

IMAGE SHADOW REMOVAL BASED ON GENERATIVE ADVERSARIAL NETWORKS

Accurate detection of shadows and removal in the image are complicated tasks, as it is difficult to understand whether darkening or gray is the cause of the shadow. This paper proposes an image shadow removal method based on generative adversarial networks. Our approach is trained in unsupervised fashion which means it does not depend on time-consuming data collection and data labeling. This together with training in a single end-to-end framework significantly raises its practical relevance.

Taking the existing method for unsupervised image transfer between different domains, we have researched its applicability to the shadow removal problem. Two networks have been used. The first network is used to add shadows in images and the second network for shadow removal. ISTD dataset has been used for evaluation clarity because it has ground truth shadow free images as well as shadow masks. For shadow removal we have used root mean squared error between generated and real shadow free images in LAB color space. Evaluation is divided into region and global where the former is applied to shadow regions while the latter to the whole images. Shadow detection is evaluated with the use of Intersection over Union, also known as the Jaccard index. It is computed between the generated and ground-truth binary shadow masks by dividing the area of overlap by the union of those two. We selected random 100 images for validation purposes. During the experiments multiple hypotheses have been tested. The majority of tests we conducted were about how to use an attention module and where to localize it.

Our network produces better results compared to the existing approach in the field. Analysis showed that attention maps obtained from auxiliary classifier encourage the networks to concentrate on more distinctive regions between domains. However, generative adversarial networks demand more accurate and consistent architecture to solve the problem in a more efficient way.

Keywords: generative adversarial networks, shadow removal, shadow generation.

1. Introduction

Shadow is a common visual phenomenon when the object overlaps illumination source. Detected shadows can provide the important clues for better visual scene understanding [1; 2]. However, they can degrade the performance of algorithms in several computer vision spheres as object detection [17], tracking [11] and intrinsic image decomposition [15]. Therefore, effective shadow removal could give a performance boost for these tasks.

Shadow removal is a very challenging task because it is not enough to detect and remove the shadow, we also need to fill the background so it looks naturally for both human and a computer system.

Current works can be divided into two groups: classical and deep learning-based. Classical solutions used user input or hand-crafted features for shadow detection [7; 3] after which they tried to make the shadow region match the background. Meanwhile, deep learning-based approaches use neural networks for extracting high level features and background filling. One of the early works [21]

used three different neural networks operating in different contexts for more quality features extraction. Later, Hu et al. [9] explored direction-aware spatial context for this task to compensate the lack of data. After that, Wang et al. [24] used adversarial learning by stacking two networks together where one is used for shadow detection and one for shadow removal. More recently, Ding et al. [5] constructed the framework which used attention maps and recurrent learning. Other approach [25] argued that earlier works were not directly constructed for shadow removal task and proposed the novel architecture with hierarchical features aggregation.

However, all these methods used the supervised data to train and thus demanded the tedious collection and annotation processes. For the similar reason these approaches are also constrained with the complexity of the scenes. It is also argued that such approach may lead to change in illumination between shadow and shadow-free images [27].

Recently, the unsupervised solution was presented [27] where the problem was formulated in unsupervised mapping learning between two domains – shadow and shadow-free – using CycleGAN

[9] framework. First CGAN detects and removes the shadow while the other tries to generate it given the image and shadow mask as the input. Shadow mask, therefore, is received by running the Otsu's algorithm on the difference between generated and input image. The final results are competitive with those outlined above and significantly higher than general image-to-image translation with CycleGAN. However, this approach is not directly constructed to fit the shadow removal task, thus it has problems such as leaving the artifacts on the shadow boundaries and using binary masks for detected shadows. One more problem that may have consequences in a real-world application is that shadow generator network chooses the mask at random, so we could receive the image with inappropriate shadow on it.

