When Shadow Removal Meets Intrinsic Image Decomposition: A Joint Learning Framework Using Unpaired Data

Rongjia Zheng[1], Qing Zhang[1,3*], Yongwei Nie[2], Wei-Shi Zheng[1,3]
[1]Sun Yat-sen University, China
[2]South China University of Technology, China
[3]The Key Laboratory of Machine Intelligence and Advanced Computing, Ministry of Education, China
zhengrj23@mail2.sysu.edu.cn, zhangq93@mail.sysu.edu.cn, nieyongwei@scut.edu.cn, wszheng@ieee.org

Abstract

We present a framework that achieves shadow removal by learning intrinsic image decomposition (IID) from unpaired shadow and shadow-free images. Although it is well-known that intrinsic images, *i.e.*, illumination and reflectance, are highly beneficial to shadow removal, IID is rarely adopted by previous work due to its inherent ambiguity and the scarcity of training data. However, we find that by properly coupling shadow removal and IID into a joint learning framework, they can reinforce each other and enable promising results on both tasks, even with unpaired training data. Our framework is comprised of an IID network for separating the shadow input image into illumination and reflectance, and an illumination recovery network for predicting shadow-free illumination with which we are able to produce the shadow removal output by recombining with the estimated reflectance. We perform extensive experiments on various benchmark datasets to demonstrate the effectiveness of our method in shadow removal, and also showcase our advantage over previous IID methods in handling images with complex shadows.

Introduction

Shadows are everywhere in photos. Although they are important clue to perceive scene lighting and geometry, their presence may degrade the overall visual quality of a photo (*e.g.*, unwanted shadows in landscape and document images), and may also challenge many fundamental computer vision tasks, such as segmentation, object detection and tracking. Hence, shadow removal has long been a fundamental problem in computer vision.

Shadow removal is a challenging task as it requires to accurately locate shadows and recover illumination for the shadow regions while restoring the texture details behind the shadows, which depends on reliable understanding of the scene lighting, geometry, and material information. There are numerous shadow removal methods in the literature. Early work addresses the problem by formulating it as layer separation based on a specific image formation model (Finlayson, Hordley, and Drew 2002; Liu and Gleicher 2008), or by transferring illumination from non-shadow regions to shadow regions (Zhang, Zhang, and Xiao 2015; Guo, Dai, and Hoiem 2011). Recent effort and progress on shadow removal is mostly learning-based, due to the emergence of deep learning and large-scale datasets (Wang, Li, and Yang 2018; Le and Samaras 2019; Qu et al. 2017).

Most learning-based shadow removal methods are trained in a fully-supervised manner (Zhang et al. 2020; Niu et al. 2022; Liu et al. 2023b; Zhang, Gu, and Zhu 2022; Zhu et al. 2022c; Sen et al. 2023; Liu et al. 2024), requiring paired shadow and shadow-free images. Although these methods demonstrate promising results, the inherent difficulty of collecting massive real-world paired images covering diverse scenes limits their generalization capability. A line of research to eliminate the dependency on paired data is to learn image-to-image translation based on unpaired data (Hu et al. 2019b; Liu et al. 2021a; Jin, Sharma, and Tan 2021; Le and Samaras 2020; Luo et al. 2023). However, methods in this category have the following two limitations: (i) they usually do not generalize well to unseen data beyond the training data, (ii) they may produce unnatural results with disturbing shadow residual or appearance inconsistency artifacts, as shown in Figure 3, 4. Besides, there are also a few attempts towards zero-shot shadow removal trained on a single image (Gandelsman, Shocher, and Irani 2019; Jiang et al. 2023), but the high inference time limits their practicability.

A recent research trend is to utilize intrinsic images for shadow removal (Liu et al. 2024; Guo et al. 2023c). Based on *paired* training data, Liu *et al.* (Liu et al. 2024) proposed to learn intrinsic image decomposition (IID) for removing shadows by recasting the lighting of the shadow regions, while Guo *et al.* (Guo et al. 2023c) firstly employed a pre-trained inpainting diffusion model for shadow removal, and then trained an IID model based on synthesized shadow images for refining context details within the shadow regions. Although these two works demonstrate impressive results, they either require paired data for training, or modelling shadow removal and IID separately. Moreover, both of them rely on high-quality masks to locate shadows during training and testing. In this paper, we present a framework for joint learning of shadow removal and intrinsic image decomposition from unpaired shadow and shadow-free images, without using shadow masks. Instead of directly learning image-to-image translation like previous unpaired learning based methods, we design our framework to learn IID for obtaining an intermediate representation that decouples the origi-

*Corresponding author.

nal shadow image into illumination and reflectance. Based on the estimated illumination, we develop an illumination recovery network to adaptively locate shadows and generate shadow-free illumination that compensates for the occluded lighting in shadow regions. The shadow removal result is then produced by recombining the shadow-free illumination and the original reflectance. To enable more effective IID, we devise a novel loss function that enforces reflectance consistency via shadow removal and shadow generation based on cycle-based adversarial training framework. Below, we summarize the major contributions of this work:

- We present a framework that jointly learns shadow removal and IID from unpaired shadow and shadow-free images without shadow masks.
- We present an IID network for decomposing an image into reflectance and illumination, and develop an illumination recovery network for locating and removing shadows in the illumination space.
- We show that our method outperforms existing unsupervised and unpaired learning based shadow removal methods, and also enables very promising IID performance on previously challenging outdoor images with shadows.

Related Work

Shadow Removal. Early shadow removal methods are primarily designed based on the observed physical properties of shadow (Drew, Finlayson, and Hordley 2003), and certain assumptions on low-level features such as edge (Shor and Lischinski 2008), intensity (Gryka, Terry, and Brostow 2015), and texture (Zhang, Zhang, and Xiao 2015). However, these methods fail to handle complex shadows, since real-world shadows cannot always be effectively modeled by simplified physical properties and assumptions. In recent years, deep-learning-based methods have become a preferable choice for shadow removal. Given the access to paired datasets such as ISTD (Wang, Li, and Yang 2018) and SRD (Qu et al. 2017), a large number of methods are developed by learning image-to-image mapping in a fully supervised manner (Guo et al. 2023b; Zhu et al. 2022b; Wan et al. 2022; Liu et al. 2023a,b, 2024; Jin et al. 2024). Different from supervised methods, some works propose to train shadow removal networks based on unpaired shadow and shadow-free images that are easy to collect (Hu et al. 2019b; Liu et al. 2021a; Le and Samaras 2020; Jin, Sharma, and Tan 2021; Luo et al. 2023). More recently, Guo *et al.* (Guo et al. 2023c) proposed to utilize an IID model trained on synthesized shadow images to help refine the context information of shadow removal result generated by a pre-trained inpainting diffusion-based model.

