MetaShadow: Object-Centered Shadow Detection, Removal, and Synthesis

Tianyu Wang[1,2,*], Jianming Zhang[1], Haitian Zheng[1], Zhihong Ding[1], Scott Cohen[1],
Zhe Lin[1], Wei Xiong[1], Chi-Wing Fu[2], Luis Figueroa[1,†], Soo Ye Kim[1,†]
[1] Adobe Research
[2] The Chinese University of Hong Kong

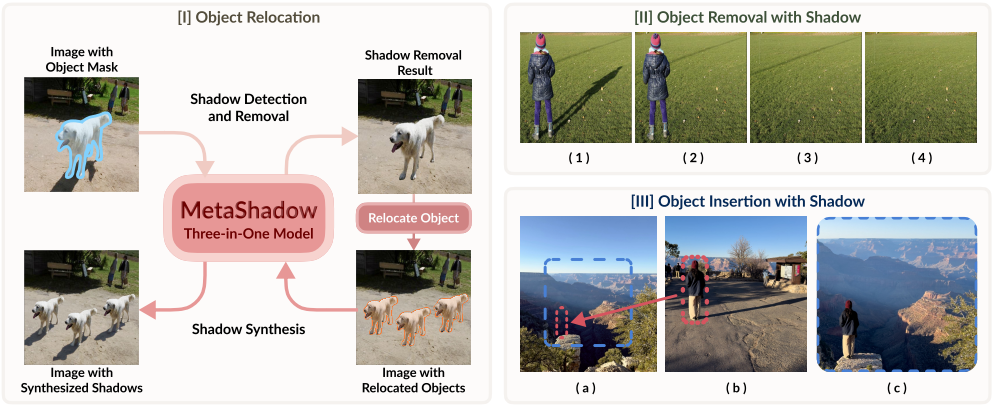


Figure 1. **MetaShadow** is a versatile three-in-one framework designed for shadow-related tasks, enabling shadow manipulation in various object-centered image editing operations such as: [I] Object Relocation: Our model can detect and remove the shadow of an existing object, then synthesize the shadow in the new location consistent with the original shadow. [II] Remove an object and its shadow: (1) Based on the mask of the unwanted object, our model can directly remove its shadow (2). After removing the object (3), we can eliminate any remaining shadows for a cleaner background (4) if we do not specify which shadow to remove. [III] Insert an object and synthesize its shadow: When inserting the person in (b) to another image (a) with similar lighting, our model can generate a realistic shadow, enhancing the final compositing quality.

Abstract

Shadows are often under-considered or even ignored in image editing applications, limiting the realism of the edited results. In this paper, we introduce MetaShadow, a three-in-one versatile framework that enables detection, removal, and controllable synthesis of shadows in natural images in an object-centered fashion. MetaShadow combines the strengths of two cooperative components: Shadow Analyzer, for object-centered shadow detection and removal, and Shadow Synthesizer, for reference-based controllable shadow synthesis. Notably, we optimize the learning of the intermediate features from Shadow Analyzer to guide Shadow Synthesizer to generate more realistic shadows that blend seamlessly with the scene. Extensive evaluations on multiple shadow benchmark datasets show significant improvements of MetaShadow over the existing state-of-the-art methods on object-centered shadow detection, removal, and synthesis. MetaShadow excels in image-editing tasks such as object removal, relocation, and insertion, pushing the boundaries of object-centered image editing.

1. Introduction

Shadows play a vital role in revealing the realism of an image, providing strong cues on the perception of the 3D space and the spatial relations between objects in the environment. However, when handling object-related image-editing tasks, such as unwanted object removal, object relocation, and object insertion, existing applications (*e.g.*, Google Magic Eraser [9]) in this field often simply neglect manipulating the shadows, greatly diminishing the overall visual coherence and realism of the edited images.

To effectively support image editing with shadows, as shown in Fig. 1, we need to collectively deal with three tasks: shadow detection, shadow removal, and shadow synthesis in an *object-centered fashion*. Our *object-centered* approach indicates that we primarily focus on instance-level object manipulation with applications to image editing workflows. Thus, when we edit objects in an image, each object should be associated with the shadow cast by itself onto the environment such that its associated shadow can be naturally manipulated together.

As shown in Tab. 1, most existing works, however, treat the three tasks separately, with several methods overlooking the need for object-centric formulations to assist image-editing workflows. General shadow detection [56, 65, 66]

* Work done during an internship at Adobe.
† Co-corresponding authors.

predicts a single binary mask for all shadows, while Wang *et al*. [48, 51] detects object-shadow pairs, akin to instance segmentation. A line of shadow removal works [22, 46, 68] require binary shadow masks as input, relying on off-the-shelf shadow detectors or user-given masks, both of which can be error-prone, to remove the shadows. Further, more advanced shadow removal methods [6, 28, 47] simultaneously detect and remove *all* image shadows, but do not support object-centered editing. In contrast, shadow synthesis aims to generate shadows for objects inserted into a new scene. One line of work [42, 43, 58] require additional estimated lighting or geometric parameters to produce convincing results. Another line of research [13, 29] utilizes another object or its shadow as a reference for synthesizing shadows. However, the absence of an effective shadow knowledge extractor prevents these methods from producing accurate shadow shapes. Recently, ObjectDrop [52] introduced a bootstrap supervision strategy to generate shadow synthesis data by training an object/shadow removal model.

Intuitively, all shadow tasks are inherently related, and should benefit from shared knowledge. For example, perfectly removing a shadow implies that one can derive an accurate shadow mask (shadow detection) from the image. Furthermore, it indicates that crucial properties like softness and intensity of the shadow have been learned. Meanwhile, knowing how to predict a binary shadow mask of an object would indicate implicit knowledge of where to synthesize the object shadow if it were not present. By design, handling each task separately limits each specialized model from benefitting from the shared knowledge in the shadow formation cycle and hinders them from achieving higher-quality results. Although we may apply multiple existing methods sequentially in the image editing pipeline, e.g., object relocation, this may lead to inconsistent outcomes. This is because existing methods for different shadow tasks do not share common shadow knowledge, leading to suboptimal visual quality due to discrepancies in shadow shape, color, and intensity.

In this work, we propose a three-in-one framework named **MetaShadow**, consisting of two synergistic components that enable object-centered shadow detection, removal, and synthesis simultaneously. Motivated by the limitation of specialized shadow-related models, we propose a novel training mechanism that successfully shares the shadow information across its task-specific components to achieve superior results. To the best of our knowledge, MetaShadow is the first framework that can jointly handle all three shadow tasks in an object-centered fashion, benefitting from the shared knowledge to achieve SOTA results.