2. Generative adversarial networks (GAN)

GAN is a framework for estimating the generative models that was firstly proposed by Ian J. Goodfellow et al. [6]. It consists of two networks where generator network G captures the data distribution and discriminator network D estimates the probability of the sample came from data distribution. In an initial variant G took the noise vector (usually, from uniform distribution) as the input trying to generate a sample that will "trick" the discriminator D . This framework corresponds to a minimax game (two-player non-cooperative game) where G is maximizing the probability of D making mistake. To bring it formal we can consider such minimax game in which we should optimize a following loss function:

$$\min_G \max_D L(D, G) = E_{x \sim p_r(x)} [\log D(x)] + \quad (1)$$

$$+ E_{z \sim p_z(z)} [1 - \log D(G(z))]$$

where p_r is real data distribution over sample x while p_z is a noise distribution over noise z . This work provoked a significant amount of research due to its capacity to generate high-quality samples. It was further improved by introducing Conditional GAN [18] which uses the label information to present the multi-modal solution. Thus, researcher can tune what kind of sample the network should generate.

CGANs were successfully used in learning the mappings between the domains, for example pix2pix [10] approach can handle multiple vision tasks as day to night, summer to winter or aerial to map in a single framework by introducing the pairs of images from each domain. However, it cannot handle the domains with no one-to-one mappings as for style transfer etc.

Hence, for unsupervised domain mapping CycleGAN [27] was firstly proposed by using the cycle consistency loss between domains. Still, it suffered from mode collapse generating one sample for different inputs. For that reason there was a research conducted [1; 14] to extend the initial solution to cope with the "many-to-many" mapping with the use of latent variables.

There has been a significant amount of work done and now GANs can generate high-quality images that are hardly distinguishable from the real ones. They are particularly good at face generation [12], style transfer [24; 3], inpainting [19; 20], domain transfer/adaptation [27; 1; 14] and are also used for shadow removal/detection [24; 5].

3. Method

We divide training into two parts: one to learn from shadow images and one to learn from shadow-free ones.

3.1. Learning from shadow images

Adversarial learning

Let I_s be an image from shadow X_s domain. We use a generator network $G_{s \rightarrow f}$ to translate an image to shadow-free X_f domain and obtain \tilde{I}_f . $G_{s \rightarrow f}$ network also includes an auxiliary classifier η_s , where $\eta_s(x)$ represents the probability of x taken from shadow domain [13].

Then, we use the corresponding discriminator D_f to discriminate whether the data comes from X_f or $G_{s \rightarrow f}(X_s)$. This network also consists the auxiliary classifier η_{D_f} that is aimed to solve the same task as the discriminator itself. We should also notice that Least Squares GAN objectives are used for more stable training, thus the adversarial loss will look like this:

$$L_{GAN}^s = E_{x \sim X_f} [(D_f(x))^2] + \quad (2)$$

$$+ E_{x \sim X_s} [(1 - D_f(G_{s \rightarrow f}(x)))^2]$$

CAM attention

Auxiliary classifier η_s is used in $G_{s \rightarrow f}$ to distinguish between domains and is inspired by CAM (Class activation maps) [24].

Let C_s^k be a k -th feature map of $G_{s \rightarrow f}^l$ output from l -th layer. Then, C_s^{kij} is the value at (i, j) position and we want to learn the importance weights for each feature map by using the global pooling layers (i.e. average, max). Thus, we obtain:

$$a_s = w_s * C_s = \{w_s^k C_s^k | 1 \leq k \leq n\} \quad (3)$$

where n is a number of feature maps and w_s^k is an importance weight for the k -th feature map.

We are learning those weights from:

$$\eta_s = \sigma\left(\sum_k w_s^k \sum_{ij} C_s^{kij}\right) \quad (4)$$

To make η_s distinguish between domains the corresponding cross-entropy loss is optimized:

$$L_{cam}^{s \rightarrow f} = -(E_{x \sim X_s}[\log(\eta_s(x))]) + E_{x \sim X_f}[\log(1 - \eta_s(x))] \quad (5)$$

Then, a_s (3.2) is transferred as the input to the following layer of the network and the learning continues.

Attention a_s is aggregated to be transferred as attention map A_s to $G_{f \rightarrow s}$:

$$A_s = \sum_{hw} \sum_c a_s^{hwc} \quad (6)$$

where we sum the values over the channels c .

