



C²ShadowGAN: cycle-in-cycle generative adversarial network for shadow removal using unpaired data

Sunwon Kang¹ · Juwan Kim² · In Sung Jang² · Byoung-Dai Lee¹

Accepted: 15 October 2022 / Published online: 8 November 2022
© The Author(s) 2022

Abstract

Recent advances in deep learning technology, and the availability of public shadow image datasets, have enabled significant performance improvements of shadow removal tasks in computer vision. However, most deep learning-based shadow removal methods are usually trained in a supervised manner, in which paired shadow and shadow-free data are required. We developed a weakly supervised generative adversarial network with a cycle-in-cycle structure for shadow removal using unpaired data. In addition, we introduced new loss functions to reduce unnecessary transformations for non-shadow areas and to enable smooth transformations for shadow boundary areas. We conducted extensive experiments using the ISTD and Video Shadow Removal datasets to assess the effectiveness of our methods. The experimental results show that our method is superior to other state-of-the-art methods trained on unpaired data.

Keywords Deep learning · Generative adversarial network · Shadow removal · Unpaired data

1 Introduction

Shadows are common in most natural images. They can be categorized into cast and self-shadows depending on the source. Cast shadows are caused by tall objects in the vicinity that block the light source, whereas self-shadows arise from object surfaces that are not directly illuminated by light sources [1]. The published literature in the field of computer vision and graphics provides evidence that the presence of shadows negatively affects the processing, analyses, and understanding of images [1–7]. Therefore, realistic shadow removal is an essential task for improving the performance of many computer vision objectives, such as image segmentation, object detection, and tracking. Early shadow removal approaches were mostly based on the development of physical models for analyzing the statistics of color and illumination and used hand-crafted features [7–9]. However, these approaches failed for complex images [10]. In recent years, public shadow image datasets, such as ISTD [11],

SBU [12], and USR [2], have enabled learning-based methods, particularly those using deep learning, to achieve state-of-the-art results for shadow removal. Current deep learning-based shadow removal methods are typically trained in a supervised manner, in which pairs of shadow and shadow-free images of identical scenes are required for learning to remove shadows. However, paired training samples are expensive to collect. They also have limitations, such as lack of diversity in terms of the collectible scenes as well as inconsistent color and luminosity between the paired images [6, 13, 14].

Several methods and algorithms have been proposed to address the limitations of paired datasets for learning-based shadow removal. On one end of the spectrum are approaches for generating large-scale, diverse synthetic shadow and shadow-free image pairs that apply various techniques, ranging from simple affine transformation to deep generative models [4, 5, 14–16]. Although these approaches help improve the performances of shadow removal networks by enriching the training samples, the quality and diversity of the generated shadow images are still limited. For instance, synthetic shadow images from the SynShadow dataset [14] are constrained by such assumptions as occluded objects being outside the camera view, and float surfaces for shadow projection. At the other end of the spectrum, a group of researchers learned to remove shadows from unpaired shadow and shadow-free images by employing cycle consistency and a generative adversarial network (GAN) [17] to translate the shadow images into shadow-free images [6, 7, 18]. These

Sunwon Kang and Juwan Kim contributed equally to this work.

✉ Byoung-Dai Lee
blee@kgu.ac.kr

¹ Division of AI & Computer Engineering, Kyonggi University, Suwon 16227, South Korea

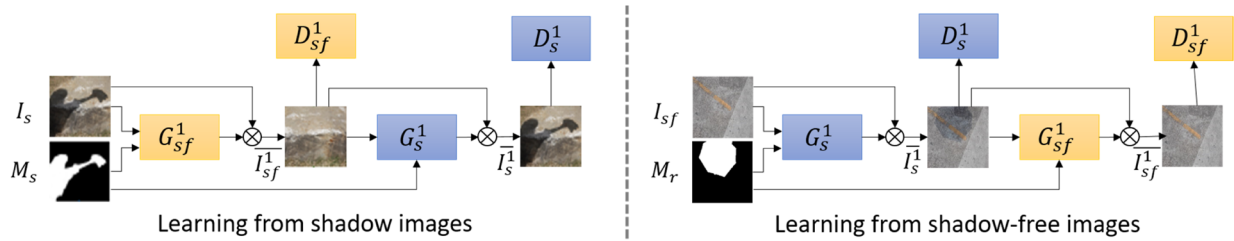
² City & Transportation ICT Research Department, Electronics and Telecommunications Research Institute, Daejeon 34129, South Korea

approaches are inferior (performance wise) to deep learning models trained in a supervised manner [14]. Therefore, there is considerable room for improvement [18].

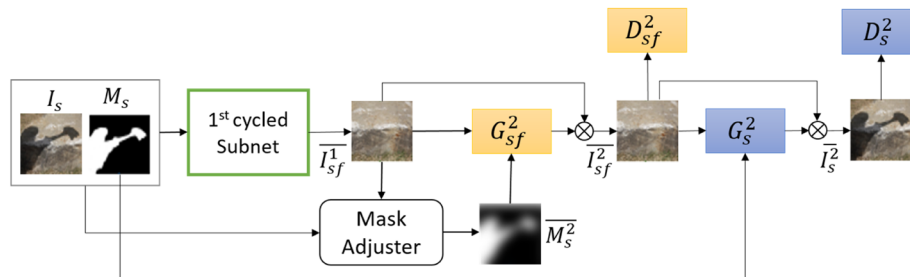
We combined the advantages of both approaches and established a novel weakly supervised GAN with a cycle-in-cycle structure for removing shadows using unpaired data, which we call the C²ShadowGAN. Our method exploits the cycle consistency constraint based on the cycle-in-cycle structure, in which multiple cycled subnetworks are stacked, to learn to remove the shadows (Fig. 1). Specifically, given an input shadow image and the corresponding shadow mask—that is, zeros which denote non-shadow pixels and ones which denote shadow pixels—the first cycled subnetwork for shadow removal learns to remove the shadows from the input images and then generates realistic shadow-free images. The resulting shadow-free images and their

corresponding auxiliary information are used in the second cycled subnetwork where the shadow-free images are refined. The auxiliary information indicates whether shadows in the input images are either removed or attenuated in the previous step. Therefore, similar to shadow masks, the auxiliary information plays the role of guiding the learning for the refinement tasks. In this manner, we can linearly stack zero or more cycled subnetworks for refinement. Each cycled subnetwork is responsible for refining the intermediate shadow-free image generated by its preceding cycled subnetwork. The entire network is thus jointly trained using adversarial learning in an end-to-end manner.

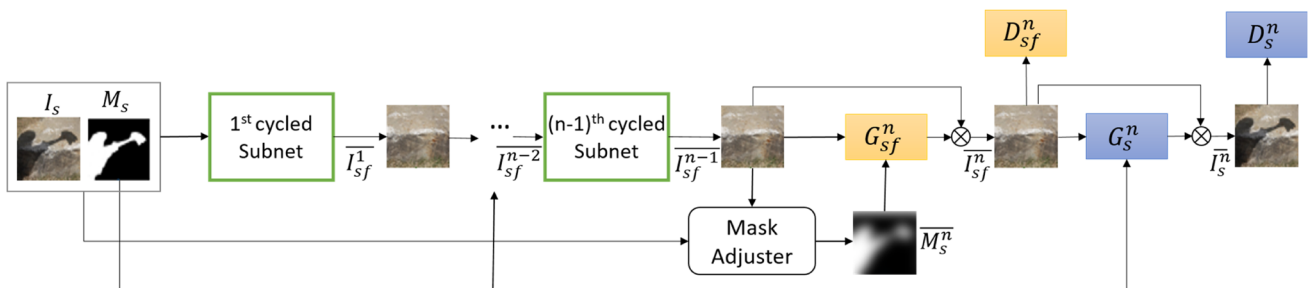
One of the limitations of the shadow removal system based on the cycle consistency constraint, such as the Mask-ShadowGAN [6], is that sufficient statistical similarity between two image domains is required [19, 20]. We adopted