Intrinsic Image Decomposition. Due to the ill-posed nature of this problem, traditional IID methods focus primarily on formulating optimization that integrates hand-crafted priors or assumptions on reflectance and illumination (Bi, Han, and Yu 2015; Li and Brown 2014). Recent IID methods are basically built upon deep neural networks. Based on the datasets with ground-truth intrinsic images (*e.g.*, MIT (Grosse et al. 2009), MPI-Sintel (Butler et al. 2012), and ShapeNet (Chang et al. 2015)), various supervised methods were developed (Narihira, Maire, and Yu 2015a; Kim et al. 2016; Lettry, Vanhoey, and Van Gool 2018). However, as MPI-Sintel and ShapeNet are synthetic datasets, and the MIT dataset contains only a very small number of images, current supervised methods do not generalize well to real-world scenes. Along with the introduction of two real-world indoor datasets with human annotations, *i.e.*, IIW (Bell, Bala, and Snavely 2014) and SAW (Kovacs et al. 2017), IID of real-world scenes have been significantly fueled in later works (Zhou, Krahenbuhl, and Efros 2015; Baslamisli, Le, and Gevers 2018; Fan et al. 2018; Narihira, Maire, and Yu 2015b; Nestmeyer and Gehler 2017; Das, Karaoglu, and Gevers 2022). However, these methods fail to handle outdoor scenes. Although there are several attempts towards unsupervised IID (Li and Snavely 2018; Zhang et al. 2021; Liu et al. 2020; Janner et al. 2017; Ma et al. 2018; Jin et al. 2023), they do not work well for outdoor scenes with complex shadows.

Method

Figure 1 presents the workflow of our framework. It is comprised of two main sub-networks, *i.e.* intrinsic image decomposition network (IID-Net) and illumination recovery network (IR-Net). Given an input shadow image, we first feed it to IID-Net to obtain its intrinsic images, *i.e.*, reflectance and illumination. Next, the IR-Net is employed to locate shadows based on the illumination and recover a shadow-free illumination that can be recombined with the reflectance to produce the final shadow removal result. The whole network is trained with unpaired shadow and shadow-free images, by jointly learning IID and shadow removal.

Shadow Removal via Intrinsic Decomposition

Given a shadow image I, intrinsic decomposition aims to factorize it into the pixel-wise product of a reflectance (also referred to as albedo) image R and an illumination (also referred to as shading) image S, *i.e.*,

$$I = R \odot S. \tag{1}$$

With the estimated reflectance R and illumination S, shadow removal can be parameterized and formulated as

$$I_f = R \odot S_f, \text{ with } S_f = \mathcal{F}_\theta(S), \tag{2}$$

where $\mathcal{F}_\theta$ denotes a deep network with learned parameters θ for illumination recovery, from which we get the shadow-free illumination S_f. I_f is the desired shadow removal result produced by combining the original reflectance R and the predicted shadow-free illumination S_f.

Our motivation. Tackling shadow removal from the perspective of IID involves many challenges. First of all, IID itself is a very challenging ill-posed problem, especially when faced with real-world outdoor images with shadows. In addition, shadow removal has low tolerance to bad decomposition that misclassifies shadows as reflectance. Finally, how to learn high-quality IID from unpaired shadow and shadow-free images remains unexplored. Despite the above challenges, we show that by coupling shadow removal and IID in a joint learning framework, the two tasks can effectively

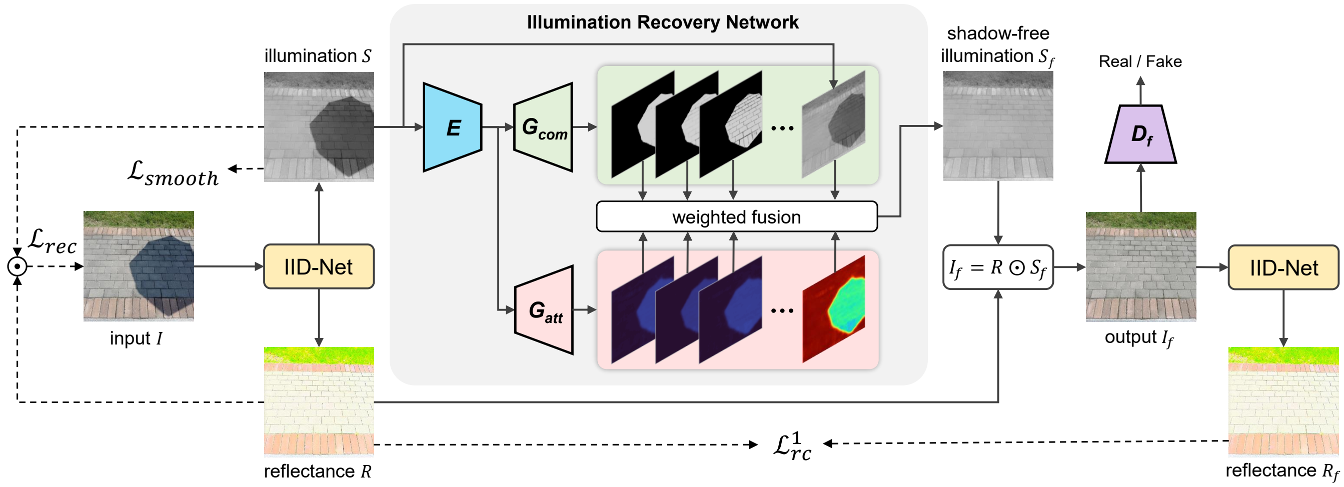


Figure 1: **Overview of our method.** Given an input shadow image I, it is first fed into the intrinsic image decomposition network (IID-Net) to obtain reflectance R and illumination S. Next, the illumination recovery network (IR-Net) takes the illumination S as input and produces a shadow-free illumination S_f, by performing weighted fusion between a series of attention maps and the corresponding illumination compensation images. Based on S_f, a shadow removal output I_f is recovered by $I_f = R \odot S_f$. The entire network is trained with unpaired data, based on a loss function consisting of various reflectance/illumination related constraints and adversarial losses.