We evaluate MetaShadow on three real-world tasks and four benchmarks. Our MetaShadow improves the mIoU from 55.8 to 71.0 for shadow mask detection, improves the bbox PSNR by 8.7dB for shadow removal, and reduces the local RMSE from 51.73 to 36.54 for shadow synthesis.

Method	Task			Condition	
	Detection	Removal	Synthesis	Object-Centered	Reference-Based
SILT [56]	✓	✗	✗	✗	✗
SSISv2 [51]	✓	✗	✗	✓	✗
DHAN [6]	✓	✓	✗	✗	✗
BMNet [67]	✗	✓	✓	✗	✗
ShadowDiffusion [11]	✗	✓	✗	✗	✗
Zhang*et al*. [58]	✗	✓	✗	✓	✗
PixHt-Lab [43]	✗	✗	✓	✓	✗
SGRNet [13]	✗	✗	✓	✓	✓
SGDiffusion [29]	✗	✗	✓	✓	✓
ObjectDrop [52]	✗	✓	✓	✓	✓
MetaShadow (Ours)	✓	✓	✓	✓	✓

Table 1. SOTA shadow-related methods and their supported task(s). Existing works handle up to two shadow-related tasks at once, with select models supporting an object-centered approach. Only one model uses other object-shadow pairs as a reference for shadow synthesis, avoiding the need for additional parameters that other models require. MetaShadow is a three-in-one framework that handles object-centered shadow detection, removal, and synthesis.

To summarize, the main contributions of this work are:

- **Three-in-one Framework**: MetaShadow adopts a novel object-centered GAN with reference-based diffusion to address the challenges of shadow understanding and manipulation to achieve object-centered image editing.
- **Shadow Knowledge Transfer**: Our approach is the first to utilize shadow-rich intermediate features from a GAN to guide the diffusion, significantly enhancing the visual quality and controllability of shadow synthesis.
- **Task-Specific Datasets for Shadow Editing**: We build a synthetic training set (MOS dataset) for shadow detection, removal, and synthesis, along with two real-world test sets, Moving DESOBA and Video DESOBA, for thorough qualitative and quantitative evaluation in target scenarios.
- **SOTA Performance on Three Tasks**: Extensive experiments on the benchmarks show that our method outperforms the baselines for object-centered shadow detection, removal, and synthesis.

2. Related works

2.1. Shadow Detection

General Shadow Detection. Existing works in this category aim to simultaneously detect *all shadow pixels in an image,* producing a single *general* shadow mask. The advent of deep learning revolutionized the task with CNNs [4, 8, 14, 17, 18, 21, 23, 33, 40, 45, 47, 53, 62–66], enabling automatic feature extraction that largely enhanced classification performance. This area remains active, as recent work [56] increased the detection performance in shadow datasets like SBU [45] by refining noisy labels through iterative label tuning.

Instance Shadow Detection. Wang *et al*. [48, 50, 51] introduced a new task to detect shadows at an instance level. These models aim to detect object-shadow pairs in a scene, leveraging complex modules to form precise associations between objects and their corresponding shadows. This line of work can support object-centered image editing but need some additional post-processing.

2.2. Shadow Removal

Explicit Shadow Removal. A large portion of shadow removal methods [10, 22, 30, 39, 46, 67, 68] require a *shadow mask* as input to explicitly guide the model to remove the shadow pixels indicated by the mask. The performance of these models highly depends on the input, as an imperfect shadow mask may negatively affect the removal quality. Recently, Guo *et al*. [11] proposed a diffusion-based method with an embedded shadow mask refinement branch that refines the input to improve the shadow removal quality.

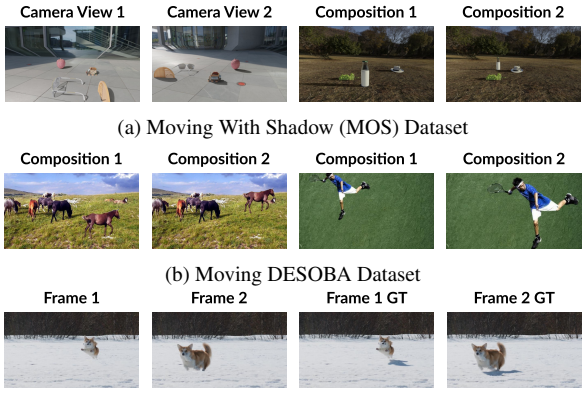
Blind Shadow Removal. More complex shadow removal works attempt to simultaneously detect and remove the shadows in the scene. This formulation allows for the blind removal of shadows and avoids the dependency on a shadow mask input. Early works by Qu *et al*. [35] and Wang *et al*. [47] introduced end-to-end network architectures combining the two tasks. Successive works [5–7, 15, 19, 28, 57] propose new architectures to boost the removal quality. While extensive works [5–7, 10, 11, 15, 19, 22, 27, 28, 30, 35, 39, 46, 47, 54, 57, 67, 68] focus on removing general shadows in a scene, Zhang *et al*. [58] proposed a method to remove an object and its associated shadow, requiring lighting, geometry, and rendering parameters as additional input to achieve realistic results.

2.3. Shadow Synthesis

Evidently, shadow detection and removal are extensively represented in literature, however, shadow synthesis is a relatively underexplored task in the natural image domain. Some methods [26, 60] have been proposed to synthesize shadows for objects in virtual environments for AR applications. Recently, Sheng *et al*. [41–43] focuses on user-driven soft shadow and reflection synthesis, considering environmental variables like light and camera position. Further, [13] introduced SGRNet and its associated DESOBA dataset, the first work to demonstrate object-centered shadow generation using object-shadow pair references without requiring explicit light parameters. SGDiffusion [29], expands on the earlier DESOBA dataset to encompass 21.5K images and adopts ControlNet [59] with a shadow intensity module to improve object-centered shadow synthesis quality. However, its shadow generation quality largely depends on other objects in the scene, which requires off-the-shelf shadow detection models to retrieve the additional input data.

2.4. Joint Frameworks

Some works [5–7, 15, 19, 28, 35, 47] combine shadow detection and removal, while others [6, 7, 16, 30, 67] employ shadow generation for data augmentation to improve the shadow removal quality. Furthermore, recent works leveraging diffusion models [44, 52] for holistic approaches such as object removal or insertion handle shadows in an implicit manner, forfeiting any controllability on the objects' effects on the scene, which is crucial in image-editing applications.