(a) Network architecture of the first cycled subnetwork for shadow removal



(b) Network architecture of C²ShadowGAN composed of the first cycled subnetwork for shadow removal and second cycled subnetwork for refinement



(c) Generalized network architecture of C²ShadowGAN composed of a cycled subnetwork for shadow removal and (n-1) cycled subnetworks for refinement

Fig. 1 Network architecture of the proposed C²ShadowGAN. G_{sf}^i and G_s^i denote generators in the i^{th} cycled subnetwork to produce shadow-free and shadow images. D_{sf}^i and D_s^i denote discriminators in the i^{th} cycled subnetwork to determine whether the generated images are real shadow-free or shadowed images. \bar{I}_{sf}^i and \bar{I}_s^i denote generated shadow-free and

shadow images in the i^{th} cycled subnetwork. M_s and M_r are shadow masks. M_s^i is a difference map that provides the information about the difference between the original input shadow image and the output shadow-free image from the $(i-1)^{\text{th}}$ step

the approach by Le et al. [13] to address this issue; here, the training dataset was prepared by cropping shadow and non-shadow patches from the same image to construct unpaired data for network training. Thus, we ensured significant statistical similarity. We then trained the proposed shadow removal system with this training set to learn mapping from patches in the shadow set to patches in the non-shadow set, which is considered an image-to-image translation task.

We conducted extensive experiments to assess the effectiveness of our approach using the ISTD [11] and Video Shadow Removal [13] datasets. The experimental results show that C²ShadowGAN is stable during training and converges faster. In addition, we demonstrate that our method achieves quantitatively and qualitatively competitive performances as compared with state-of-the-art methods.

The main contributions of this work are as follows:

- We propose a weakly unsupervised single image shadow removal system based on the cycle-in-cycle structure, in which multiple cycled subnetworks can be stacked linearly to learn to remove shadows.
- We introduce new loss functions to reduce unnecessary transformations for non-shadow areas and to enable smooth transformations of the shadowed boundary areas.
- We conducted experiments using public datasets and demonstrated that the proposed C²ShadowGAN could achieve comparable performance to state-of-the-art methods.

2 Related works

Shadow removal is an essential task in improving the performances of many computer vision tasks and has been heavily studied in recent times. Our review of the related research focuses primarily on methods involving deep learning-based shadow removal because the objective of this study is to investigate these methods. Comprehensive surveys on shadow detection and removal methods can be found in previously published literature [21–24].

Hu et al. [6] proposed the Mask-ShadowGAN for learning to remove shadows from unpaired training data by extending CycleGAN [25]; they modified CycleGAN to learn the underlying relationships between the shadow and shadow-free domains with the guidance of shadow masks, which are also learned from shadow images automatically. The Mask-ShadowGAN is the first data-driven shadow removal method that uses unpaired data for training. The method proposed by Vasluianu et al. [7] was similar to the Mask-ShadowGAN. Both these methods were based on the vanilla CycleGAN approach. However, they formulated a component in the training objective to generate more sophisticated synthetic shadow

masks; instead of shadow masks being computed as binarized differences between the real shadow images and generated shadow-free images. In addition, they used perceptual losses rather than pixelwise fidelity losses. Liu et al. [18] developed the LG-ShadowNet to improve the performance of the Mask-ShadowGAN by introducing a new lightness guided strategy; the core aspect of this approach was to learn the lightness information from the input images by separate training and by using this information to guide the learning of shadow removal. All these methods required shadow-free images. In addition, a small domain difference was required between the unpaired shadow and the shadow-free images for stable learning, making it challenging to acquire shadow-free images in some cases.

Le et al. [13] proposed a patch-based shadow removal system to prevent the dependency on paired training data, where unpaired patches were cropped from the same image used for the network training. In addition, they introduced three different deep neural networks to learn a set of physics-based constraints that define a transformation closely modeling shadow removal. The G2R-ShadowNet proposed by Liu et al. [16] addressed the issues related to the patch-based shadow removal system, such as heavy computational load and strict physics-based constraints. They constructed paired shadow and non-shadow images using only shadow images and their corresponding masks to form training data. The shadow removal subnetwork removes the shadows from the images, and the shadow refinement subnetwork refined intermediate shadow-free images by leveraging contextual information. Since both methods generated synthetic training data using the same shadow images, their domain gaps were small and well-controlled.

In contrast to the CycleGAN-based shadow removal methods mentioned above, our method introduces a novel cycle-in-cycle structure. Multiple cycled subnetworks are stacked linearly and jointly trained in our approach. In addition, our method eliminates the weakness of the CycleGAN-based systems by adopting a state-of-the-art patch-based training strategy, where unpaired data are used for network training.

3 Methodology

According to previous observations [2, 26, 27], a shadow image I_s can be generated from the pixelwise product of a shadow-free image I_{sf} and a shadow matte α , as shown in (1).

$$I_s = \alpha \otimes I_{sf} \quad (1)$$

Similarly, from (1), we can deduce that a shadow-free image I_{sf} can be considered a pixelwise product of a shadow

image I_s and another shadow matte β . Thus, we used shadow matting instead of generating the shadow-free image directly for both patch-level and image-level shadow removal. In particular, both shadow mattes (α and β) are learned by the cycle consistency constraints and adversarial training of the proposed C²ShadowGAN.

3.1 Cycled subnetwork for shadow removal

The first cycled subnetwork of the C²ShadowGAN is based on the Mask-ShadowGAN approach with two generators and two discriminators for the shadow and shadow-free domains, respectively. In detail, the shadow image I_s is transformed to the shadow-free image \overline{I}_{sf}^1 by the generator G_{sf}^1 , which is further transformed to the shadow image \overline{I}_s^1 by the generator G_s^1 , as illustrated on the left-hand side of Fig. 1a. Similarly, the shadow-free image I_{sf} is transformed to the shadowed image \overline{I}_s^1 by the generator G_s^1 , which is further transformed to the shadow-free image \overline{I}_{sf}^1 by the generator G_{sf}^1 , as illustrated on the right-hand side of Fig. 1a. Note that both generators are trained to produce a shadow matte that is multiplied in a pixelwise manner with the input shadowed image for the shadow removal or with the input shadow-free image for the shadow addition. Furthermore, shadow masks M_s and M_r are used to guide the shadow removal and shadow addition. The shadow mask M_s corresponds with the shadowed areas of the input image, which can be obtained either manually, semi-interactively, or automatically using shadow detection methods [13]. The shadow mask M_r is randomly selected from the masks of the training set. The discriminators D_{sf}^1 and D_s^1 learn to distinguish between the synthetic shadow-free and shadowed images (e.g., \overline{I}_{sf}^1 and \overline{I}_s^1) and the randomly selected real shadow-free and shadowed images, helping generators G_{sf}^1 and G_s^1 to produce better outputs. The adversarial losses to optimize the generator G_{sf}^1 and the discriminator D_{sf}^1 for shadow removal and the generator G_s^1 and discriminator D_s^1 for shadow addition are given as

$$\begin{aligned} L_{GAN}^1(G_{sf}^1, D_{sf}^1, G_s^1, D_s^1) &= L_{GAN}^{sf,1}(G_{sf}^1, D_{sf}^1) + L_{GAN}^{s,1}(G_s^1, D_s^1) \\ L_{GAN}^{sf,1}(G_{sf}^1, D_{sf}^1) &= \frac{1}{N} \sum -[\log(D_{sf}^1(I_{sf})) + \log(1 - D_{sf}^1(G_{sf}^1(I_s, M_s) \otimes I_s))] \\ L_{GAN}^{s,1}(G_s^1, D_s^1) &= \frac{1}{N} \sum -[\log(D_s^1(I_s)) + \log(1 - D_s^1(G_s^1(I_{sf}, M_r) \otimes I_{sf}))], \end{aligned} \quad (2)$$

where N represents the number of training samples.