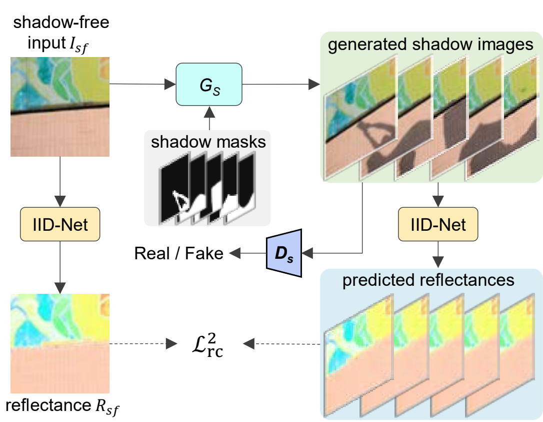


Figure 2: Learning reflectance consistency from shadow generation.

reinforce each other, giving our approach various advantages over prior works: (i) Thanks to the simplicity and similarity nature of illumination in natural images, our method has stronger generalization capability in shadow removal, as demonstrated in Figure 9. (ii) As shadows are restored in a material-independent illumination space, the common appearance inconsistency artifacts encountered by previous methods are well-suppressed in our method (see Figures 3 and 4). (iii) As a byproduct, our method enables surprisingly good IID performance on previously challenging outdoor images with shadows (see Figure 5).

Intrinsic Image Decomposition Network

To guarantee the success of our IID-Net, we first define the following reconstruction loss to constrain that the estimated reflectance R and illumination S can reconstruct the original image I:

$$\mathcal{L}_{rec} = \| I - R \odot S \|_1 . \tag{3}$$

Besides, we enforce reflectance consistency between the original shadow image I and the generated shadow-free image I_f to ensure that the learned reflectance is independent of illumination, which is expressed as:

$$\mathcal{L}^1_{rc} = \| R - R_f \|_1, \tag{4}$$

where R_f is reflectance estimated from the output shadow-free image I_f. To enable more effective reflectance learning, unlike previous unpaired learning methods that use the shadow generator G_s solely for cycle consistency constraint, we further utilize G_s to produce a sequence of shadow images based on a shadow-free input image I_{sf} and a set of shadow masks (see Figure 2). We then impose reflectance consistency between the reflectance of the generated shadow images and the original shadow-free input image by:

$$\mathcal{L}^2_{rc} = \sum_{i=1}^{K} \| R_{sf} - R'_i \|_1, \tag{5}$$

where R_{sf} is reflectance of the shadow-free input I_{sf}, while $\{R'_i\}_{i=1}^K$ denotes the set of reflectance estimated from the generated shadow images. Note, K is empirically set as 5, and the shadow masks used for generation are produced by (Otsu 1979) based on the input shadow image I and predicted shadow-free image I_f during training. The reflectance consistency loss for IID-Net is now defined as:

$$\mathcal{L}_{rc} = \mathcal{L}^1_{rc} + \mathcal{L}^2_{rc}, \tag{6}$$

Considering that illumination is generally locally smooth in natural images (Li and Brown 2014), we propose to encourage spatially smoothness of illumination by:

$$\mathcal{L}_{smooth} = \|\nabla_x S\|_1 + \|\nabla_y S\|_1, \tag{7}$$

Figure 3: Visual comparison of shadow removal on ISTD+ (first two rows) and SRD dataset (last two rows).

Figure 4: Visual comparison of shadow removal on USR (first two rows) and SBU datasets (last two rows).

where ∇_x and ∇_y refer to first-order derivatives along horizontal and vertical directions, respectively. Note that, to accelerate training, we initialize illumination S to be the Y channel of the input shadow image I in YUV color space.

Illumination Recovery Network

To remove shadows while avoiding unwanted modifications to the original non-shadow regions, we design IR-Net to be comprised of three components, an encoder E for feature extraction, a generator G_{att} to predict a group of normalized attention maps $\{\alpha_i\}_{i=1}^{N}$, and another generator G_{com} for producing the same number of illumination compensation images $\{C_i\}_{i=1}^{N}$ where C_N is fixed to the original illumination S. In this way, we are able to obtain the desired shadow-free illumination S_f by

$$S_f = \sum_{i=1}^{N} \alpha_i C_i, \tag{8}$$

where N is the number of attention maps, which is empirically set as 10. The reason behind setting $C_N = S$ is to encourage the attention map α_N to locate non-shadow regions while enforcing the remaining attention maps to focus on shadows. Multiple attention maps utilized for shadows is to account for possible inhomogeneous shadows involving complex illumination variations.

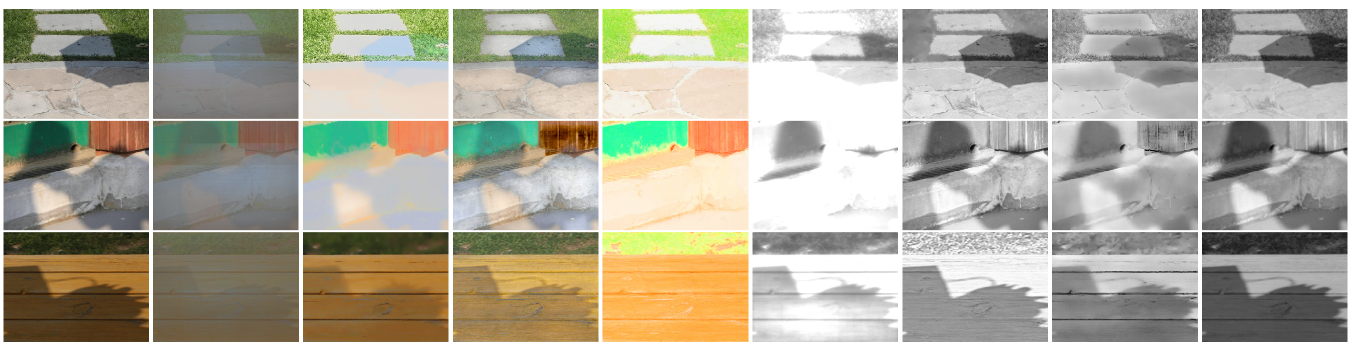
It is worth mentioning that a similar fusion strategy was explored in (Fu et al. 2021), which achieves shadow removal by weighted fusion of the original shadow image and its multiple over-exposure images. However, their design is not suitable for our task because it requires firstly training an exposure estimation model based on paired data, and then incorporating the model to train another weighted fusion framework in a fully supervised manner. As shown in Figure 8, naively transferring the design in (Fu et al. 2021) to our framework yields unsatisfactory results.

