction aware spatial context features for shadow detection. *IEEE Conference on Computer Vision and Pattern Recognition*, pp.7454-7462.
- Isola, P.; Zhu, J. Y.; Zhou, T.; Efros, A. A.** (2017): Image-to-image translation with conditional adversarial networks. *IEEE Conference on Computer Vision and Pattern Recognition*.
- Kahu, S. Y.; Bhurchandi, K. M.** (2018): JPEG-based variable block-size image compression using CIE  $L^*a^*b^*$  color space. *KSII Transactions on Internet and Information Systems*, vol. 12, no. 10, pp. 5056-5078.
- Kim, T. H.; Jang, H. S.; Yoo, S. H.; Go, J. H.** (2017): 3D tile application method for improvement of performance of V-world 3D map service. *Journal of Korean Society for Geospatial Information System*, vol. 25, no. 1, pp. 55-61.

- Lee, J. M.; Lee, J. H.; Lim, J. Y.; Kim, M.** (2019): Bandwidth-efficient live virtual reality streaming scheme for reducing view adaptation delay. *KSII Transactions on Internet and Information Systems*, vol. 13, no. 1.
- Mu, R.; Zeng, X.** (2019): A review of deep learning research. *KSII Transactions on Internet and Information Systems*, vol. 13, no. 4.
- Murali, S.; Govindan, V. K.** (2013): Shadow detection and removal from a single image using LAB color space. *Cybernetics and Information Technologies*, vol. 13, no. 1, pp. 95-103.
- Qu, L.; Tian, J.; He, S.; Tang, Y.; Lau, R. W. H.** (2017): Deshadownet: a multi-context embedding deep network for shadow removal. *IEEE Conference on Computer Vision and Pattern Recognition*.
- Sidorov, O.** (2019): Conditional GANs for multi-illuminant color constancy: revolution or yet another approach? *IEEE Conference on Computer Vision and Pattern Recognition*.
- Soe, K. B.; Song, K. D.** (1996): Development algorithms to predict the luminous flux transfer rates of vertical rectangular daylight duct systems with the consideration of direct sunlight incidence. *Journal the Korean Society of Living Environmental System*, vol. 3, no. 3, pp. 21-34.
- Wang, J.; Li, X.; Yang, J.** (2018): Stacked conditional generative adversarial networks for jointly learning shadow detection and shadow removal. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*.
- Wu, Z.; Su, K.; Huang, Q.** (2019): Cascaded partial decoder for fast and accurate salient object detection. arXiv:1904.08739.
- Yago, T.; Hou, L.; Yu, C. P.; Hoai, M.; Samaras, D.** (2016): Large-scale training of shadow detectors with noisily-annotated shadow examples. *European Conference on Computer Vision, Springer*, pp. 816-832.
- Zhao, T.; Wu, X.** (2019): Pyramid feature attention network for saliency detection, *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*.
- Zhu, L.; Deng, Z.; Hu, X.; Fu, C. W.; Xu, X. et al.** (2018): Bidirectional feature pyramid network with recurrent attention residual modules for shadow detection. *European Conference on Computer Vision*.