(a) Moving With Shadow (MOS) Dataset

(b) Moving DESOBA Dataset

(c) Video DESOBA Dataset

Figure 2. We construct one training set, *i.e*., MOS Dataset, and two real-world evaluation sets, *i.e*., Moving DESOBA Dataset (without ground truths) and Video DESOBA Dataset (with ground truths), to train and evaluate the effectiveness of MetaShadow.

Yet, none of the existing works handle three object-centered shadow tasks jointly in a knowledge-sharing and mutually beneficial manner, resulting in a limited performance. In contrast, by jointly performing object-centered shadow detection, removal, and synthesis, MetaShadow can greatly boost performance on shadow detection and removal, while our shadow knowledge transfer mechanism leads to more realistic and consistent shadow synthesis.

3. Datasets

Contemporary shadow-related datasets [13, 17, 22, 26, 35, 45, 49] were built for specific shadow tasks. Paired data preparation for shadow removal and shadow synthesis are notably challenging and expensive to collect, leading to the scarcity of real-world datasets. Additionally, only the DESOBA [13] and Shadow AR [26] datasets support object-centered shadow detection, removal, and synthesis. Yet, DESOBA provides only 840 images for training and Shadow-AR provides only 13 3D models for rendering. Their limited scale can severely limit a model's generalizability. Moreover, neither dataset provides samples for object relocation, which is a highly demanded image editing.

To address the limitations of existing datasets for model training, we introduce the **Moving Object with Shadow (MOS) Dataset**, synthesized using the Blender Cycles rendering engine [2], shown in Fig. 2 (a). The dataset consists of 200 scenes, each with eight camera views. In addition, there are five object relocation cases for each scene, resulting in a total of 8,000 image/ground truth pairs. Furthermore, to evaluate the applicability of MetaShadow for real-world scenes, we also introduce two evaluation sets: (i) **Moving DESOBA** and (ii) **Video DESOBA**. (i) For each image in

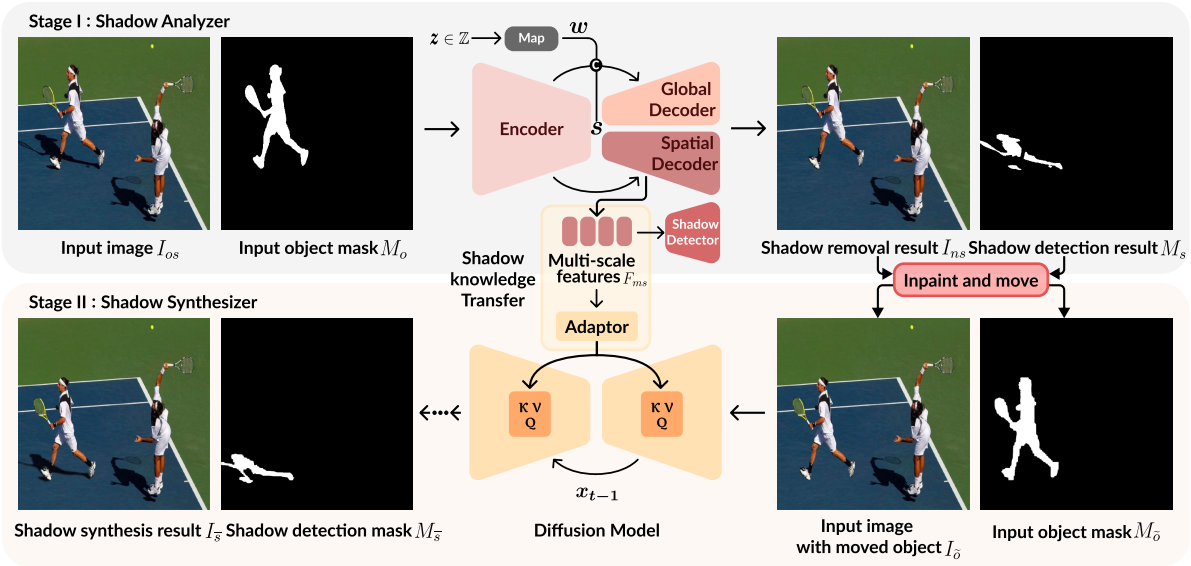


Figure 3. The schematic illustration of our MetaShadow framework. In Stage I, the Shadow Analyzer takes the input image with object mask (left player) to perform object-centered shadow detection and removal. After that, the selected player, together with the detected shadow region, will be moved to a new location. Our Stage II then takes these as input and synthesizes a shadow for this object. To achieve realistic shadow synthesis, we transfer the shadow knowledge extracted from the Shadow Analyzer to Shadow Synthesizer as reference. Note that s represents the global style code, w denotes the intermediate latent space, and "K Q V" stand for key, query, and value in UNet's cross attention layer.

the DESOBA test set [13], we randomly choose an object and reposition it to a different location; see Fig. 2 (b) for examples. (ii) This test set consists of twelve tripod-captured videos with static backgrounds, featuring moving objects casting shadows. Examples are shown in Fig. 2 (c). We will release Moving DESOBA and Video DESOBA for future evaluation; see Supplementary Materials for details.

4. Methodology

As shown in Fig. 3, we design two cooperative components in MetaShadow: (i) **Shadow Analyzer**, an object-centered GAN model that jointly detects and removes an object's shadow by taking an *object mask* and an RGB image as input, and (ii) **Shadow Synthesizer**, a reference-based diffusion model that synthesizes shadows using an object mask from the Shadow Analyzer.

A reference object casting a shadow is often available in image editing scenarios involving natural images. For object insertion, some shadows may already exist in the scene. For object relocation, if the original object shadow is available, such shadow can be used as a reference to improve the consistency in the edited result. Furthermore, the reference shadow can be a way to manipulate the generated shadow as desired, which can be useful for creative editing.

4.1. Shadow Analyzer

Unlike existing shadow removal models that remove shadows within regions in the *shadow mask*, our Shadow Analyzer needs a higher level understanding of the image, especially on the lighting and geometry of the scene, so as to enable it to *identify the object shadow* and remove it. To this end, we base our model architecture on CM-GAN [61], a state-of-the-art image inpainting model, and finetune it from the pretrained CM-GAN weights.

Specifically, as shown in Fig. 3 (top), our Shadow Analyzer has four parts: an Encoder, two parallel cascaded decoders, and a shadow detector. The Encoder extracts multi-scale features F_e^i ($i \in [1, L]$) and global style code $\boldsymbol{s}$ from F_e^L. In the parallel decoders, the cascade of global and spatial modulations utilizes the global code $\boldsymbol{s}$ with style code $\boldsymbol{w}$, mapping from noise $\boldsymbol{z}$, to ensure structural coherence and a spatial code for fine-grained detail, producing output features F_g^i and F_s^i. For details, including the discriminator architecture, please refer to the Supplementary Materials.