Loss Function

In addition to the losses related to IID-Net, akin to (Hu et al. 2019b), we employ various other losses to enable more effective learning from unpaired data. Specifically, a discrim-

Method	Training data	Params	Flops	ISTD+			SRD		
				Shad.↓	Non-Shad.↓	All↓	Shad.↓	Non-Shad.↓	All↓
Input Image	-	-	-	40.20	2.60	8.50	36.69	4.83	14.05
DeshadowNet	Paired	-	-	15.90	6.00	7.60	11.78	4.84	6.64
ST-CGAN	Paired+Mask	29M	18G	13.40	7.70	8.70	18.64	6.37	8.23
DHAN	Paired+Mask	22M	263G	11.20	7.10	7.70	8.73	5.81	6.62
Auto-Exposure	Paired+Mask	143M	160G	6.50	3.80	4.20	8.56	5.75	6.51
Inpaint4shadow	Paired+Mask	15M	81G	5.93	2.90	3.39	6.09	2.97	3.83
ShadowFormer	Paired+Mask	11M	64G	5.40	2.40	2.80	6.05	3.55	4.09
ShadowDiff	Paired+Mask	64M	9100G	4.90	2.30	2.70	4.98	3.44	3.63
RRL-Net	Paired+Mask	172M	6100G	5.69	2.31	2.87	5.49	3.00	3.66
HomoFormer	Paired+Mask	18M	1140G	5.42	2.26	2.76	4.25	2.85	3.33
LFG-Diffusion	Paired+Mask	82.6M	2100G	5.15	2.47	2.90	-	-	-
DeS3	Paired	82.9M	11875G	-	-	-	5.88	2.83	3.72
MaskGAN	Unpaired	11M	57G	12.40	4.00	5.30	9.89	5.74	6.78
DC-Net	Unpaired	11M	105G	11.02	4.96	5.92	11.59	4.91	6.43
LG-Net	Unpaired	6M	52G	9.72	3.58	4.52	12.69	5.88	7.61
G2R-Net	Shadow image+Mask	23M	113G	12.37	3.14	4.60	13.20	5.50	8.10
BADC	Shadow image+Mask	40M	28750G	7.60	2.40	3.30	-	-	-
Ours	Unpaired	13M	100G	7.06	3.65	4.13	7.82	4.85	5.61
Ours w/ test-time mask	Unpaired	13M	100G	7.06	2.60	3.21	7.82	4.83	5.60

Table 1: Quantitative comparison of shadow removal on ISTD+ and SRD datasets.



Input LIID-R UID-R PIE-R Ours-R LIID-S UID-S PIE-S Ours-S

Figure 5: Comparison of IID on outdoor images with shadows. R and S refer to reflectance and illumination, respectively.

inator D_f is used to determine whether the output image I_f belongs to the shadow-free domain. To achieve shadow generation, a shadow generator G_s and a shadow domain discriminator D_s are also utilized. By denoting our shadow removal method as G_f, we define the following adversarial loss

$$\mathcal{L}_{GAN} = \mathcal{L}_{adv}(G_f, D_f) + \mathcal{L}_{adv}(G_s, D_s), \tag{9}$$

where $\mathcal{L}_{adv}(G_f, D_f)$ and $\mathcal{L}_{adv}(G_s, D_s)$ are defined as:

$$\begin{aligned}\mathcal{L}_{adv}(G_f, D_f) = \mathbb{E}_{I_{sf}}[log\big(D_f(I_{sf})\big)] + \\ \mathbb{E}_I[log\big(1 - D_f(G_f(I))\big)],\end{aligned} \tag{10}$$

$$\begin{aligned}\mathcal{L}_{adv}(G_s, D_s) = \mathbb{E}_I[log\big(D_s(I)\big)] + \\ \mathbb{E}_{I_{sf}}[log\big(1 - D_s(G_s(I_{sf}, \mathcal{M}_r))\big)],\end{aligned} \tag{11}$$

where I and I_{sf} are real shadow and shadow-free images. Note, we generate binary mask $\mathcal{M}$ using (Otsu 1979) to guide shadow generation. $\mathcal{M}_r$ denotes a randomly selected shadow mask learned from previous iterations. Besides, we adopt the cycle consistency loss (Zhu et al. 2017):

$$\begin{aligned}\mathcal{L}_{cycle} = \parallel I - G_s(G_f(I), \mathcal{M}) \parallel_1 + \\ \parallel I_{sf} - G_f(G_s(I_{sf}, \mathcal{M}_r)) \parallel_1 .\end{aligned} \tag{12}$$

Finally, akin to (Hu et al. 2019b), we define the identity loss:

$$\mathcal{L}_{id} = \parallel I - G_s(I, \mathcal{M}_0) \parallel_1 + \parallel I_{sf} - G_f(I_{sf}) \parallel_1, \tag{13}$$