To preserve the cycle consistency between the input and reconstructed images, the network is trained to ensure that $G_s^1(G_{sf}^1(I_s, M_s), M_s)$ is identical to the shadowed input image I_s

and $G_{sf}^1(G_s^1(I_{sf}, M_r), M_r)$ is identical to the shadow-free input image I_{sf} :

$$\begin{aligned} L_{cycle}^1(G_{sf}^1, G_s^1) &= L_{cycle}^{sf,1}(G_{sf}^1, G_s^1) + L_{cycle}^{s,1}(G_s^1, G_{sf}^1) \\ L_{cycle}^{sf,1}(G_{sf}^1, G_s^1) &= \frac{1}{N} \sum (\|G_s^1(G_{sf}^1(I_s, M_s) \otimes I_s, M_s) \otimes \overline{I}_{sf}^1 - I_s\|_1) \\ L_{cycle}^{s,1}(G_s^1, G_{sf}^1) &= \frac{1}{N} \sum (\|G_{sf}^1(G_s^1(I_{sf}, M_r) \otimes I_{sf}, M_r) \otimes \overline{I}_s^1 - I_{sf}\|_1), \end{aligned} \quad (3)$$

where $\|\cdot\|_1$ represents L₁ loss.

Furthermore, G_{sf}^1 is regularized to produce the output \overline{I}_{sf}^1 that is close to the shadow-free input image I_{sf} with the guidance of the all-zero shadow-free mask M_0 . Similarly, using the shadow mask M_0 and shadow input image I_s , G_s^1 is trained to generate the shadowed image \overline{I}_s^1 , which contains no newly added shadows:

$$\begin{aligned} L_{identity}^1(G_{sf}^1, G_s^1) &= L_{identity}^{sf,1}(G_{sf}^1) + L_{identity}^{s,1}(G_s^1) \\ L_{identity}^{sf,1}(G_{sf}^1) &= \frac{1}{N} \sum (\|G_{sf}^1(I_{sf}, M_0) \otimes I_{sf} - I_{sf}\|_1) \\ L_{identity}^{s,1}(G_s^1) &= \frac{1}{N} \sum (\|G_s^1(I_s, M_0) \otimes I_s - I_s\|_1) \end{aligned} \quad (4)$$

In addition to the losses described thus far, we introduce the non-shadow-area loss (i.e., L_{nsa}) to reduce unnecessary transformation of the non-shadow areas; and boundary loss (i.e., L_{ba}) to enable smooth transformation of the shadowed boundary areas (BAs), such as the penumbra areas. The umbra, non-shadow, and penumbra areas can be roughly identified with the shadow mask given. Therefore, we train the generators G_{sf}^1 and G_s^1 so that the non-shadow areas of the reconstructed images \overline{I}_{sf}^1 (i.e., $G_{sf}^1(I_s, M_s)$) and \overline{I}_s^1 (i.e., $G_s^1(I_{sf}, M_r)$), which are indicated by the corresponding shadow masks M_s and M_r , are identical to those of their input images I_s and I_{sf} respectively:

$$\begin{aligned} L_{nsa}^1(G_{sf}^1, G_s^1) &= L_{nsa}^{sf,1}(G_{sf}^1) + L_{nsa}^{s,1}(G_s^1) \\ L_{nsa}^{sf,1}(G_{sf}^1) &= \frac{1}{N} \sum \left(\frac{1}{|NSA|} \sum_{i \in NSA} \|p(I_s, i) - p(G_{sf}^1(I_s, M_s) \otimes I_s, i)\|_1 \right) \\ L_{nsa}^{s,1}(G_s^1) &= \frac{1}{N} \sum \left(\frac{1}{|NSA|} \sum_{j \in NSA} \|p(I_{sf}, j) - p(G_s^1(I_{sf}, M_r) \otimes I_{sf}, j)\|_1 \right) \end{aligned} \quad (5)$$

where $p(I, x)$ represents a pixel value at position x in image I , and $|NSA|$ denotes the number of pixels in the non-shadow areas according to the shadow masks (M_s and M_r). The

objective of L_{nsa} loss is similar to that of Le et al. [13]. In the case of a non-shadow pixel, both approaches force the pixel value in the output image to be equal to that in the input image by controlling the shadow mattes. However, unlike the approach of Le et al., in which the values of the shadow matte are manipulated directly, we apply the reconstruction error to the non-shadow areas between the two images. Thus, we en-

sure a more natural overall output image. This is a similar effect to those of $L_{mat - \alpha}$ [13] and $L_{sm - \alpha}$ [13].

The shadow effects are assumed to vary smoothly across the shadow boundaries. Thus, we applied local variation regularization on the shadow boundary areas, which are defined as areas within B_{step} pixels from the shadow boundaries in the shadow mask:

$$\begin{aligned}
 L_{ba}^1(G_{sf}^1, G_s^1) &= L_{ba}^{sf,1}(G_{sf}^1) + L_{ba}^{s,1}(G_s^1) \\
 L_{ba}^{sf,1}(G_{sf}^1) &= \frac{1}{N} \sum \left(\frac{1}{|BA|} \sum_{i \in BA} \left(\|\nabla_{hp}(G_{sf}^1(I_s, M_s) \otimes I_s, i)\|_1 + \|\nabla_{wp}(G_{sf}^1(I_s, M_s) \otimes I_s, i)\|_1 \right) \right) \\
 L_{ba}^{s,1}(G_s^1) &= \frac{1}{N} \sum \left(\frac{1}{|BA|} \sum_{j \in BA} \left(\|\nabla_{hp}(G_s^1(I_{sf}, M_r) \otimes I_{sf}, j)\|_1 + \|\nabla_{wp}(G_s^1(I_{sf}, M_r) \otimes I_{sf}, j)\|_1 \right) \right)
 \end{aligned} \tag{6}$$

where $\nabla_h(\cdot)$ and $\nabla_w(\cdot)$ are operations to compute the horizontal and vertical gradients of a given pixel, and $|BA|$ denotes the number of pixels in the shadow boundary areas. When B_{step} is set to the image size, it is equal to the total variation regularization. In this study, the configurable parameter B_{step} was set to 2 for all experiments. The sensitivity to B_{step} will be presented in a later section. In summary, the final objective loss for the first cycled subnetwork for shadow removal is the weighted sum of the five loss functions.

$$L_{total}^1 = \omega_1^1 L_{GAN}^1 + \omega_2^1 L_{cycle}^1 + \omega_3^1 L_{identity}^1 + \omega_4^1 L_{nsa}^1 + \omega_5^1 L_{ba}^1 \tag{7}$$

where $\omega_i^1, i \in \{1, \dots, 5\}$ controls the relative importance of each loss. We follow the previously reported result [25] and empirically set $\omega_1^1, \omega_2^1, \omega_3^1, \omega_4^1, \omega_5^1$ as 1.0, 10.0, 5.0, 10.0, and 5.0, respectively.

3.2 Cycled subnetworks for refinement

Although the first cycled subnetwork is aimed at realistic shadow removal, some residual shadows or blurry regions may remain in the generated shadow-free images. Therefore, we use multiple cycled subnetworks to refine the shadow-free images step by step, and every step introduces a new larger cycle that encompassed the previous one. Cycled subnetworks for refinement consist of two generators and two discriminators for the shadowed and shadow-free domains, similar to the cycled subnetwork for shadow removal. Learned shadow mattes are used to generate synthetic shadow-free and shadowed images. All additional cycled subnetworks for refinement have the same architecture.