More importantly, we integrate a shadow detector alongside the Spatial Decoder, which processes multi-scale features F_s^i. This integration, under the shadow detection supervision, encourages the encoder and parallel decoders to accurately identify the shadow regions. The detector upsamples high-level features (size from 8 to 64) to a uniform size (64×64), concatenating them into a single feature map. It comprises a sequence of convolution layers, batch normalization, and GELU layers, interspersed with transpose convolution layers. The final output is a 256×256 shadow mask, obtained via a sigmoid layer, and subsequently interpolated to match the size of the input image.

The Shadow Analyzer is trained with the original combi-

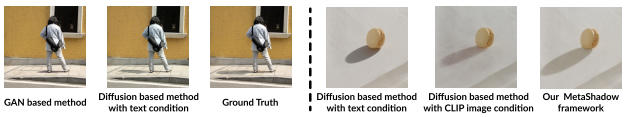


Figure 4. Respective limitations of GAN-based and diffusion-based methods on shadow synthesis. For more discussion, please see Sec. 5.2.

nation of adversarial loss, perceptual loss, masked-R_1 regularization, and the L1 loss from [61]. Also, we adopt the dice loss [32] to compute the losses between the predicted shadow mask and the ground truth.

4.2. Shadow Synthesizer

We illustrate our Shadow Synthesizer in the bottom part of Fig. 3. It is adapted from an inpainting diffusion model based on the DDPM architecture [12] trained similarly to the StableDiffusion inpainting model [38]. To support shadow synthesis, it is modified in the following key ways: (i) We feed an object mask $M_{\widetilde{o}}$ along with the image I_o that contains the moved objects as input. This combination enables the model to identify the specific object for which a shadow needs to be synthesized and understand the desired shape of the shadow. (ii) We incorporate multi-task training into the diffusion model, so that it predicts an additional shadow mask M_s at each diffusion step. (iii) To transfer the shadow knowledge from Shadow Analyzer and align the dimensions from F_{ms}, the multi-scale features with dimensions of $[N, 1348, 32, 32]$, to the original text embeddings, we insert an adaptor $T(\cdot)$ with a 2D convolution layer followed by a 1D convolution layer. Additionally, we employ a Multilayer Perceptron (MLP) layer to increase the embedding dimension from 1344 to 2048, so that the final shadow embedding E_s has dimensions of $[N, 1024, 2048]$, where N denotes the batch size. We then inject it into the diffusion model through a cross-attention mechanism.

The loss function for training our Shadow Synthesizer is

$$\mathcal{L}_{syn} = \mathbb{E}_{T,\epsilon\sim\mathcal{N}(0,1)}\left[\left\|\epsilon - \epsilon_\theta\left(I_o^t, M_{\widetilde{o}}, M_{\widetilde{s}}, t, T(F_{ms})\right)\right\|_2^2\right], \tag{1}$$

where $\epsilon \sim \mathcal{N}(0,1)$ is an initial noise, ϵ_θ denotes the denoising U-Net, and I_o^t is a noisy version of I_o at timestep t. Note that $M_{\widetilde{s}}$ is an optional shadow mask, which is further explained in the supplemental materials.

Discussion. In this work, we unveil a unique insight: the integration of GANs and diffusion models overcomes their respective limitations, enabling a more controlled and realistic object-related image editing. As shown in Fig. 4, we find the GAN excels in effectively and efficiently detecting and removing specific shadows but struggles with synthesizing reasonable shadow shapes [13, 61], as shown in the left part of Fig. 4, whereas diffusion models excel in generating realistic contents but lack precise control for the light direction, color, and intensity of the shadow as shown in the right part of Fig. 4. By conditioning diffusion models with GAN features, we can enable controllable and realistic object-centered shadow editing.

4.3. Training Strategies

Multi-source Dataset Training. As mentioned in Sec. 3, existing shadow-related datasets are limited at scale. As we aim for more general and realistic image editing, we employ multiple datasets to train the MetaShadow framework.

For Shadow Analyzer, we adopt two types of datasets. (i) **Datasets with full annotations**: DESOBA [13] and our MOS dataset contain shadow images, object masks, shadow masks, and shadow-free images. (ii) **Datasets with partial annotations**: ISTD+ [22] and SRD [35] contain shadow images, shadow masks, and shadow-free images. When training on this dataset type, we simply feed an empty object mask and make the model predict general shadows and shadow masks. Also, we randomly make the object mask empty for datasets with full annotations in training. With this data combination strategy, Shadow Analyzer is able to detect an object's shadow with a non-empty object mask, and detect general cast shadows with an empty object mask.

For training the Shadow Synthesizer, we combine MOS, DESOBA [13], and Shadow-AR [26]. During the training, we randomly choose another object as the reference when there are multiple objects in the image. For the MOS dataset, we also use the moved object as the reference.

Shadow-Specific Data Augmentations. We perform three shadow-specific data augmentations to improve the model's generalizability and controllability: (i) Random shadow intensity augmentation, (ii) Curve-based shadow color grading, and (iii) Random shadow dropping. For more details, please refer to the Supplementary Materials.

5. Experiments and Results

Implementation details. We train MetaShadow in two stages. In Stage I, we train Shadow Analyzer for 100 epochs with a learning rate of 0.001 and batch size of 16. We iterate on the DESOBA dataset [13] ten times to balance the number of samples in the multi-dataset training. The training and inference resolution are both 512×512. In Stage II, we freeze the Shadow Analyzer and fine-tune the diffusion U-Net. We also train the Adaptor in Shadow Synthesizer from scratch. The inputs and outputs of Shadow Synthesizer are all at 128×128 resolution with a batch size of 64. In addition, we employ different learning rate strategies for the U-Net and the Adaptor. The learning rate for the U-Net begins at $1e-4$ and is multiplied by 0.01 after 200 epochs (totaling 400 epochs), while the learning rate for the Adaptor remains constant at $1e-4$ to strengthen its ability to gain shadow knowledge. All training stages are conducted on an eight A100-GPU server with the Adam optimizer.

5.1. Comparison with Existing Methods

Though no existing methods aim for the same goal as ours, we evaluate our MetaShadow on four benchmark datasets, including SOBA [48], the DESOBA [13] test set, Moving DESOBA, and Video DESOBA with different methods for different sub-tasks: object-centered shadow detection, removal, and synthesis.