Auxiliary classifier η_{D_f} is also integrated in D_f to decide whether the data comes from X_f or $G_{s \rightarrow f}(X_s)$:

$$L_{cam}^{D_f} = E_{x \sim X_f}[(\eta_{D_f}(x))^2] + E_{x \sim X_s}[(1 - \eta_{D_f}(G_{s \rightarrow f}(x)))^2] \quad (7)$$

Cycle consistency and identity loss

If we only use adversarial loss for learning then the mapping is highly under-constrained. That is why we present the inverse transformation $G_{f \rightarrow s}$ to transform the images back and encourage the contents to be the same.

As we outlined above generator network $G_{f \rightarrow s}$ additionally takes attention map A_s and generated mask M_s [9] as the input (concatenating to the image as additional channels). To preserve the consistency between the generated shadow image and the original one we take the same attention map and shadow mask extracted from shadow removal generator $G_{s \rightarrow f}$. This allows to produce multiple shadow images from one shadow-free raising the generalization capacity. Shadow mask is a binary map where -1 indicates non-shadow region while 1 – the shadow region. Attention map is also normalized to $[-1,1]$ and is used for complementing the shadow mask.

Then, we formulate following cycle-consistency loss:

$$L_{cycle}^{s \rightarrow f} = E_{x \sim X_s}[||G_{f \rightarrow s}(G_{s \rightarrow f}(x), A_s, M_s) - x||_1] \quad (8)$$

However, using only adversarial and cycle-consistency losses gives the generators freedom to change colors on images without being penalized. That is why researches in original work [12] introduced an identity loss to regularize the generators to be near an identity mapping when the inputs from

target domain are provided. Furthermore, this approach allows our solution to remove/generate shadows only when the image from proper domain is given:

$$L_{idt}^s = E_{x \sim X_s}[||G_{f \rightarrow s}(x, A_n, M_n) - x||_1] \quad (9)$$

where A_n, M_n are constructed only from -1 (non-shadow) which penalizes the network for generating the shadows on images where the shadow is already presented.

3.2. Learning from non-shadow images

Given the generator network $G_{f \rightarrow s}$ and also attention map A_s together with shadow mask M_s we can define corresponding losses for inverse transformation. We have the same adversarial loss where generator is maximizing the probability of discriminator to make mistake:

$$L_{GAN}^f = E_{x \sim X_s}[(D_s(x))^2] + E_{x \sim X_f}[(1 - D_s(G_{f \rightarrow s}(x)))^2] \quad (10)$$

However, we do not integrate the CAM module into the inverse transformation networks due to stability issues and because this approach gives better results in experiments.

The cycle-consistency constraint also stays the same: we generate shadow image from shadow-free X_f and then using $G_{s \rightarrow f}$ to restore the image back and optimize the networks:

$$L_{cycle}^{f \rightarrow s} = E_{x \sim X_f}[||G_{s \rightarrow f}(G_{f \rightarrow s}(x, A_s, M_s)) - x||_1] \quad (11)$$

Finally, we adopt the $G_{s \rightarrow f}$ to produce shadow free image given the real shadow free image from X_f . That means that we encourage the network to not remove anything if there is nothing to remove.

$$L_{idt}^f = E_{x \sim X_f}[||G_{s \rightarrow f}(x) - x||_1] \quad (12)$$

3.3. Maps generation

Shadow mask generation

Our generator $G_{f \rightarrow s}$ uses the shadow mask as the input, so we can condition network with it and generate multiple shadows from one shadow-free image. We follow the same approach as [9] and construct the threshold binarizer B between generated shadow free image \tilde{I}_f and original image I_s :

$$M_l = B(\tilde{I}_f, I_s) \quad (13)$$

Thus, when we obtain a pair of images we compute the difference $\tilde{I}_f - I_s$ and compute the threshold to assign the values greater than it as 1 and those less – with -1. The threshold is computed using Otsu's algorithm which separates the shadow from non-shadow regions by maximizing the intra-class variance.