For the sake of brevity, we explain an instance of C²ShadowGAN composed of a subnetwork for shadow removal and another subnetwork for refinement, as shown in Fig. 1b. The generator G_{sf}^2 is trained to generate a shadow matte for transforming the shadow-free image \overline{I}_{sf}^1 generated by the first cycled subnetwork to the refined shadow-free image \overline{I}_{sf}^2 , which is indistinguishable to the discriminator D_{sf}^2 between the synthetic and real shadow-free images. The model can be confused if the original shadow mask M_s is used as auxiliary information for network training in the refinement task, as in the first cycled subnetwork. For instance, the shadow mask eventually provides false information if the generated shadow-free image is close to the real shadow-free image. Therefore, the auxiliary information for the refinement task, called the difference map, needs to provide information on how well the shadow was removed along with the shadow's location information. For this purpose, we adopt a simple assumption that the smaller the differences in the shadow pixels between the input image and generated shadow-free image, the less is the shadow removed. Thus, $\overline{M}_s^2[i, j]$ is computed as $\overline{M}_s^2[i, j] = \min\left(\frac{1}{\left(|I_s[i, j] - \overline{I}_{sf}^1[i, j]|\right) + \epsilon}\right), 1.0$ if $M_s[i, j] = 1$. Otherwise, $\overline{M}_s^2[i, j] = 0$. Furthermore, adversarial learning for G_{sf}^2 alone may generate artifacts in the generated shadow-free images [6]. Therefore, another generator G_s^2 is employed to generate a shadow matte for transforming the generated shadow-free image \overline{I}_{sf}^2 back to the original shadow image I_s . This also ensures cycle consistency between the original and reconstructed shadow images. The introduction of discriminators D_{sf}^2 and D_s^2 enables adversarial training for both generators (G_{sf}^2 and G_s^2), which also affects the optimization of the first cycled subnetwork. Thus, we formulate the

adversarial, cycle consistency, identity, non-shadowed area, and boundary area losses for the second cycled subnetwork as follows:

$$\begin{aligned}
 L_{GAN}^1(G_{sf}^2, D_{sf}^2, G_s^2, D_s^2) &= L_{GAN}^{sf,2}(G_{sf}^2, D_{sf}^2) + L_{GAN}^{s,2}(G_s^2, D_s^2) \\
 L_{GAN}^{sf,2}(G_{sf}^2, D_{sf}^2) &= \frac{1}{N} \sum -[\log(D_{sf}^2(I_{sf})) + \log(1 - D_{sf}^2(G_{sf}^2(\overline{I}_{sf}^1, \overline{M}_s^2) \otimes \overline{I}_{sf}^1))] \\
 L_{GAN}^{s,2}(G_s^2, D_s^2) &= \frac{1}{N} \sum -[\log(D_s^2(I_s)) + \log(1 - D_s^2(G_s^2(\overline{I}_{sf}^2, M_s) \otimes \overline{I}_{sf}^2))]
 \end{aligned} \tag{8}$$

$$\begin{aligned}
 L_{cycle}^2(G_{sf}^2, G_s^2) &= L_{cycle}^{sf,2}(G_{sf}^2, G_s^2) \\
 L_{cycle}^{sf,2}(G_{sf}^2, G_s^2) &= \frac{1}{N} \sum (\|G_s^2(\overline{I}_{sf}^2, M_s) \otimes \overline{I}_{sf}^2 - I_s\|_1)
 \end{aligned} \tag{9}$$

$$\begin{aligned}
 L_{identity}^2(G_{sf}^2, G_s^2) &= L_{identity}^{sf,2}(G_{sf}^2) + L_{identity}^{s,2}(G_s^2) \\
 L_{identity}^{sf,2}(G_{sf}^2) &= \frac{1}{N} \sum (\|G_{sf}^2(\overline{I}_{sf}^1, M_0) \otimes \overline{I}_{sf}^1 - I_{sf}\|_1) \\
 L_{identity}^{s,2}(G_s^2) &= \frac{1}{N} \sum (\|G_s^2(I_s, M_0) \otimes I_s - I_s\|_1)
 \end{aligned} \tag{10}$$

$$\begin{aligned}
 L_{ba}^2(G_{sf}^2, G_s^2) &= L_{ba}^{sf,2}(G_{sf}^2) + L_{ba}^{s,2}(G_s^2) \\
 L_{ba}^{sf,2}(G_{sf}^2) &= \frac{1}{N} \sum \left(\frac{1}{|BA|} \sum_{i \in BA} (\|\nabla_{hp} (G_{sf}^2(\overline{I}_{sf}^1, \overline{M}_s^2) \otimes \overline{I}_{sf}^1, i)\|_1 + \|\nabla_{wp} (G_{sf}^2(\overline{I}_{sf}^1, \overline{M}_s^2) \otimes \overline{I}_{sf}^1, i)\|_1) \right) \\
 L_{ba}^{s,2}(G_s^2) &= \frac{1}{N} \sum \left(\frac{1}{|BA|} \sum_{j \in BA} (\|\nabla_{hp} (G_s^2(\overline{I}_{sf}^2, M_s) \otimes \overline{I}_{sf}^2, j)\|_1 + \|\nabla_{wp} (G_s^2(\overline{I}_{sf}^2, M_s) \otimes \overline{I}_{sf}^2, j)\|_1) \right),
 \end{aligned} \tag{12}$$

$$\begin{aligned}
 L_{total}^2 &= \omega_1^2 L_{GAN}^2 + \omega_2^2 L_{cycle}^2 + \omega_3^2 L_{identity}^2 + \omega_4^2 L_{nsa}^2 \\
 &\quad + \omega_5^2 L_{ba}^2
 \end{aligned} \tag{13}$$

where $\omega_i^2, i \in \{1, \dots, 5\}$ are the weights of the corresponding loss functions. Similar to the first cycled subnetwork, we empirically set $\omega_1^2, \omega_2^2, \omega_3^2, \omega_4^2, \omega_5^2$ as 1.0, 10.0, 5.0, 10.0, and 5.0, respectively. We can also stack zero or more cycled subnetworks onto the first subnetwork, as illustrated in Fig. 1c.

Unlike the forward-backward cycle consistency loss of the first cycled subnetwork for shadow removal, the cycled subnetworks for refinement exploit the cycle consistency loss in only one direction to encourage the reconstructed image \overline{I}_s^2 to be identical to the original input shadow image I_s , that is, $I_s \rightarrow G_{sf}^1 \overline{I}_{sf}^1 \rightarrow G_{sf}^2 \overline{I}_{sf}^2 \rightarrow G_s^2 \overline{I}_s^2$. As the objective of the refinement step is to improve shadow removal, the direction for shadow addition is not included. However, we plan to incorporate the forward-backward cycle consistency loss in future work and we will evaluate its performance.

$$\begin{aligned}
 L_{nsa}^2(G_{sf}^2, G_s^2) &= L_{nsa}^{sf,2}(G_{sf}^2) + L_{nsa}^{s,2}(G_s^2) \\
 L_{nsa}^{sf,2}(G_{sf}^2) &= \frac{1}{N} \sum \left(\frac{1}{|NSA|} \sum_{i \in NSA} \|p(I_s, i) - p(G_{sf}^2(\overline{I}_{sf}^1, \overline{M}_s^2) \otimes \overline{I}_{sf}^1, i)\|_1 \right) \\
 L_{nsa}^{s,2}(G_s^2) &= \frac{1}{N} \sum \left(\frac{1}{|NSA|} \sum_{j \in NSA} \|p(I_{sf}, j) - p(G_s^2(\overline{I}_{sf}^2, M_s) \otimes \overline{I}_{sf}^2, j)\|_1 \right),
 \end{aligned} \tag{11}$$

3.3 Network architecture

We consider the network architecture of the Mask-ShadowGAN [6] as the general architecture of the generators and discriminators for the cycled subnetworks for both shadow removal and refinement. The original architecture of Mask-ShadowGAN was drawn from CycleGAN [25], which is a seminal architecture for general image-to-image translation, instead of a specific design for shadow removal. Each generator consists of three convolutional layers, followed by nine residual blocks with stride-two convolutions and two deconvolutional layers for up-sampling and output generation. Furthermore, instance normalization [28] is used after each convolution and deconvolution operation. The generators concatenate the shadow or shadow-free image with the corresponding auxiliary information (e.g., shadow mask or difference map), which have four channels in total. For the discriminators, we used PatchGAN [29].