Evaluation on object-centered shadow detection. To evaluate this task, we utilize the common mIoU metric at different sizes, following the COCO [25] definitions with an additional extra small category. Tab. 2 reports the comparison results on the SOBA test set on shadow segmentation quality. As SSIS [50, 51] simultaneously detect all object masks, shadow masks, and their associations, we extract the shadow instance predictions corresponding to each ground-truth object mask for evaluation. Our Shadow Analyzer (with or without using MOS Dataset) significantly outperforms both methods across various shadow scales.

Evaluation on object-centered shadow removal. We employ Masked MAE, Masked RMSE in LAB color space, Bbox PSNR, Bbox SSIM, and PSNR to evaluate the performance on this task. For clarity, *masked* denotes only computing the error inside the ground-truth shadow mask region, and *Bbox* means we compute the error inside the bounding box retrieved from the shadow mask. We join two recent SOTA methods [11, 51] in cascade and also finetune the method [11] on our dataset setting for fair comparisons. The results are reported in Tab. 3. It is evident that training on the original SRD dataset [35] does not result in good generalization on more complex datasets, such as DESOBA [13]. Furthermore, our method outperforms [11] even when using ground-truth masks. We provide comparisons on general shadow removal on the ISTD+ [22] test set with [11, 27, 54] in Tab. 4, which shows that our MetaShadow outperforms most SOTA methods on ISTD+: even without shadow masks, which most SOTA methods still require.

Fig. 5 reveals that before fine-tuning, ShadowDiffusion [11] inadequately recovers shadow regions, leaving residual shadows (1st and 3rd row), or alters other shadows not corresponding to the given object mask (2nd row). After finetuning, it removes shadows but loses detail, causing over-smoothing. Additionally, SSISv2's [51] erroneous detections can lead to incorrect shadow region removal (see 3rd row). In contrast, our Shadow Analyzer preserves details under shadows, ensuring high-quality visuals.

Evaluation on object-centered shadow synthesis. We follow [13] and utilize Global RMSE, Local RMSE, and our Bbox PSNR and Bbox SSIM as the metrics to evaluate the methods on this task with three baselines [13, 29, 34]. In order to compare on the same image size, we upsample our results to 256x256. Please refer to Supplementary Materials for the upsample process. As shown in Tab. 5, our method demonstrates superior shadow-synthesis quality on the DESOBA and Video DESOBA test sets, significantly reducing the Local RMSE to 36.84. Our Shadow Synthesizer also exhibits enhanced performance for real moving object scenarios in Video DESOBA, reducing the Local RMSE to 36.54. Note that since SGRNet [13] utilizes ground-truth shadow parameters for additional supervision. As our datasets lack these shadow parameters, we cannot fine-tune SGRNet using them. We use the official release of SGDiffusion [29] and Libcom [34], trained on the 22K DESOBAv2 dataset. However, we found that SGDiffusion tends to darken images and changes the details, leading to lower performance.

To further show the advantage of our method, we provide various visual comparisons in Fig. 6, presenting results on DESOBA test set [13], our Moving DESOBA, and Video DESOBA. As the reference image: *another* object in the image is used for DESOBA; the original object before relocation is used for Moving DESOBA; the first frame in the

Methods	mIoU	mIoU_{xs}	mIoU_{s}	mIoU_{m}	mIoU_{l}
SSIS [50]	51.6	37.2	46.0	66.7	81.4
SSISv2 [51]	55.8	42.4	49.5	70.4	82.5
Ours wo MOS	67.2	54.5	70.3	79.1	86.5
Ours	**71.0**	**60.4**	**72.6**	**81.1**	**87.8**

Table 2. Comparison with the SOTA shadow-detection methods on the SOBA test set. Note that SSISv2 automatically detects shadow-object instance pairs in the image, whereas our method uses object masks to detect shadows of objects.

Method	Masked MAE ↓	Masked RMSE ↓	Bbox PSNR ↑	Bbox SSIM ↑	PSNR ↑
ShadowDiffusion [11] with GT shadow mask	60.71	17.50	19.42	54.70	25.04
ShadowDiffusion† [11] with SSISv2 [51] detected shadow mask	39.53	12.49	23.41	68.04	39.13
ShadowDiffusion† [11] with GT shadow mask	35.45	11.44	24.28	70.17	40.04
Ours	**21.32**	**6.62**	**32.97**	**96.49**	**42.20**

Table 3. Comparison with SOTA shadow-removal methods on the DESOBA test set. Note that ShadowDiffusion [11] needs a shadow mask as input, so we use the instance shadow mask detected by SSISv2 and the ground-truth shadow mask together to evaluate this method. Note that our MetaShadow, takes an object mask as input. † denotes fine-tuning on our joint datasets.

Method	PSNR / LPIPS / RMSE	Shadow PSNR / RMSE	Non-Shadow PSNR / RMSE
ShadowDiffusion [11]	29.86 / 0.110 / 2.75	26.45 / 3.82	31.46 / 2.46
HomoFormer [54]	30.63 / 0.079 / 2.53	27.25 / 3.42	32.07 / 2.30
RASM [27]	34.66 / 0.061 / 1.64	31.00 / 2.46	36.33 / 1.41
Ours wo shadow mask	31.38 / 0.097 / 2.14	28.22 / 3.12	32.63 / 1.89

Table 4. Comparison with SOTA shadow-removal methods on ISTD+ with image size 512×512.

	Method	Global RMSE ↓	Local RMSE ↓	Bbox PSNR↑	Bbox SSIM↑
DESOBA	SGRNet [13]	4.91	56.44	27.29	91.08
	SGDiffusion [29]	15.03	64.90	21.53	73.57
	Libcom [34]	7.88	67.21	23.90	87.48
	Ours (256 × 256)	**3.12**	**36.84**	**29.16**	**93.56**
	Ours (128 × 128)	**2.93**	**30.92**	**30.73**	**93.49**
Video DESOBA	SGRNet [13]	9.89	51.73	20.77	79.14
	SGDiffusion [29]	12.40	54.15	36.54	76.12
	Libcom [34]	12.52	58.29	19.73	75.60
	Ours (256 × 256)	**8.07**	**36.54**	**23.14**	**82.41**

Table 5. Comparison with the SOTA shadow-synthesis method on the DESOBA subset of the test set with multiple objects in an image and Video DESOBA.

Figure 5. Visual comparison for object-centered shadow detection and removal tasks on the DESOBA test set. † means fine-tuned on our multi-source dataset strategy. Zoom in to see the details. For more results, please refer to the supplementary materials.