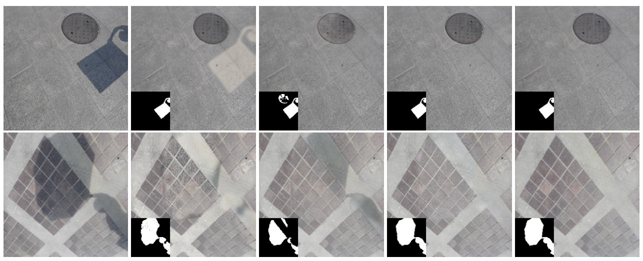
where $\mathcal{M}_0$ is a full-zero mask which indicates no shadow to be added. In summary, the overall loss function for our framework is formulated as:

$$\begin{aligned}\mathcal{L}_{total} = \mathcal{L}_{rec} + \lambda_1 \mathcal{L}_{rc} + \lambda_2 \mathcal{L}_{smooth} + \lambda_3 \mathcal{L}_{GAN} + \\ \lambda_4 \mathcal{L}_{cycle} + \lambda_5 \mathcal{L}_{id},\end{aligned} \tag{14}$$

where we empirically set $\lambda_1 = 0.25$, $\lambda_2 = 1.25$, $\lambda_3 = 0.05$, $\lambda_4 = 0.5$, and $\lambda_5 = 0.25$.

Method	Training setting (dataset)	WHDR↓
Retinex-Color	-	26.9%
DI	supervised (Sintel+MIT)	37.3%
Shi *et al.*	supervised (ShapeNet)	59.4%
SIID	unsupervised (IIW)	28.1%
UID	unsupervised (IIW)	18.2%
PIE	supervised (NED)	34.0%
Ours	unsupervised (SRD)	26.3%

Table 2: Quantitative evaluation of IID on the IIW dataset.



Input w/o IID-Net w/o IR-Net Ours GT

Figure 6: Effect of IID-Net and IR-Net on shadow removal. We show that IID-Net plays an important role in avoiding appearance inconsistency artifacts and IR-Net helps accurately identify shadows (see masks in bottom-left corner).

Implementation and Training Details

The ResNet-based generator in (Hu et al. 2019b) is used for the IID-Net and the shadow generator G_s, which consists of three convolution layers, followed by five residual blocks and two transposed convolution layers for feature map upsampling. As for IR-Net, the encoder contains one input convolution layer, two downsampling convolution layers, and another five residual blocks, while the other two generators (G_{com} and G_{att}) consist of two transposed convolution layers and one output convolution layer. The PatchGAN in (Isola et al. 2017) is used as our discriminators (D_s and D_f).

We build our model on Pytorch and train it for 200 epochs with a mini-batch size of 1. The entire network is optimized using Adam optimizer with the first and second momentum values set as 0.5 and 0.999. The initial learning rate is set as 2×10^{-4} for the first 100 epochs and gradually reduced to zero with a linear decay rate in the other 100 epochs.

Experiments

Datasets. We evaluate the shadow removal performance of our method on six public datasets including ISTD+ (Le and Samaras 2019; Wang, Li, and Yang 2018), SRD (Qu et al. 2017), USR (Hu et al. 2019b), SBU (Vicente et al. 2016), WSRD (Vasluianu, Seizinger, and Timofte 2023), VSR (Le and Samaras 2020). Following previous works (Jin, Sharma, and Tan 2021), to evaluate on paired datasets, *i.e.*, ISTD+ or SRD, are randomly shuffled to provide unpaired data. As SRD does not provide shadow masks, we follow (Fu et al. 2021) to use masks provided by (Cun, Pun, and Shi 2020) for evaluation. For the USR dataset with unpaired data, we only conduct visual comparisons due to the lack of ground-truth shadow removal results. We do not train our method on the SBU dataset because it does not contain shadow-free images, and instead provide qualitative results produced by our model trained on the ISTD+ dataset.

Method	Shadow↓	Non-shadow↓	All↓
w/o IID-Net	10.51	3.96	4.99
w/o IR-Net	7.13	4.41	4.67
w/o $\mathcal{L}_{rc}^1$	7.18	3.82	4.37
w/o $\mathcal{L}_{rc}^2$	7.21	3.80	4.35
Ours	7.06	3.65	4.13

Table 3: Ablation study of different components in our method on the ISTD+ dataset.

Metric	ST-CGAN	BDRAR	DSC	SDCM	Ours
IOU↑	0.71	0.78	0.64	0.89	0.84
BER↓	4.24	3.26	3.42	1.44	3.12

Table 4: Quantitative comparison of shadow detection on the ISTD+ dataset.

Evaluation Metrics. We follow previous works to evaluate the shadow removal results at the resolution of 256×256, and use mean absolute error (MAE) [1] defined in LAB color space for quantitative evaluation.

Comparison on Shadow Removal

Baselines. We compare our method with various state-of-the-art shadow removal methods, which can be roughly divided into two categories: (i) *supervised methods*, including DeshadowNet (Qu et al. 2017), ST-CGAN (Wang, Li, and Yang 2018), DHAN (Cun, Pun, and Shi 2020), Auto-Exposure (Fu et al. 2021), Inpaint4shadow (Li et al. 2023), ShadowFormer (Guo et al. 2023a), ShadowDiff (Guo et al. 2023b), RRL-Net (Liu et al. 2024), HomoFormer (Xiao et al. 2024), LFG-Diffusion (Mei et al. 2024), and DeS3 (Jin et al. 2024). (ii) *Unsupervised and unpaired learning based methods*, including MaskGAN (Hu et al. 2019b), DC-Net (Jin, Sharma, and Tan 2021), LG-Net (Liu et al. 2021a), G2R-Net (Liu et al. 2021b), and BADC (Guo et al. 2023c). Note, as RRL-Net and BADC do not release their codes, their quantitative results for shadow removal are from their paper, and we are unable to compare with them on IID.

Evaluation on ISTD+ and SRD. Table 1 reports the quantitative comparison results on the ISTD+ and SRD datasets. As shown, our method outperforms all methods trained on unpaired data and produces results comparable to leading supervised methods. Compared to the latest unsupervised method BADC (Guo et al. 2023c) that also utilize IID for shadow removal, our method achieves better results on shadow region even with far fewer parameters. Note, the

[1]Note, most previous shadow removal methods claim to report RMSE but actually compute MAE in their implementations.

Figure 7: Effect of reflectance consistency loss $\mathcal{L}_{rc}$.