4 Experiments

4.1 Dataset and evaluation metrics

The ISTD dataset [11] is a standard benchmark used for shadow detection and removal experiments. It contains 1870 triplets of shadows, shadow masks, and shadow-free images under 135 different scenarios. Although the ISTD dataset exhibits various illumination properties, shapes, and scenes, it has an illumination inconsistency problem between the paired shadow and shadow-free images owing to slight changes in ambient light [11]. Le et al. [5] addressed this problem in previously published results using a color correction method to mitigate the color inconsistencies between the shadow and shadow-free image pairs. We also applied a color-corrected test dataset because the color-adjusted shadow-free images significantly affect the experimental results.

The Video Shadow Removal dataset [13] contains a set of eight videos, each containing a static scene without visible moving objects. For each video, the dataset provides a single pseudo shadow-free frame (i.e., pseudo ground truth) as well as a moving-shadow mask for each frame of the video. The moving-shadow mask marks the pixels appearing in both the shadow and non-shadow areas in the video. Following previous works [13, 17], we set a threshold of 80 to determine if a pixel is included in the moving-shadow mask.

Following the approach by Le et al. [13] for preparing the training dataset, patches of size 128×128 were cropped from a real shadow image of size 640×480 with a step size of 32. These were grouped into two sets according to the corresponding shadow masks: a non-shadow set containing patches without shadow pixels and a shadow set containing patches with both shadow and non-shadow pixels. In particular, we set the minimum percentage of shadow pixels of each patch in the shadow set to 10% to ensure differences between the shadow and non-shadow areas within the patch. In total, we created 110,201 non-shadow patches and 110,201 shadow patches from 1330 triplets. The remaining 540 triplets were used for testing. Following previous works [13, 16, 17], all shadow removal results with a resolution of 256×256 were used for the performance evaluations. However, our method can accept input images of any size.

We used the root-mean-squared error (RMSE) and mean absolute error (MAE) between the ground truth and generated shadow-free images in the LAB color space as evaluation metrics. Their formulas are as follows:

$$\begin{aligned} \text{MAE} &= \frac{1}{N} \sum_{j=1}^N |y_j - \hat{y}_j| \\ \text{RMSE} &= \sqrt{\frac{1}{N} \sum_{j=1}^N (y_j - \hat{y}_j)^2}, \end{aligned} \quad (14)$$

where N represents the number of test data items, and y_j and \hat{y}_j denote the predicted value and corresponding ground truth, respectively. MAE is a linear score that weighs all individual differences equally on average, whereas RMSE assigns a higher weight to large errors [30]. RMSE is most useful when large errors are particularly undesirable. We computed the RMSE and MAE for each test image and then averaged the scores over all test images, emphasizing the quality of each image for the shadow and non-shadow areas [16]. In general, the smaller these values are, the better is the performance.

4.2 Training details

Each patch image of size 128×128 in the training dataset was resized to 112×112 , and a random crop of 100×100 was used for training. All generators and discriminators were initialized using a zero-mean Gaussian distribution with a standard deviation of 0.02. The entire network is jointly trained in an end-to-end manner using the Adam optimizer, with the first and second momentum values set as 0.5 and 0.000, respectively. The learning rate was set to 2×10^{-4} for the first half of the epochs and then gradually decreased to zero with a linear decay rate in the next half of the epochs. The mini-batch size was set to one, and the training epoch was set to 60 for all experiments.

Our model was implemented using the PyTorch framework with a CUDA backend. We used a single NVIDIA GeForce GTX 2080ti graphics processing unit for training and testing. It took approximately 178 h to train C²ShadowGAN with a single cycled subnetwork for shadow removal and a single cycled subnetwork for refinement.

4.3 Results

We first conducted an experiment to assess the effectiveness of the shadow-matting approach over direct generation of shadow-free images for C²ShadowGAN. For this experiment, we modified only the generators of C²ShadowGAN to produce shadow and shadow-free images directly, as in Mask-ShadowGAN. The same training strategy and hyperparameter settings were used for both networks. In addition, we configured both networks to contain only the first cycled subnetwork

Table 1 Performance comparisons between two different methods for generating shadow-free images

| Method | Shadow area | | Non-shadow area | | All | |
|-------------------|-------------|------|-----------------|------|------|------|
| | MAE | RMSE | MAE | RMSE | MAE | RMSE |
| Direct generation | 16.32 | 5.28 | 4.67 | 2.88 | 6.50 | 3.47 |
| Shadow matting | 7.31 | 3.83 | 3.65 | 1.80 | 4.22 | 2.21 |

Table 2 Comparative performances according to the configuration of C²ShadowGAN

| # of cycles (<i>NC</i>) | Shadow area | | Non-shadow area | | All | |
|---------------------------|-------------|------|-----------------|------|------|------|
| | MAE | RMSE | MAE | RMSE | MAE | RMSE |
| <i>NC</i> =1 | 7.31 | 3.83 | 3.65 | 1.79 | 4.22 | 2.21 |
| <i>NC</i> =2 | 7.18 | 3.76 | 3.04 | 1.58 | 3.68 | 2.04 |
| <i>NC</i> =3 | 7.69 | 3.97 | 3.05 | 1.58 | 3.78 | 2.10 |
| <i>NC</i> =4 | 8.41 | 4.22 | 3.69 | 1.82 | 4.42 | 2.34 |

for shadow removal. Finally, we used DSDNet++ [14], pretrained on the SynShadow dataset and fine-tuned on the ISTD dataset, to obtain the shadow masks for our settings for all experiments. The quantitative results for the shadow areas, non-shadow areas, and all areas are shown in Table 1.

For shadow removal, the shadow-matting-based method outperformed that based on direct image generation by 35.1% for MAE and 36.3% for RMSE. The former method exhibited better performances in both shadow and shadow-free areas.

Comparative performance of C²ShadowGAN according to the configuration (i.e., number of cycles) is shown in Table 2. When the number of cycles in C²ShadowGAN is 1 (i.e., *NC* = 1), the network consists of a cycled subnetwork for shadow removal only. In contrast, when the number of cycles in C²ShadowGAN is $n > 1$ (i.e., *NC* = *n*), the network consists of the first cycled subnetwork for shadow removal and (*n*-1) additional linearly stacked cycled subnetworks for refinement.

C²ShadowGAN achieved the best performance with the configuration of a cycled subnetwork for shadow removal and a cycled subnetwork for refinement (i.e., *NC* = 2), reducing the RMSE in the shadow and non-shadow areas to 3.76 and 1.58, respectively. However, the overall performance and per-area performance decrease owing to overfitting as the number of cycled subnetworks increase. Figure 2 shows the

visual results generated from individual generators (e.g., G_{sf}^i) of the cycled subnetworks for both shadow removal and refinement. This demonstrates the negative impact of including more cycles than necessary. In the case of C²ShadowGAN with two cycled subnetworks, we observed that the final shadow-free images were refined from the intermediate shadow-free images, resulting in improved image quality (see Fig. 2c–d). However, in the case of C²ShadowGAN with four cycled subnetworks, we observed that the quality of the output image deteriorated with each additional cycle (see Fig. 2e–g). For instance, the colors of the shadow areas are not consistent with those of adjacent non-shadow areas, residuals on the shadow boundary remain sharp, and visually obvious artifacts are generated in the shadow and shadow-free areas.

We also conducted an ablation study to investigate the effectiveness of the proposed loss function. Starting from the original model with all the proposed losses, we trained the new models by removing the specific loss terms each time. For this experiment, we also used C²ShadowGAN with only the first cycled subnetwork for shadow removal because its relatively simple architecture makes it easier to analyze how individual loss functions affect network training.

Combinations of L_{GAN} , L_{cycle} , and $L_{identity}$ have been frequently adopted as objective functions by other state-of-the-art shadow removal models based on image-to-image translation [13, 16, 18]. From row 1 of Table 3, we observe that a significant level of shadow removal performance can be achieved using only these loss functions, and we use it as the baseline performance for comparison. Adding L_{BA} or L_{NSA} achieves overall performance gains of 5.7% and 7.8% for MAE and 3.75% and 7.5% for RMSE, respectively.