Attention map generation

Attention map A_s is received from auxiliary classifier η_s by applying the pooling operation to feature maps. In our approach we are using average (GAP) together with max pooling (GMP) layers to get the complete picture. GAP is able to find all discriminative regions on the image while GMP is encouraged to find only one [26]. So we decided to combine the best from two worlds by applying both GAP and GMP and concatenating the corresponding results.

So, additionally we have:

$$\eta_m = \sigma(\sum_k w_s^k \max(C_s^k)) \quad (14)$$

After that we concatenate the outputs of η_s and η_m and feed them to 1×1 convolutional layer with the following ReLU non-linearity to restore the input dimensions due to channel axis concatenation.

The output from these operation is followed by aggregation to obtain an attention map A_s which we additionally scale to $[-1, 1]$ range for invariance.

During the training process those maps are generated for each shadow image in the same way as binary shadow masks.

3.4. Losses

To conclude, we present the final loss function which is a weighted sum of adversarial, CAM, cycle-consistency and identity losses outlined above in both architectures:

$$\begin{aligned} \min_{G_s \rightarrow f, G_f \rightarrow s, \eta_s, D_s, D_f, \eta_D} \max &= \lambda_{adv}(L_{GAN}^s + L_{GAN}^f) + \\ &+ \lambda_{cam}(L_{cam}^{s \rightarrow f} + L_{cam}^{D_f}) \\ &+ \lambda_{cycle}(L_{cycle}^{s \rightarrow f} + L_{cycle}^{f \rightarrow s}) + \lambda_{idt}(L_{idt}^s + L_{idt}^f) \end{aligned} \quad (15)$$

where $\lambda_{adv} = 1, \lambda_{cam} = 500, \lambda_{cycle} = 10, \lambda_{idt} = 5$.

3.5. Shadow removal generator network

The network architecture is following a Johnson et al. [11] and reminds the encoder-decoder architecture without skip-connections. Encoder is constructed from two down-sample convolutional layers, it is important that there are no pooling layers and down-sampling is implemented using convolutions with stride 2.

Then, we have a bottleneck layer where most of work takes place. It includes nine residual blocks

with linear dilation growth starting from the sixth layer. Dilation factors should be tuned depending on the receptive field size. In our experiments, we have made an assumption that the first part of the network would extract the shadow region operating on the local level while the second part will be responsible for filling this region, thus it will need a background information. For that reason we added receptive field growth at the end but in the way it does not exceed the input image size.

Bottleneck layer also integrates an attention CAM module that we described above. We inserted it before the receptive field growth (i.e. shadow removal process takes place) so it would help to localize the shadow in a more efficient way.

Finally, decoder is here to restore the image back to initial size by the use of transposed convolutions, it is important that network is learning to make the downsample and upsample operations itself.

Shadow free discriminator

We will remind that discriminator D_f network is used for discriminating the real shadow free images from those generated. Architecture for it is following the idea of PatchGAN [16] where the network is not looking at the whole image but on patches (usually, 70×70) of it deciding whether the patch is real or not. We additionally complemented it with CAM attention module which is trained to solve the same task as the discriminator itself. CAM attention is operating before the final layer. Discriminators are not using dilated convolutions.

Shadow generator network

This network is also following the Johnson et al. [11] architecture and has dilated convolutions in it which may help in shadow generation, however we have not seen any difference in experiments. We did not use an attention module here because it exposes an unstable training.

Generator uses shadow free image together with attention map and shadow mask (binary map) where three of them are concatenated by channel axis. Attention and mask are scaled to $[-1, 1]$ to improve invariance.

Discriminator D_s is also a PatchGAN with 70×70 patches with no attention module in it.

Training strategy

We used Rectified Linear Unit (ReLU) non-linearity and reflection padding for generator networks.

Instance normalization (IN) is utilized for all networks just after the convolutional layer. The exceptions are input and output layers where we want to encourage the networks to learn the normalization by themselves. We are adding hyperbolic tangent (Tanh) function in generator network to output the values from $[-1,1]$ range.

All networks are using spectral normalization [19] as this proved to improve GAN learning. It constrains the Lipschitz constant by restricting the spectral norm of each layer.