Figure 6. Visual comparison for object-centered shadow synthesis on the DESOBA test set [13], our Moving DESOBA test set, and Video DESOBA. Zoom in to see the details. For more results, please refer to the supplementary materials.

video clip is used for Video DESOBA, showing the versatility of our framework. We compare with SGRNet [13] using identical reference objects, and SGDiffusion [29] using the reference shadow masks. Even so, our MetaShadow excels in creating realistic shadows for complex shapes, such as airplanes, with precise color (like the first case in Moving DESOBA), intensity, and direction matching, highlighting its effectiveness. Also, as shown in Fig. 7, SGDiffusion [29] generates inconsistent shadows depending on the random seed. On the contrary, our MetaShadow framework achieves consistent shadow synthesis owing to our shadow knowledge transfer mechanism. See more results, including GIFs of Video DESOBA, in the Supplementary Materials.

Figure 7. Visual comparison on different random seeds reveals a critical issue with the previous diffusion-based method [29]: inconsistent shadow generation across various sampled noises. Empirically, our model does not exhibit this weakness.

Method	Global RMSE ↓	Bbox PSNR↑	Bbox SSIM↑
Baseline 1: SSDM-Text [12, 37]	3.36	29.80	92.21
Baseline 2: SSDM-CLIP [12, 36]	4.51	29.72	93.17
Ours	**2.93**	**30.73**	**93.49**

Table 6. Ablation study on the DESOBA test set.

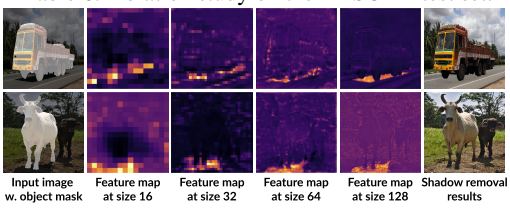


Figure 8. Visualization of the intermediate multi-scale feature maps of our Shadow Analyzer. Given an object mask, Shadow Analyzer detects and removes the shadow for specific objects. Thus, the feature maps effectively capture the shadow knowledge, especially in the target shadow regions.

5.2. Analysis on Shadow Knowledge Transfer

Recently, text-to-image generation models [20, 24, 38] have achieved great success in synthesizing realistic images by injecting the text embedding from the text encoder of a large language model (LLM) (like T5 [37]) or a vision-language model (VLM) (like CLIP [36]) into the diffusion model. Some recent methods [3, 44, 55] replace the text embedding with an image embedding or even combine text and image embeddings [1]. Yet, it is hard for LLMs and VLMs to represent/extract fine-grained features for degradation tasks, e.g., shadowed image, as they are typically trained on diverse web-scale data without specific captions for degradation scenarios [31].

Especially for our Shadow Synthesizer, the condition embeddings should ideally include shadow characteristics such as intensity, softness, color, and direction of the original shadow. Thus, a task-specific encoder for shadow feature extraction would better serve the purpose than a general image encoder. We empirically verify this by comparing the following: (i) "SSDM-Text" representing the Shadow Synthesis Diffusion Model, which has the base architecture [12] of our Shadow Synthesizer but takes the text embedding from T5 [37] as the condition with the word "shadow" as the text prompt, and (ii) "SSDM-CLIP" representing SSDM with CLIP [36] image embeddings as condition, replacing F_{ms}. Note that as the original CLIP [36] image encoder faces the challenge of extracting sufficient shadow knowledge directly from the image, we finetune it with the diffusion model to strengthen its ability. Note that both baselines are trained with the same dataset as Ours until convergence. Table 6 reports the comparison results, showing that image embeddings are more effective than text embeddings. Even though CLIP [36] is widely used to encode image information [44, 55], it performs sub-optimally compared to our task-specific shadow knowledge transfer.

We further visualize feature maps from Shadow Analyzer, which distinctly highlight the response of the shadow regions (Fig. 8), demonstrating the Analyzer's effectiveness in capturing shadow characteristics. However, we further observed that larger resolution features gradually include texture information within the shadow region, which is not desired, as we solely want to transfer the shadow properties, not the texture from previous locations. Based on this observation, we use features of varying sizes (16 to 128) and resize to a uniform 32×32 size as mentioned in Sec. 4.2.

Our design additionally offers a significant advantage by requiring only four steps to generate the final result, in contrast to SSDM's 30 steps or SGDiffusion's [29] 50 steps. All results presented in this paper and Supplemental Materials are based on this four-step setting.

In the Supplemental Material, we delve deeper into detailed analyses of the dataset and data augmentation techniques through ablation studies, along with more comparisons and visualizations on different sub-tasks as well as the limitation and potential solution.

6. Conclusion

In this work, we introduced MetaShadow, a novel framework for enhancing realism in image editing through advanced shadow manipulation. By integrating Shadow Analyzer for precise shadow detection and removal, and Shadow Synthesizer for controllable shadow generation, MetaShadow achieves a significant leap in object-centered image processing. This synergy ensures that shadows are not only realistic but also contextually harmonized with the scene, eliminating the need for complex systems requiring lighting and geometry parameters. Our evaluations demonstrate MetaShadow's superior performance over existing methods, with notable improvements in object-centered shadow detection, removal, and synthesis. This framework enables various image-editing tasks, such as object removal, relocation, and insertion, showcasing its potential to advance object-centered image editing techniques.

References

[1] Yogesh Balaji, Seungjun Nah, Xun Huang, Arash Vahdat, Jiaming Song, Qinsheng Zhang, Karsten Kreis, Miika Aittala, Timo Aila, Samuli Laine, Bryan Catanzaro, Tero Karras, and Ming-Yu Liu. eDiff-I: Text-to-image diffusion models with ensemble of expert denoisers. *arXiv preprint arXiv:2211.01324*, 2022. 8

[2] Blender Foundation. Blender. https://www.blender.org, 2023. 3

[3] Xi Chen, Lianghua Huang, Yu Liu, Yujun Shen, Deli Zhao, and Hengshuang Zhao. AnyDoor: Zero-shot object-level image customization. *arXiv preprint arXiv:2307.09481*, 2023. 8

[4] Zhihao Chen, Lei Zhu, Liang Wan, Song Wang, Wei Feng, and Pheng-Ann Heng. A multi-task mean teacher for semi-supervised shadow detection. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 5611–5620, 2020. 2

[5] Zipei Chen, Chengjiang Long, Ling Zhang, and Chunxia Xiao. CANet: A context-aware network for shadow removal. In *IEEE International Conference on Computer Vision*, 2021. 3

[6] Xiaodong Cun, Chi-Man Pun, and Cheng Shi. Towards ghost-free shadow removal via dual hierarchical aggregation network and shadow matting GAN. In *AAAI Conference on Artificial Intelligence*, 2020. 2, 3