As shown in rows 2–3, the use of L_{BA} loss ($L_{baseline + BA}$) is more effective in transforming shadow pixels to non-shadow pixels. In contrast, the use of L_{NSA} loss ($L_{baseline + NSA}$) is useful for reducing unnecessary transformation of non-shadow areas. These results are consistent with the original purpose of the loss functions. As L_{BA} loss considers only

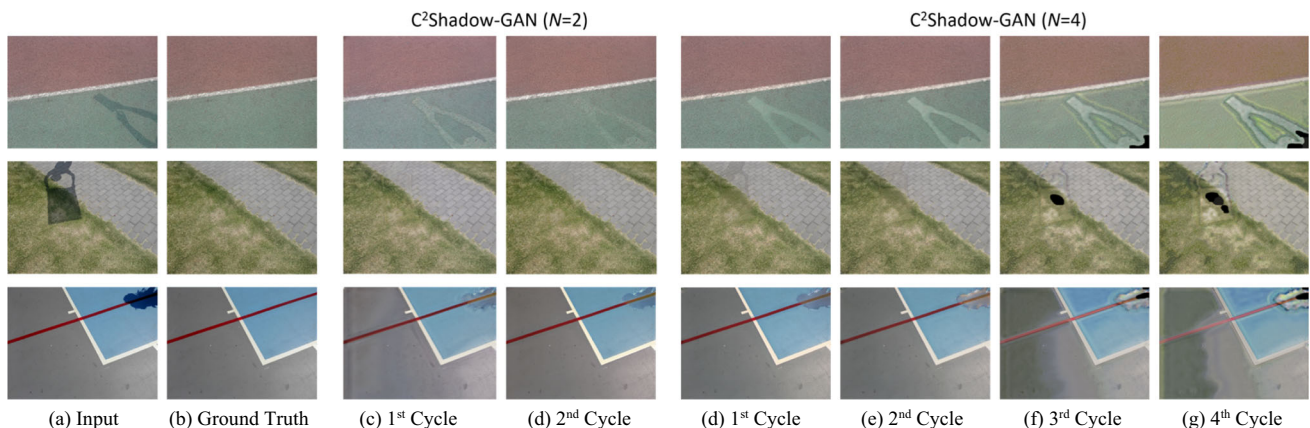
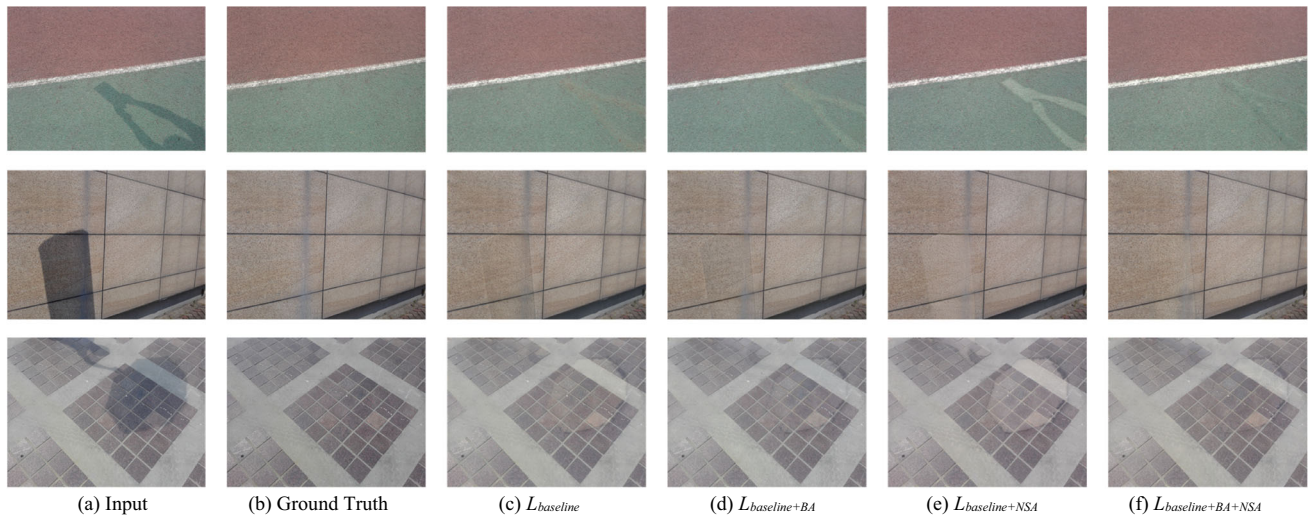
**Fig. 2** Visual results generated from generators (G_{sf}^i) of individual cycled subnetworks of C²ShadowGAN with (c,d) *NC* = 2 and (d–g) *NC* = 4

Table 3 Ablation study to investigate effectiveness of loss functions

| Loss | Constituent loss terms | | | | | Shadow area | | Non-shadow area | | All | |
|-----------------------|------------------------|-------------|----------------|----------|-----------|-------------|------|-----------------|------|------|------|
| | L_{GAN} | L_{cycle} | $L_{Identity}$ | L_{BA} | L_{NSA} | MAE | RMSE | MAE | RMSE | MAE | RMSE |
| $L_{baseline}$ | ✓ | ✓ | ✓ | – | – | 7.98 | 4.10 | 4.13 | 1.97 | 4.73 | 2.40 |
| $L_{baseline+BA}$ | ✓ | ✓ | ✓ | ✓ | – | 7.52 | 3.99 | 3.89 | 1.92 | 4.46 | 2.31 |
| $L_{baseline+NSA}$ | ✓ | ✓ | ✓ | – | ✓ | 8.13 | 4.06 | 3.65 | 1.76 | 4.36 | 2.22 |
| $L_{baseline+BA+NSA}$ | ✓ | ✓ | ✓ | ✓ | ✓ | 7.31 | 3.83 | 3.65 | 1.79 | 4.22 | 2.21 |

**Fig. 3** Visual results based on the loss functions

pixels within the shadow boundary area, the network can be trained to transform these pixels to look natural in the final shadow-free image by minimizing L_{BA} .

On the contrary, the network is optimized using L_{NSA} loss to generate outputs similar to the non-shadow areas of the input shadow image by handling the shadow and non-shadow areas separately. Although both MAE and RMSE for the non-shadow area increased slightly compared to the best case ($L_{baseline} + NSA$), the best overall performance was achieved when both L_{BA} and L_{NSA} were used together

Table 4 Comparative performances according to B_{step} for the shadow boundary area

| B_{step} | Shadow area | | Non-shadow area | | All | |
|------------|-------------|------|-----------------|------|------|------|
| | MAE | RMSE | MAE | RMSE | MAE | RMSE |
| 2 | 7.31 | 3.83 | 3.65 | 1.79 | 4.22 | 2.21 |
| 10 | 10.99 | 5.47 | 4.97 | 2.42 | 5.91 | 3.05 |
| 20 | 9.65 | 4.83 | 4.81 | 2.21 | 5.57 | 2.71 |
| 50 | 8.64 | 4.25 | 4.07 | 1.86 | 4.78 | 2.38 |
| 256 | 8.04 | 4.16 | 3.76 | 1.83 | 4.43 | 2.31 |