Figure 8: Effect of two alternative illumination recovery designs, *i.e.*, learning illumination residual and adopting the exposure fusion strategy (Exp. fusion) in (Fu et al. 2021).

advantage of BADC on non-shadow regions mainly comes from the use of shadow masks. As shown in the last row in Table 1, if masks are similarly adopted during testing, our method outperforms BADC on both shadow and non-shadow regions. Figure 3 provides visual comparison results, where our method produces high-quality results free of visual artifacts.

Evaluation on USR and SBU. Due to the lack of ground-truth shadow-free images, we can only perform qualitative comparisons on USR and SBU. Figure 4 provides the visual comparison results on the two datasets, where our results exhibit clear advantages.

Comparison on Intrinsic Image Decomposition

Besides shadow removal, we also compare our method with several IID methods, including Retinex-Color (Grosse et al. 2009), DI (Narihira, Maire, and Yu 2015a), Shi *et al.* (Shi et al. 2017), SIID (Ma et al. 2018), LIID (Li and Snavely 2018), UID (Zhang et al. 2021) and PIE (Das, Karaoglu, and Gevers 2022) trained on NED dataset (Baslamisli et al. 2018). As shown in Figure 5, our model trained on the SRD dataset effectively separates reflectance and illumination without mistaking shadows as reflectance, significantly outperforming the compared specialized IID methods. We also quantitatively evaluate our IID performance on the IIW dataset in Table 2, where our method outperforms most of the IID methods (except for UID which is trained on IIW), manifesting its effectiveness in IID.

More Analysis

Ablation studies. Besides the visual ablation study results in Figures 6, 7, 8, we also conduct ablation studies on the ISTD+ dataset to quantitatively evaluate the effectiveness of different designs in our method. As shown in Table 3, omitting the IID-Net and learning illumination recovery in the original RGB color space leads to a large performance drop from 7.06 to 10.51 for the shadow regions, validating the necessity of IID. On the other hand, the performance on non-shadow regions degrades clearly from 3.65 to 4.41 on ISTD+ when the IR-Net is replaced with the generator in (Hu et al. 2019b). The numerical result also demonstrates that the reflectance consistency loss $\mathcal{L}_{rc}^1$ and $\mathcal{L}_{rc}^2$ are beneficial to shadow removal.

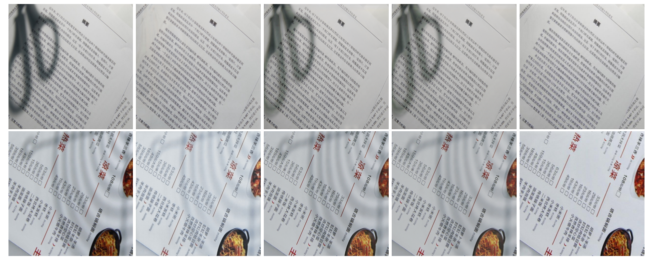


Figure 9: Generalization to unseen document images.

Analysis of generalization ability. Thanks to the simplicity nature of illumination, by adopting IID, our method achieves strong generalization capability. As shown in Figure 9, although our model is trained on the ISTD+ dataset, it produces visually compelling shadow removal results for document images, which are even better than results produced by a specialized document image shadow removal method BEDSR (Lin, Chen, and Chuang 2020).

Evaluation on shadow detection. To examine the effectiveness of our method in shadow detection, we in Table 4 compare our predicted shadow attention map with shadow masks generated by supervised shadow detection methods including BDRAR (Zhu et al. 2018), DSC (Hu et al. 2019a), and SDCM (Zhu et al. 2022a). As shown, our method produces results comparable to the compared methods in terms of intersection over union (IOU) and balance error rate (BER).

Conclusion

We have presented a framework for joint learning of shadow removal and intrinsic image decomposition (IID) from unpaired data. Our core idea is to learn IID and transfer shadow removal to the decomposed illumination space, so as to embrace the simplicity nature of illumination to boost the robustness and generalization of shadow removal. To this end, a novel unpaired learning framework is devised to train shadow removal and IID jointly in a reciprocal manner. Extensive experiments show that our method outperforms existing unsupervised and unpaired learning based shadow removal methods, and enables superior IID performance on previously challenging outdoor images with shadows.

Acknowledgments

This work is supported by the National Natural Science Foundation of China (62471499), Guangdong Basic and Applied Basic Research Foundation (2023A1515030002).

References

Baslamisli, A. S.; Groenestege, T. T.; Das, P.; Le, H.-A.; Karaoglu, S.; and Gevers, T. 2018. Joint learning of intrinsic images and semantic segmentation. In *ECCV*, 286–302.

Baslamisli, A. S.; Le, H.-A.; and Gevers, T. 2018. CNN based learning using reflection and retinex models for intrinsic image decomposition. In *CVPR*, 6674–6683.

Bell, S.; Bala, K.; and Snavely, N. 2014. Intrinsic images in the wild. *ACM Transactions on Graphics*, 33(4): 1–12.

Bi, S.; Han, X.; and Yu, Y. 2015. An l1 image transform for edge-preserving smoothing and scene-level intrinsic decomposition. *ACM Transactions on Graphics*, 34(4): 1–12.

Butler, D. J.; Wulff, J.; Stanley, G. B.; and Black, M. J. 2012. A naturalistic open source movie for optical flow evaluation. In *ECCV*, 611–625.

Chang, A. X.; Funkhouser, T.; Guibas, L.; Hanrahan, P.; Huang, Q.; Li, Z.; Savarese, S.; Savva, M.; Song, S.; Su, H.; et al. 2015. Shapenet: An information-rich 3d model repository. *arXiv preprint arXiv:1512.03012*.

Cun, X.; Pun, C.-M.; and Shi, C. 2020. Towards ghost-free shadow removal via dual hierarchical aggregation network and shadow matting gan. In *AAAI*, volume 34, 10680–10687.

Das, P.; Karaoglu, S.; and Gevers, T. 2022. Pie-net: Photometric invariant edge guided network for intrinsic image decomposition. In *CVPR*, 19790–19799.

Drew, M. S.; Finlayson, G. D.; and Hordley, S. D. 2003. Recovery of chromaticity image free from shadows via illumination invariance. In *ICCVW*, 32–39.