[7] Bin Ding, Chengjiang Long, Ling Zhang, and Chunxia Xiao. ARGAN: Attentive recurrent generative adversarial network for shadow detection and removal. In *IEEE International Conference on Computer Vision*, 2019. 3

[8] Xianyong Fang, Xiaohao He, Linbo Wang, and Jianbing Shen. Robust shadow detection by exploring effective shadow contexts. In *Proceedings of the 29th ACM International Conference on Multimedia*, pages 2927–2935, 2021. 2

[9] Google. Magic Eraser. https://blog.google/products/photos/magic-eraser-android-ios-google-one/, 2023. 1

[10] Lanqing Guo, Siyu Huang, Ding Liu, Hao Cheng, and Bihan Wen. ShadowFormer: Global context helps image shadow removal. In *AAAI Conference on Artificial Intelligence*, 2023. 3

[11] Lanqing Guo, Chong Wang, Wenhan Yang, Siyu Huang, Yufei Wang, Hanspeter Pfister, and Bihan Wen. ShadowDiffusion: When degradation prior meets diffusion model for shadow removal. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2023. 2, 3, 6

[12] Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising diffusion probabilistic models. In *Advances in Neural Information Processing Systems*, 2020. 5, 8

[13] Yan Hong, Li Niu, and Jianfu Zhang. Shadow generation for composite image in real-world scenes. In *AAAI Conference on Artificial Intelligence*, 2022. 2, 3, 4, 5, 6, 7

[14] Xiaowei Hu, Lei Zhu, Chi-Wing Fu, Jing Qin, and Pheng-Ann Heng. Direction-aware spatial context features for shadow detection. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 7454–7462, 2018. 2

[15] Xiaowei Hu, Chi-Wing Fu, Lei Zhu, Jing Qin, and Pheng-Ann Heng. Direction-aware spatial context features for shadow detection and removal. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2019. 3

[16] Xiaowei Hu, Yitong Jiang, Chi-Wing Fu, and Pheng-Ann Heng. Mask-ShadowGAN: Learning to remove shadows from unpaired data. In *IEEE International Conference on Computer Vision*, pages 2472–2481, 2019. 3

[17] Xiaowei Hu, Tianyu Wang, Chi-Wing Fu, Yitong Jiang, Qiong Wang, and Pheng-Ann Heng. Revisiting shadow detection: A new benchmark dataset for complex world. *IEEE Transactions on Image Processing*, 30:1925–1934, 2021. 2, 3

[18] Leiping Jie and Hui Zhang. RMLANet: Random multi-level attention network for shadow detection. In *IEEE International Conference on Multimedia and Expo*, pages 1–6. IEEE, 2022. 2

[19] Yeying Jin, Aashish Sharma, and Robby T. Tan. DC-ShadowNet: Single-image hard and soft shadow removal using unsupervised domain-classifier guided network. In *IEEE International Conference on Computer Vision*, 2021. 3

[20] Bahjat Kawar, Shiran Zada, Oran Lang, Omer Tov, Huiwen Chang, Tali Dekel, Inbar Mosseri, and Michal Irani. Imagic: Text-based real image editing with diffusion models. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2023. 8

[21] Salman Hameed Khan, Mohammed Bennamoun, Ferdous Sohel, and Roberto Togneri. Automatic feature learning for robust shadow detection. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 1939–1946, 2014. 2

[22] Hieu Le and Dimitris Samaras. Shadow removal via shadow image decomposition. In *ICCV*, 2019. 2, 3, 5, 6

[23] Hieu Le, Tomás F. Yago Vicente, Vu Nguyen, Minh Hoai, and Dimitris Samaras. A+D Net: Training a shadow detector with adversarial shadow attenuation. In *European Conference on Computer Vision*, pages 662–678, 2018. 2

[24] Yuheng Li, Haotian Liu, Qingyang Wu, Fangzhou Mu, Jianwei Yang, Jianfeng Gao, Chunyuan Li, and Yong Jae Lee. GLIGEN: Open-set grounded text-to-image generation. *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2023. 8

[25] Tsung-Yi Lin, Michael Maire, Serge J. Belongie, Lubomir D. Bourdev, Ross B. Girshick, James Hays, Pietro Perona, Deva Ramanan, Piotr Doll'a r, and C. Lawrence Zitnick. Microsoft COCO: common objects in context. *CoRR*, abs/1405.0312, 2014. 6

[26] Daquan Liu, Chengjiang Long, Hongpan Zhang, Hanning Yu, Xinzhi Dong, and Chunxia Xiao. ARShadowGAN: Shadow generative adversarial network for augmented reality in single light scenes. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 8139–8148, 2020. 3, 5

[27] Hengxing Liu, Mingjia Li, and Xiaojie Guo. Regional attention for shadow removal. 2024. 3, 6

[28] Jiawei Liu, Qiang Wang, Huijie Fan, Wentao Li, Liangqiong Qu, and Yandong Tang. A decoupled multi-task network for shadow removal. *IEEE Transactions on Multimedia*, 2023. 2, 3

[29] Qingyang Liu, Junqi You, Jianting Wang, Xinhao Tao, Bo Zhang, and Li Niu. Shadow generation for composite image using diffusion model. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2024. 2, 3, 6, 7, 8

[30] Zhihao Liu, Hui Yin, Xinyi Wu, Zhenyao Wu, Yang Mi, and Song Wang. From shadow generation to shadow removal. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2021. 3

[31] Ziwei Luo, Fredrik K. Gustafsson, Zheng Zhao, Jens Sjölund, and Thomas B. Schön. Controlling vision-language models for universal image restoration. *arXiv preprint arXiv:2310.01018*, 2023. 8

[32] Fausto Milletari, Nassir Navab, and Seyed-Ahmad Ahmadi. V-Net: Fully convolutional neural networks for volumetric medical image segmentation. In *International Conference on 3D Vision*, 2016. 5

[33] Vu Nguyen, Tomás F. Yago Vicente, Maozheng Zhao, Minh Hoai, and Dimitris Samaras. Shadow detection with conditional generative adversarial networks. In *IEEE International Conference on Computer Vision*, pages 4510–4518, 2017. 2

[34] Li Niu, Wenyan Cong, Liu Liu, Yan Hong, Bo Zhang, Jing Liang, and Liqing Zhang. Making images real again: A comprehensive survey on deep image composition. *arXiv preprint arXiv:2106.14490*, 2021. 6

[35] Liangqiong Qu, Jiandong Tian, Shengfeng He, Yandong Tang, and Rynson W.H. Lau. DeshadowNet: A multi-context embedding deep network for shadow removal. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 4067–4075, 2017. 3, 5, 6