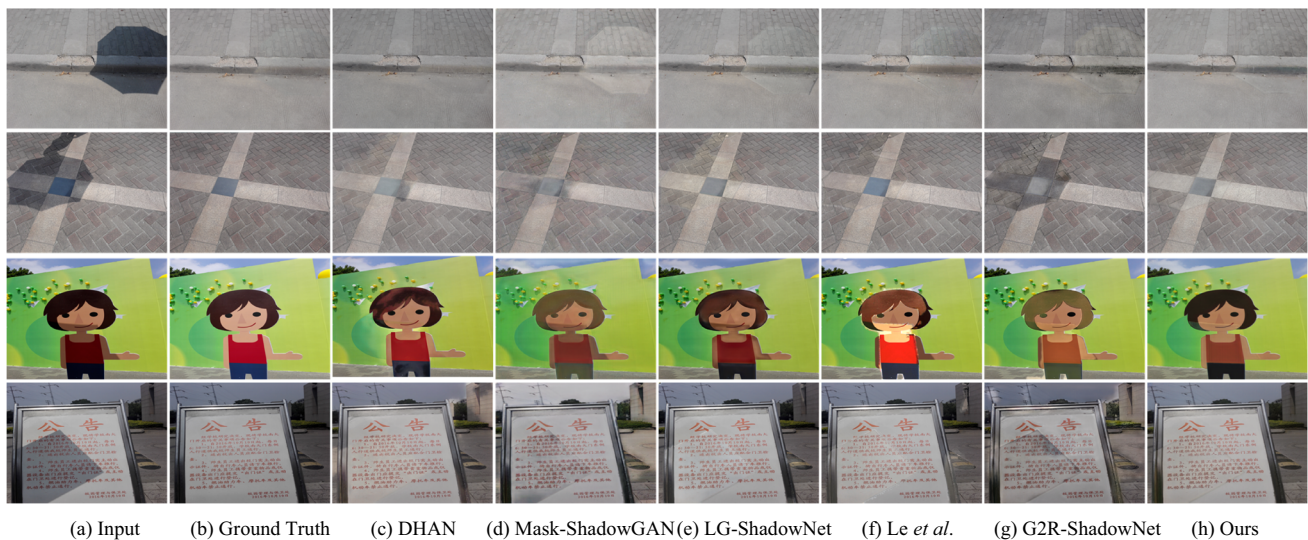
($L_{baseline} + BA + NSA$). Figure 3 shows the qualitative results of C²ShadowGAN trained with different loss functions. It also demonstrates that compared with the model trained using all of the proposed loss functions, the other models that are trained with a subset of the loss terms may cause obvious artifacts in the generated shadow-free images. Overall, L_{BA} and L_{NSA} losses are crucial for learning appropriate shadow removal as they restrict the model to process individual pixels based on their characteristics.

In Table 4, we summarize the impact of choosing different B_{step} pixels, which determines the shadow boundary area for L_{BA} loss. The best performance was achieved when $B_{step} = 2$. However, when B_{step} was set to 10 or 20, the corresponding performances were worse than those of models trained without L_{BA} loss (i.e., $L_{baseline}$ and $L_{baseline} + NSA$). This suggests that careful selection of B_{step} is required for the performance gains.

We also compared the proposed method with recent state-of-the-art supervised and unsupervised methods on the ISTD dataset: DHAN [4], Mask-ShadowGAN [6], LG-ShadowNet [18], Le et al. [13], and G2R-ShadowNet [16]. DHAN is one of the representative supervised shadow removal systems, in which shadow and shadow-free image pairs are required to

Table 5 Comparative performance of the proposed method with state-of-the-art methods on the ISTD dataset

| Methods | Training Data | Shadow area | | Non-shadow area | | All | |
|--------------------|--|-------------|------|-----------------|------|------|------|
| | | MAE | RMSE | MAE | RMSE | MAE | RMSE |
| DHAN [4] | Paired shadow and shadow-free images | 6.95 | 3.52 | 2.72 | 1.56 | 3.29 | 1.87 |
| Mask-ShadowGAN [6] | Unpaired shadow and shadow-free images | 10.65 | 5.29 | 5.58 | 2.89 | 6.37 | 3.47 |
| LG-ShadowNet [18] | | 9.81 | 4.89 | 3.27 | 1.94 | 4.29 | 2.69 |
| Le et al. [13] | Shadow images + Shadow masks | 9.96 | 4.86 | 2.87 | 1.83 | 3.93 | 2.57 |
| G2R-ShadowNet [17] | | 7.36 | 3.74 | 2.95 | 1.85 | 3.64 | 2.28 |
| Ours ($NC=2$) | | 7.18 | 3.76 | 3.04 | 1.58 | 3.68 | 2.04 |

**Fig. 4** Visual results comparison of shadow removal on the ISTD dataset

learn to remove shadows. Mask-ShadowGAN and LG-ShadowNet require unpaired shadow and shadow-free images to train their models, while the other weakly supervised methods, including ours, require shadow images and shadow masks for network training. The results are either produced by us using the officially available codes or are provided by the authors of the original publications. For this experiment, we used C^2 ShadowGAN with a cycled subnetwork for shadow

removal and a cycled subnetwork for refinement, which showed the best performance (Table 2).

As shown in Table 5, DHAN showed the best performance as it was trained on paired data with pixel-level annotations [18]. Among the unsupervised methods, our method achieves a competitive performance advantage over other methods. Compared with methods that use the same type of training data as ours, our method outperformed the G2R-ShadowNet by 10.5% and the approach of Le et al. by 20.6% in terms of RMSE for the complete image. Although the MAE value of the C^2 ShadowGAN was slightly higher than that of the G2R-ShadowNet, we observed that the output images generated by the C^2 ShadowGAN contained fewer artifacts. Figure 4 shows the qualitative results of our method and other state-of-the-art methods, which present some challenging cases such as large shadow areas (first row) and shadows across backgrounds with complex textures and colors (second, third, and fourth rows). In most cases, except for the third row, where all methods produced results with visible artifacts, our method generated more realistic shadow-free images and restore the

Table 6 Comparison of generalization capability of the proposed method and state-of-the-art methods

| Method | MAE | RMSE |
|--------------------|-------|-------|
| Mask-ShadowGAN [6] | 33.99 | 31.53 |
| LG-ShadowNet [18] | 22.71 | 21.15 |
| G2R-ShadowNet [16] | 20.76 | 20.09 |
| Ours ($NC=2$) | 19.64 | 18.02 |



Fig. 5 Visual results comparison of shadow removal on the Video Shadow Removal dataset

texture details occluded by shadows. This is because our model exploits effective constraints for learning appropriate shadow removal based on pixel characteristics and learns to avoid unrealistic output images through adversarial learning.

Finally, we compared the generalization capability of our model with Mask-ShadowGAN [6], LG-ShadowNet [18], and G2R-ShadowNet [16]. All models were trained on the ISTD dataset and tested on the Video Shadow Removal dataset without additional training or fine-tuning. We applied the pretrained shadow detection model used in Le et al. [13] to obtain a set of shadow masks for each video. By following previous works [13, 16], we measured the MAE and RMSE in the LAB color space between the output frame and pseudo ground truth on the moving-shadow area marked by the moving-shadow mask.

The quantitative results are listed in Table 6. Our method exhibited the best performance for all the metrics. In particular, our method outperformed the G2R-ShadowNet—which showed comparable performance to ours in terms of shadow removal—reducing the MAE and RMSE by 5.39% and 10.30%, respectively. This result demonstrates that our method has better generalizability for other unseen environments.

Figure 5 shows the visual comparison results for four samples that present close (first and third rows) or distant shots (second and fourth rows). The first and second rows show examples where shadow removal was performed relatively well. Although our method showed consistent performance in both cases, the performances of the other methods fluctuated. For instance, G2R-ShadowNet recovered the shadow area of the forest with fewer artifacts, as in our study (first row), but

failed to generate a shadow-free image with an occlusion object of a complex shape (second row). In rows 3–4, although all methods failed to remove shadows completely, we observe that our method preserves the texture details of the shadow area better than other methods. However, we expect that accurate shadow mask generation and additional fine-tuning processes will suppress these artifacts considerably.

5 Conclusion

In this study, we developed a novel weakly supervised GAN with a cycle-in-cycle structure for shadow removal using unpaired data, called the C²ShadowGAN. Our method leverages the cycle consistency constraint based on the cycle-in-cycle structure, in which multiple cycled subnetworks are stacked to learn to remove shadows. We also introduced loss functions for learning shadow removal based on pixel characteristics. Thus, the network was able to restrict simple modifications of all parts of the image to fool the discriminator. We conducted extensive experiments to assess the effectiveness of our method and showed that our method achieved competitive performance against recent state-of-the-art shadow removal methods with training on unpaired data. In the future, we plan to extend this method to accommodate higher resolution real-life images, such as high-resolution drone images. We also plan to exploit state-of-the-art technologies, such as self-supervision, to enhance deep learning methods trained on unpaired data.