We also have used different heuristics to stabilize training by smoothing the target label from 1 to 0.9 and dividing the discriminator loss by 2 to further address slow learning. Moreover, we are updating the discriminator with the history of previously generated samples as it is stated to reduce the model oscillations [27]. The same approach is used for shadow mask, attention pairs which are added to a separate Queue.

All the parameters are initialized using zero-centered Gaussian distribution with 0.02 standard deviation. For data augmentation, we resize the images to 286×286 cropping them randomly to be 256×256 and flipping them horizontally with 0.5 probability. The method is implemented using PyTorch framework.

4. Evaluation

Dataset. There are many community datasets for shadow detection and removal. Mostly, they are supervised and not large enough for deep learning solution but there are some which fits just well: ISTD [24], SRD [14]. There is also an unpaired one where more complex scenes are presented USR [13].

In our experiments we only use ISTD dataset for evaluation clarity because it has ground truth shadow free images as well as shadow masks. We does not correct our evaluation method to cover the issues with illumination and color change between shadow and shadow free images. In future, we will extend our solution to USR dataset as being more appropriate to our solution.

Metrics. In our evaluations we aim to estimate how good our network is in removing the shadows as well as detecting them. We are also concerned about the global image consistency. That is why we present three metrics.

For shadow removal we follow recent works [14; 9; 4] and use RMSE (root mean squared error)

between generated and real shadow free images in LAB color space. Evaluation is divided into region and global where the former is applied to shadow regions while the latter to the whole images. In general, lower RMSE score tells about better results.

Shadow detection is evaluated with the use of IOU (Intersection over Union), also known as the Jaccard index, which is a widely used metric for image segmentation and object detection tasks. It is computed between the generated and ground-truth binary shadow masks by dividing the area of overlap by the union of those two. Greater IOU indicates better shadow detection.

5. Experimental results

Our method is trained on supervised dataset that is why the unpaired strategy is used: the first one is sampled from shadow domain while the second one is randomly chosen from the shadow free. We also selected random 100 images for validation purposes. During the experiments multiple hypothesis were tested and the most successful of them are shown in Table 1. Global RMSE tells the quality of shadow removal coupled with background restoration. Shadow region RMSE directly estimates the removal comparing shadow regions. IOU is aimed to evaluate the shadow detection quality comparing with ground-truth shadow mask. All scores are evaluated after the models are trained on validation set of images. Other solutions from the field except the MaskShadowGAN [9] are not included but would be tested in the future.

As soon as we added two components to this work we wanted to test how they affect the generation quality.

We saw that method with dilated convolutions stays on roughly the same level as the one without them. However, we decided to take the one with dilated convolution as we expected it may improve the results after the attention would be added.

The majority of tests we conducted were about how to use an attention module and where to localize it. At first, we have added the attention to all networks following the original approach [13]. The results improved drastically given by RMSE score in both global and local contexts.













Table 1

Quantitative comparison of algorithms

Methods	Global RMSE	Shadow region RMSE	IOU
MaskShadowGAN [26]	3.0099	23.6703	80.0894
Dilated generators dilation in $G_{S \rightarrow f}$ and $G_{f \rightarrow S}$	3.1253	26.9753	75.5382
CAM attention for all networks in all G and D	2.3902	15.0150	73.8277
CAM attention for shadow-removal networks in $G_{S \rightarrow f}$ and D_f giving A_S as input to D_S (*)	2.3139	15.7047	71.1142
CAM attention for shadow-removal networks. Shadow generator using attention map and binary mask A_S with no M_S to (*)	2.3261	15.5398	70.3951
CAM attention for shadow-removal networks. Shadow generator uses weights from shadow removal for re-balancing CAM weights from $G_{S \rightarrow f}$ to $G_{f \rightarrow S}$	2.2436	15.6087	70.7965

Table 2

Examples of images

Shadow image	Ground truth	Initial	CAM attention
			
			
			

However, IOU showed lower results for the reason we will go into details below. Adding attention module presented new instability issues while converging faster. We also encountered with messy outputs compared to initial methods where the shadow removal generator not only identified the shadow but the background behind it. Thus, it could be seen that

shadow detection quality decreased significantly indicated by lower IOU.