Fan, Q.; Yang, J.; Hua, G.; Chen, B.; and Wipf, D. 2018. Revisiting deep intrinsic image decompositions. In *CVPR*, 8944–8952.

Finlayson, G. D.; Hordley, S. D.; and Drew, M. S. 2002. Removing Shadows from Images. In *ECCV*.

Fu, L.; Zhou, C.; Guo, Q.; Juefei-Xu, F.; Yu, H.; Feng, W.; Liu, Y.; and Wang, S. 2021. Auto-exposure fusion for single-image shadow removal. In *CVPR*, 10571–10580.

Gandelsman, Y.; Shocher, A.; and Irani, M. 2019. Double-DIP: unsupervised image decomposition via coupled deep-image-priors. In *CVPR*, 11026–11035.

Grosse, R.; Johnson, M. K.; Adelson, E. H.; and Freeman, W. T. 2009. Ground truth dataset and baseline evaluations for intrinsic image algorithms. In *ICCV*, 2335–2342.

Gryka, M.; Terry, M.; and Brostow, G. J. 2015. Learning to remove soft shadows. *ACM Transactions on Graphics*, 34(5): 1–15.

Guo, L.; Huang, S.; Liu, D.; Cheng, H.; and Wen, B. 2023a. ShadowFormer: Global Context Helps Shadow Removal. In *AAAI*. ISBN 978-1-57735-880-0.

Guo, L.; Wang, C.; Yang, W.; Huang, S.; Wang, Y.; Pfister, H.; and Wen, B. 2023b. Shadowdiffusion: When degradation prior meets diffusion model for shadow removal. In *CVPR*, 14049–14058.

Guo, L.; Wang, C.; Yang, W.; Wang, Y.; and Wen, B. 2023c. Boundary-Aware Divide and Conquer: A Diffusion-based Solution for Unsupervised Shadow Removal. In *ICCV*, 13045–13054.

Guo, R.; Dai, Q.; and Hoiem, D. 2011. Single-image shadow detection and removal using paired regions. In *CVPR*, 2033–2040.

Hu, X.; Fu, C.-W.; Zhu, L.; Qin, J.; and Heng, P.-A. 2019a. Direction-aware spatial context features for shadow detection and removal. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 42(11): 2795–2808.

Hu, X.; Jiang, Y.; Fu, C.-W.; and Heng, P.-A. 2019b. Mask-shadowgan: Learning to remove shadows from unpaired data. In *ICCV*, 2472–2481.

Isola, P.; Zhu, J.-Y.; Zhou, T.; and Efros, A. A. 2017. Image-to-image translation with conditional adversarial networks. In *CVPR*, 1125–1134.

Janner, M.; Wu, J.; Kulkarni, T. D.; Yildirim, I.; and Tenenbaum, J. 2017. Self-supervised intrinsic image decomposition. In *NIPS*, volume 30.

Jiang, H.; Zhang, Q.; Nie, Y.; Zhu, L.; and Zheng, W.-S. 2023. Learning to Remove Shadows from a Single Image. *International Journal of Computer Vision*, 1–18.

Jin, Y.; Li, R.; Yang, W.; and Tan, R. T. 2023. Estimating reflectance layer from a single image: Integrating reflectance guidance and shadow/specular aware learning. In *AAAI*, volume 37, 1069–1077.

Jin, Y.; Sharma, A.; and Tan, R. T. 2021. DC-ShadowNet: Single-image hard and soft shadow removal using unsupervised domain-classifier guided network. In *ICCV*, 5027–5036.

Jin, Y.; Ye, W.; Yang, W.; Yuan, Y.; and Tan, R. T. 2024. DeS3: Adaptive Attention-Driven Self and Soft Shadow Removal Using ViT Similarity. In *AAAI*, volume 38, 2634–2642.

Kim, S.; Park, K.; Sohn, K.; and Lin, S. 2016. Unified depth prediction and intrinsic image decomposition from a single image via joint convolutional neural fields. In *ECCV*, 143–159.

Kovacs, B.; Bell, S.; Snavely, N.; and Bala, K. 2017. Shading annotations in the wild. In *CVPR*, 6998–7007.

Le, H.; and Samaras, D. 2019. Shadow removal via shadow image decomposition. In *ICCV*, 8578–8587.

Le, H.; and Samaras, D. 2020. From shadow segmentation to shadow removal. In *ECCV*, 264–281.

Lettry, L.; Vanhoey, K.; and Van Gool, L. 2018. DARN: a deep adversarial residual network for intrinsic image decomposition. In *WACV*, 1359–1367.

Li, X.; Guo, Q.; Abdelfattah, R.; Lin, D.; Feng, W.; Tsang, I.; and Wang, S. 2023. Leveraging inpainting for single-image shadow removal. In *ICCV*, 13055–13064.

Li, Y.; and Brown, M. S. 2014. Single image layer separation using relative smoothness. In *CVPR*, 2752–2759.

Li, Z.; and Snavely, N. 2018. Learning intrinsic image decomposition from watching the world. In *CVPR*, 9039–9048.

Lin, Y.-H.; Chen, W.-C.; and Chuang, Y.-Y. 2020. BEDSR-Net: A Deep Shadow Removal Network From a Single Document Image. In *CVPR*, 12902–12911.

Liu, F.; and Gleicher, M. 2008. Texture-consistent shadow removal. In *ECCV*, 437–450.

Liu, J.; Wang, Q.; Fan, H.; Li, W.; Qu, L.; and Tang, Y. 2023a. A Decoupled Multi-Task Network for Shadow Removal. *IEEE Transactions on Multimedia*.

Liu, J.; Wang, Q.; Fan, H.; Tian, J.; and Tang, Y. 2023b. A Shadow Imaging Bilinear Model and Three-Branch Residual Network for Shadow Removal. *IEEE Transactions on Neural Networks and Learning Systems*.

Liu, Y.; Ke, Z.; Xu, K.; Liu, F.; Wang, Z.; and Lau, R. W. 2024. Recasting Regional Lighting for Shadow Removal. In *AAAI*.