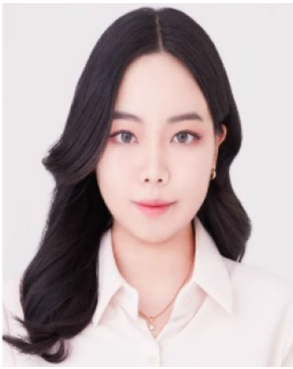
Acknowledgments This research was supported by a grant (22DRMS-B147287-05) for the development of a customized realistic 3D geospatial information update and utilization technology based on consumer demand, funded by the Ministry of Land, Infrastructure and Transport of the Korean government

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Zhang Y, Chen G, Vukomanovic J, Singh KK, Liu Y, Holden S, Meetemeyer RK (2020) Recurrent shadow attention model (RSAM) for shadow removal in high-resolution urban land-cover mapping. *Remote Sens Environ* 247:111945–111959
- Qu L, Tian J, He S, Tang Y, Lau R (2017) DeshadowNet: a multi-context embedding deep network for shadow removal. In: *Proc. CVPR, Honolulu, Hawaii, USA*, pp 2308–2316
- Wei J, Long C, Zou H, Xiao C (2019) Shadow inpainting and removal using generative adversarial networks with slice convolutions. *Comput Graph Forum* 38(7):381–392
- Cun X, Pun C, Shi C (2020) Towards ghost-free shadow removal via dual hierarchical aggregation network and shadow matting GAN. In: *Proc. AAAI Conf. Artif. Intell., New York, NY, USA*, pp 10680–10687
- Le H, Samaras D (2019) Shadow removal via shadow image decomposition. In: *Proc. ICCV, Seoul, South Korea*, pp 8578–8587
- Hu X, Jiang Y, Fu C, Heng P (2019) Mask-ShadowGAN: learning to remove shadows from unpaired data. In: *Proc. ICCV, Seoul, South Korea*, pp 2472–2481
- Vasluianu F, Romero A, Gool L, Timofte R (2021) Self-supervised shadow removal. In: *Proc. CVPRW*, pp 826–835
- Hu X, Fu C, Zhu L, Qin J, Heng P (2020) Direction-aware spatial context features for shadow detection and removal. *IEEE Trans Pattern Anal Mach Intell* 42(11):2795–2808
- Zheng Q, Qian X, Cao Y, Lau R (2019) Distraction-aware shadow detection. In: *Proc. CVPR, Long Beach, LA, USA*, pp 5167–5176
- Khan SH, Bennamoun M, Sohel F, Togneri R (2016) Automatic shadow detection and removal from a single image. *IEEE Trans Pattern Anal Mach Intell* 38(3):431–446
- Wang J, Li X, Yang J (2018) Stacked conditional generative adversarial networks for jointly learning shadow detection and shadow removal. In: *Proc. CVPR, Salt Lake City, UT, USA*, pp 1788–1797
- Vincente TF, Hou L, Yu C, Hoai M, Samaras D (2016) Large-scale training of shadow detectors with noisily-annotated shadow examples. In: *Proc. ECCV, Amsterdam, Netherlands*, pp 816–832
- Le H, Samaras D (2020) From shadow segmentation to shadow removal. In: *Proc. ECCV*, pp 264–281
- Inoue N, Yamasaki T (2020) Learning from synthetic shadows for shadow detection and removal. *IEEE Trans Circuits Syst Video Technol* 31(11):4187–4197
- Kim J, Jang I (2021) Dual hierarchical aggregation network based enhanced shadow detection and removal. *J Korean Soc Geospatial Info Sci* 29(2):27–34
- Liu Z, Yin H, Wu X, Wu Z, Mi Y, Wang S (2021) From shadow generation to shadow removal. In: *Proc. CVPR*, pp 4927–4936
- Goodfellow I, Pouget-Abadie J, Mirza M, Xu B, Warde-Farley D, Ozair S, Courville A, Bengio Y (2014) Generative adversarial nets. *Adv Neural Inf Process Syst*:2672–2680
- Liu Z, Yin H, Mi Y, Pu M, Wang S (2021) Shadow removal by a lightness-guided network with training on unpaired data. *IEEE Trans Image Process* 30:1853–1865
- Li Y, Tang S, Zhang R, Zhang Y, Li U, Yan S (2019) Asymmetric GAN for unpaired image-to-image translation. *IEEE Trans Image Process* 28:5881–5896
- Choi Y, Choi M, Kim M, Ha J, Kim S, Choo J (2018) StarGAN: unified generative adversarial networks for multi-domain image-to-image translation. In: *Proc. CVPR, Salt Lake City, UT, USA*, pp 8789–8797
- Kumar PC (2019) A survey on various shadow detection and removal methods. In: *Proc. ICCVBIC, Coimbatore, India*, pp 395–401
- Tiwari A, Singh PK, Amin S (2016) A survey on shadow detection and removal in images and video sequences. In: *Proc. CONFLUENCE, Noida, India*, pp 518–523
- Murali S, Govindan V, Kalady S (2016) A survey on shadow removal techniques for single image. *Int J Image Graph Signal Process* 8:38–46
- Shilpa M, Gopalakrishnan MT, Naveena C (2020) Approach for shadow detection and removal using machine learning techniques. *IET Image Process* 14(13):2998–3005
- Zhu J, Park T, Isola P, Efros AA (2017) Unpaired image-to-image translation using cycle-consistent adversarial networks. In: *Proc. ICCV, Venice, Italy*, pp 1033–1038
- Arbel E, Helor H (2011) Shadow removal using intensity surfaces and texture anchor points. *IEEE Trans Pattern Anal Mach Intell* 33:1202–1216
- Fan H, Han M, Li J (2019) Image shadow removal using end-to-end deep convolutional neural networks. *Appl Sci* 9:1–18
- Ulyanov D, Vedaldai A, Lempitsky V (2016) Instance normalization: the missing ingredient for fast stylization. *arXiv preprint arXiv:1701.02096*
- Isola P, Zhu J, Zhou T, Efros A (2017) Image-to-image translation with conditional adversarial networks. In: *Proc. CVPR, Honolulu, Hawaii, USA*, pp 1125–1134
- Son S, Song Y, Kim N, Do Y, Kwak N, Lee M, Lee B (2019) TW3-based fully automated bone age assessment system using deep neural networks. *IEEE Access* 7:33346–33358

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Sunwon Kang Song received her BS from the Department of Computer Science, Kyonggi University, South Korea, in 2020, where she is currently pursuing her MS at the Department of Computer Science. Her current research interests include computer vision, deep learning, and geospatial artificial intelligence.



In Sung Jang received his BS and MS in Computer Engineering from Pusan National University, South Korea, in 1999 and 2001, respectively. Since 2001, he has been a principal member of the research staff at ETRI. His main research areas are platforms for digital twins, 3D GIS service, GeoAIoT service, location based services, and moving object DBMS.



Juwan Kim received his BS and MS in Computer Engineering from Pusan National University, Pusan, South Korea, in 1993 and 1995, respectively, as well as his PhD in Computer Science from Chungnam National University, Daejeon, in 2004. Since 1995, he has been a research member of Electronic and Telecommunications Research Institute (ETRI). His research interests include geospatial information science, computer vision, and geospatial artificial intelligence.



Byoung-Dai Lee received his BS and MS in Computer Science from Yonsei University, South Korea, in 1996 and 1998, respectively. He received his PhD in Computer Science and Engineering from the University of Minnesota, Minneapolis, USA, in 2003. He is currently a full professor at the Division of AI and Computer Engineering, Kyonggi University, Korea. Before joining Kyonggi University, he worked at Samsung Electronics, Co., Ltd. as a senior engineer from 2003 to 2010. His research interests include computer vision, medical imaging analysis, and geospatial artificial intelligence.