For these reasons, we removed an attention module from $G_{f \rightarrow S}$ and D_S which helped to suppress the instability but still it was a way higher than solutions without attention. The qualitative results improved a little but still there was a problem with over-detection.

Hence, the next solution added an attention map transfer from shadow removal generator to the input of the opposite one. We aimed to complement the binary mask with an attention map to make the learning more consistent. As the result, it reduced the model oscillations and raised the generated samples quality. However the shadow detection performance declined even more.

Attention map transfer showed the capacity to improve the results that is why we researched other ways to share this information with shadow generator network. For instance, shadow mask M_s was removed from the input of $G_{f \rightarrow s}$ so the network should have used the information from attention map A_s only. This resulted in again messy results with roughly the same metric values.

Rather than integrating the attention information to the input of the network, we researched the ways to insert it inside. After solid amount of experimenting we came up with transferring the CAM weights directly to rebalance the corresponding feature maps in $G_{f \rightarrow s}$. This improved the

shadow removal performance, however increased the problem with over-detection.

6. Conclusions

In this paper, we presented a solution to unsupervised shadow removal problem with the use of generative adversarial networks with attention modules and multi context aggregation. Our network produces better results compared to the existing approach in the field. Analysis showed that attention maps obtained from auxiliary classifier encourage the networks to concentrate on more distinctive regions between domains. However, GANs demand more accurate and consistent architecture to solve the problem in a more efficient way. We have also showed how attention modules can improve the quality of shadow removal while introducing the problems with the shadow over-detection. For that reason we will research further to address the problem of more consistent architecture in the future work.

References

- Almahairi, A., Rajeshwar, S., Sordoni, A., & other. (2018). Courville. Augmented CycleGAN: Learning many-to-many mappings from unpaired data. In *ICML* (pp. 195–204).
- Barrow, H., & Tenenbaum, J. (1978). *Recovering intrinsic scene characteristics from images*. Comp. Vision Systems.
- Chuang, Y., Goldman, D. B., & other. (2003). Shadow matting and compositing. *ACM ToG*, 22 (3), 494–500.
- Cucchiara, R., Grana, C., Piccardi, M., & other. (2001). Improving shadow suppression in moving object detection with HSV color information. In *IEEE Intelligent Transportation Systems* (pp. 334–339). Proceedings, Oakland, CA.
- Ding, B., Long, C., Zhang, L., & Xiao, C. (2019). ARGAN: Attentive Recurrent Generative Adversarial Network for Shadow Detection and Removal. In *IEEE/CVF International Conference on Computer Vision (ICCV), Seoul, Korea (South)* (pp. 10212–10221).
- Goodfellow, I. J., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., & Bengio, Y. (2014). Generative adversarial nets. In *NIPS*.
- Guo, R., Dai, Q., & Hoiem, D. (2013). Paired Regions for Shadow Detection and Removal. In *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 35 (12), 2956–2967.
- Hu, X., Fu, C., Zhu, L., Qin, L., & Heng, P. (2019). Direction-aware Spatial Context Features for Shadow Detection and Removal. In *IEEE Transactions on Pattern Analysis and Machine Intelligence*.
- Hu, X., Jiang, Y., Fu, C., & Heng, P. (2019). Mask-Shadow GAN: Learning to Remove Shadows From Unpaired Data. In *IEEE/CVF International Conference on Computer Vision (ICCV)* (pp. 2472–2481). Seoul, Korea (South).
- Isola, P., Zhu, J., Zhou, T., & Efros, A. A. (2017). Image-to-Image Translation with Conditional Adversarial Networks. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (pp. 5967–5976).
- Johnson, J., Alahi, A., & Fei-Fei, Li. (2016). Perceptual losses for real-time style transfer and super-resolution. In *ECCV* (pp. 694–711).
- Karras, T., Laine, S., & Aila, T. (2019). A Style-Based Generator Architecture for Generative Adversarial Networks. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* (pp. 4396–4405). Long Beach, CA, USA.
- Kim, J., Kim, M., Kang, H., & Lee, K. (2020). U-gat-it: Unsupervised generative attentional networks with adaptive layer-instance normalization for image-to-image translation. In *ICLR*.
- Lee, Hsin-Ying, Tseng, Hung-Yu, Huang, Jia-Bin, Singh, Maneesh, Yang, & Ming-Hsuan. (2018). Diverse image-to-image translation via disentangled representations. In *ECCV* (pp. 35–51).
- Li, Z., & Snavely, N. (2018). Learning Intrinsic Image Decomposition from Watching the World. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 9039–9048). Salt Lake City, UT.
- Maxwell, B. A., Friedhoff, R. M., & Smith, C. A. (2008). An illuminant dichromatic reflection model for understanding images. In *CVPR*.
- Mikic, I., Cosman, P. C., & other. (2000). Moving shadow and object detection in traffic scenes. In *Proceedings 15th International Conference on Pattern Recognition. ICPR-2000* (Vol. 1, pp. 321–324). Barcelona, Spain.
- Mirza, M., & Osindero, S. (2014). Conditional generative adversarial nets. *arXiv preprint arXiv*.
- Miyato, T., Kataoka, T., Koyama, M., & Yoshida, Y. (2018). Spectral normalization for generative adversarial networks. In *ICLR*.
- Pathak, D., Krahenbuhl, P., Donahue, J., & other. (2016). Context encoders: Feature learning by inpainting. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*.
- Qu, L., Tian, J., He, S., & other. (2017). *DeshadowNet: A Multi-context Embedding Deep Network for Shadow Removal*. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (pp. 2308–2316).
- Shin, Y., Sagong, M., Yeo, Y., Kim, S., & Ko, S. (2020). PEPSI+: Fast and Lightweight Network for Image Inpainting. In *IEEE Transactions on Neural Networks and Learning Systems*.