Liu, Y.; Li, Y.; You, S.; and Lu, F. 2020. Unsupervised learning for intrinsic image decomposition from a single image. In *CVPR*, 3248–3257.

Liu, Z.; Yin, H.; Mi, Y.; Pu, M.; and Wang, S. 2021a. Shadow Removal by a Lightness-Guided Network With Training on Unpaired Data. *IEEE Transactions on Image Processing*, 30: 1853–1865.

Liu, Z.; Yin, H.; Wu, X.; Wu, Z.; Mi, Y.; and Wang, S. 2021b. From shadow generation to shadow removal. In *CVPR*, 4927–4936.

Luo, W.; Xie, X.; Deng, K.; Yang, L.; and Lai, J. 2023. Learning Shadow Removal from Unpaired Samples via Reciprocal Learning. *IEEE Transactions on Image Processing*.

Ma, W.-C.; Chu, H.; Zhou, B.; Urtasun, R.; and Torralba, A. 2018. Single image intrinsic decomposition without a single intrinsic image. In *ECCV*, 201–217.

Mei, K.; Figueroa, L.; Lin, Z.; Ding, Z.; Cohen, S.; and Patel, V. M. 2024. Latent feature-guided diffusion models for shadow removal. In *WACV*, 4313–4322.

Narihira, T.; Maire, M.; and Yu, S. X. 2015a. Direct intrinsics: Learning albedo-shading decomposition by convolutional regression. In *ICCV*, 2992–2992.

Narihira, T.; Maire, M.; and Yu, S. X. 2015b. Learning lightness from human judgement on relative reflectance. In *CVPR*, 2965–2973.

Nestmeyer, T.; and Gehler, P. V. 2017. Reflectance adaptive filtering improves intrinsic image estimation. In *CVPR*, 6789–6798.

Niu, K.; Liu, Y.; Wu, E.; and Xing, G. 2022. A Boundary-Aware Network for Shadow Removal. *IEEE Transactions on Multimedia*.

Otsu, N. 1979. A Threshold Selection Method from Gray-Level Histograms. *IEEE Transactions on Systems, Man, and Cybernetics*, 9(1): 62–66.

Qu, L.; Tian, J.; He, S.; Tang, Y.; and Lau, R. W. 2017. Deshadownet: A multi-context embedding deep network for shadow removal. In *CVPR*, 4067–4075.

Sen, M.; Chermala, S. P.; Nagori, N. N.; Peddigari, V.; Mathur, P.; Prasad, B.; and Jeong, M. 2023. Shards: Efficient shadow removal using dual stage network for high-resolution images. In *WACV*, 1809–1817.

Shi, J.; Dong, Y.; Su, H.; and Yu, S. X. 2017. Learning non-lambertian object intrinsics across shapenet categories. In *CVPR*, 1685–1694.

Shor, Y.; and Lischinski, D. 2008. The shadow meets the mask: Pyramid-based shadow removal. In *Computer Graphics Forum*, volume 27, 577–586.

Vasluianu, F.-A.; Seizinger, T.; and Timofte, R. 2023. Wsrd: A novel benchmark for high resolution image shadow removal. In *CVPR*, 1825–1834.

Vicente, T. F. Y.; Hou, L.; Yu, C.-P.; Hoai, M.; and Samaras, D. 2016. Large-scale training of shadow detectors with noisily-annotated shadow examples. In *ECCV*, 816–832.

Wan, J.; Yin, H.; Wu, Z.; Wu, X.; Liu, Y.; and Wang, S. 2022. Style-guided shadow removal. In *ECCV*, 361–378.

Wang, J.; Li, X.; and Yang, J. 2018. Stacked conditional generative adversarial networks for jointly learning shadow detection and shadow removal. In *CVPR*, 1788–1797.

Xiao, J.; Fu, X.; Zhu, Y.; Li, D.; Huang, J.; Zhu, K.; and Zha, Z.-J. 2024. HomoFormer: Homogenized Transformer for Image Shadow Removal. In *CVPR*, 25617–25626.

Zhang, L.; Long, C.; Zhang, X.; and Xiao, C. 2020. Risgan: Explore residual and illumination with generative adversarial networks for shadow removal. In *AAAI*, volume 34, 12829–12836.

Zhang, L.; Zhang, Q.; and Xiao, C. 2015. Shadow remover: Image shadow removal based on illumination recovering optimization. *IEEE Transactions on Image Processing*, 24(11): 4623–4636.

Zhang, Q.; Zhou, J.; Zhu, L.; Sun, W.; Xiao, C.; and Zheng, W.-S. 2021. Unsupervised intrinsic image decomposition using internal self-similarity cues. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 44(12): 9669–9686.

Zhang, X. F.; Gu, C. C.; and Zhu, S. Y. 2022. SpA-Former: Transformer image shadow detection and removal via spatial attention. *arXiv preprint arXiv:2206.10910*.

Zhou, T.; Krahenbuhl, P.; and Efros, A. A. 2015. Learning data-driven reflectance priors for intrinsic image decomposition. In *ICCV*, 3469–3477.

Zhu, J.-Y.; Park, T.; Isola, P.; and Efros, A. A. 2017. Unpaired image-to-image translation using cycle-consistent adversarial networks. In *ICCV*, 2223–2232.

Zhu, L.; Deng, Z.; Hu, X.; Fu, C.-W.; Xu, X.; Qin, J.; and Heng, P.-A. 2018. Bidirectional feature pyramid network with recurrent attention residual modules for shadow detection. In *ECCV*, 121–136.

Zhu, Y.; Fu, X.; Cao, C.; Wang, X.; Sun, Q.; and Zha, Z.-J. 2022a. Single image shadow detection via complementary mechanism. In *Proceedings of the 30th ACM International Conference on Multimedia*, 6717–6726.

Zhu, Y.; Huang, J.; Fu, X.; Zhao, F.; Sun, Q.; and Zha, Z.-J. 2022b. Bijective mapping network for shadow removal. In *CVPR*, 5627–5636.

Zhu, Y.; Xiao, Z.; Fang, Y.; Fu, X.; Xiong, Z.; and Zha, Z.-J. 2022c. Efficient model-driven network for shadow removal. In *AAAI*, volume 36, 3635–3643.