23. Waltz, D. L. (1972). *Generating semantic descriptions from drawings of scenes with shadows*. Cambridge, MA, USA, Tech. Rep.
24. Wang, J., Li, X., & Yang, J. (2018). Stacked Conditional Generative Adversarial Networks for Jointly Learning Shadow Detection and Shadow Removal. *IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 1788–1797). Salt Lake City.
25. Xiaodong, Cun, Chi-Man, Pun, & Cheng, Shi1. (2020). *Towards Ghost-free Shadow Removal via Dual Hierarchical Aggregation Network and Shadow Matting GAN*. Association for the Advancement of Artificial Intelligence.
26. Zhou, B., Khosla, A., Lapedriza, A., Oliva, A., & Torralba, A. (2016). Learning Deep Features for Discriminative Localization. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (pp. 2921–2929). Las Vegas, NV.
27. Zhu, J., Park, T., Isola, P., & Efros, A. A. (2017). Unpaired Image-to-Image Translation Using Cycle-Consistent Adversarial Networks. In *IEEE International Conference on Computer Vision (ICCV)* (pp. 2242–2251). Venice.

Андронік В. П., Бучко О. А.

ВИДАЛЕННЯ ТІНЕЙ НА ЗОБРАЖЕННІ З ВИКОРИСТАННЯМ ГЕНЕРАТИВНИХ ЗМАГАЛЬНИХ МЕРЕЖ

Точне виявлення тіні на зображенні є складним завданням, оскільки досить важко зрозуміти, чи затемнення або сірий колір є причиною тіні. У цій статті запропоновано метод видалення тіней на зображенні з використанням генеративних змагальних нейронних мереж. Навчання мережі відбувається без нагляду, тобто не залежить від трудомісткого збирання даних і маркування даних. Метод видалення тіней на зображенні базується на методі невідконтрольного передання зображень між різними доменами. Було використано дві мережі: першу – для додавання тіней у зображення, а другу – для видалення тіней. Набір даних ISTD використовували для чіткості оцінювання, оскільки він містить основні зображення, що не мають тіні, а також тінюві маски.

Ключові слова: генеративні змагальні мережі, видалення тіней, генерування тіней.



Creative Commons Attribution 4.0 International License (CC BY 4.0)

Матеріал надійшов 10.06.2020