[36] Alec Radford, Jong Wook Kim, Chris Hallacy, Aditya Ramesh, Gabriel Goh, Sandhini Agarwal, Girish Sastry, Amanda Askell, Pamela Mishkin, Jack Clark, et al. Learning transferable visual models from natural language supervision. In *International Conference on Machine Learning*, 2021. 8

[37] Colin Raffel, Noam Shazeer, Adam Roberts, Katherine Lee, Sharan Narang, Michael Matena, Yanqi Zhou, Wei Li, and Peter J. Liu. Exploring the limits of transfer learning with a unified text-to-text transformer. *Journal of Machine Learning Research*, 2020. 8

[38] Robin Rombach, Andreas Blattmann, Dominik Lorenz, Patrick Esser, and Björn Ommer. High-resolution image synthesis with latent diffusion models. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022. 5, 8

[39] Mrinmoy Sen, Sai Pradyumna Chermala, Nazrinbanu Nurmohammad Nagori, Venkat Peddigari, Praful Mathur, B. H. Pawan Prasad, and Moonhwan Jeong. SHARDS: Efficient shadow removal using dual stage network for high-resolution images. In *IEEE/CVF Winter Conference on Applications of Computer Vision*, 2023. 3

[40] Li Shen, Teck Wee Chua, and Karianto Leman. Shadow optimization from structured deep edge detection. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 2067–2074, 2015. 2

[41] Yichen Sheng, Jianming Zhang, and Bedrich Benes. SSN: Soft shadow network for image compositing. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2021. 3

[42] Yichen Sheng, Yifan Liu, Jianming Zhang, Wei Yin, A. Cengiz Oztireli, He Zhang, Zhe Lin, Eli Shechtman, and Bedrich Benes. Controllable shadow generation using pixel height maps. In *European Conference on Computer Vision*, 2022. 2

[43] Yichen Sheng, Jianming Zhang, Julien Philip, Yannick Hold-Geoffroy, Xin Sun, He Zhang, Lu Ling, and Bedrich Benes. PixHt-Lab: Pixel height based light effect generation for image compositing. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2023. 2, 3

[44] Yizhi Song, Zhifei Zhang, Zhe Lin, Scott Cohen, Brian Price, Jianming Zhang, Soo Ye Kim, and Daniel Aliaga. Objectstitch: Object compositing with diffusion model. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2023. 3, 8

[45] Tomás F. Yago Vicente, Le Hou, Chen-Ping Yu, Minh Hoai, and Dimitris Samaras. Large-scale training of shadow detectors with noisily-annotated shadow examples. In *European Conference on Computer Vision*, pages 816–832, 2016. 2, 3

[46] Jin Wan, Hui Yin, Zhenyao Wu, Xinyi Wu, Yanting Liu, and Song Wang. Style-guided shadow removal. In *European Conference on Computer Vision*, 2022. 2, 3

[47] Jifeng Wang, Xiang Li, and Jian Yang. Stacked conditional generative adversarial networks for jointly learning shadow detection and shadow removal. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 1788–1797, 2018. 2, 3

[48] Tianyu Wang, Xiaowei Hu, Qiong Wang, Pheng-Ann Heng, and Chi-Wing Fu. Instance shadow detection. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020. 2, 6

[49] Tianyu Wang*, Xiaowei Hu*, Qiong Wang, Pheng-Ann Heng, and Chi-Wing Fu. Instance shadow detection. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 1880–1889, 2020. * Joint first authors. 3

[50] Tianyu Wang, Xiaowei Hu, Chi-Wing Fu, and Pheng-Ann Heng. Single-stage instance shadow detection with bidirectional relation learning. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 1–11, 2021. 2, 6

[51] Tianyu Wang, Xiaowei Hu, Pheng-Ann Heng, and Chi-Wing Fu. Instance shadow detection with a single-stage detector. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pages 1–14, 2022. 2, 6

[52] Daniel Winter, Matan Cohen, Shlomi Fruchter, Yael Pritch, Alex Rav-Acha, and Yedid Hoshen. Objectdrop: Bootstrapping counterfactuals for photorealistic object removal and insertion, 2024. 2, 3

[53] Wen Wu, Kai Zhou, Xiao-Diao Chen, and Jun-Hai Yong. Light-weight shadow detection via GCN-based annotation strategy and knowledge distillation. *Computer Vision and Image Understanding*, 216:103341, 2022. 2

[54] Jie Xiao, Xueyang Fu, Yurui Zhu, Dong Li, Jie Huang, Kai Zhu, and Zheng-Jun Zha. HomoFormer: Homogenized transformer for image shadow removal. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2024. 3, 6

[55] Binxin Yang, Shuyang Gu, Bo Zhang, Ting Zhang, Xuejin Chen, Xiaoyan Sun, Dong Chen, and Fang Wen. Paint by example: Exemplar-based image editing with diffusion models.

In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2023. 8

[56] Han Yang, Tianyu Wang, Xiaowei Hu, and Chi-Wing Fu. SILT: Shadow-aware iterative label tuning for learning to detect shadows from noisy labels. In *IEEE International Conference on Computer Vision*, 2023. 1, 2

[57] Mehmet Kerim Yücel, Valia Dimaridou, Bruno Manganelli, Mete Ozay, Anastasios Drosou, and Albert Saà-Garriga. LRA&LDRA: Rethinking residual predictions for efficient shadow detection and removal. In *IEEE/CVF Winter Conference on Applications of Computer Vision*, 2023. 3

[58] Edward Zhang, Ricardo Martin-Brualla, Janne Kontkanen, and Brian L. Curless. No shadow left behind: Removing objects and their shadows using approximate lighting and geometry. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2021. 2, 3

[59] Lvmin Zhang, Anyi Rao, and Maneesh Agrawala. Adding conditional control to text-to-image diffusion models. In *IEEE International Conference on Computer Vision*, pages 3836–3847, 2023. 3

[60] Shuyang Zhang, Runze Liang, and Miao Wang. ShadowGAN: Shadow synthesis for virtual objects with conditional adversarial networks. *CVM*, 2019. 3

[61] Haitian Zheng, Zhe Lin, Jingwan Lu, Scott Cohen, Eli Shechtman, Connelly Barnes, Jianming Zhang, Ning Xu, Sohrab Amirghodsi, and Jiebo Luo. CM-GAN: Image inpainting with cascaded modulation GAN and object-aware training. In *European Conference on Computer Vision*, 2022. 4, 5

[62] Quanlong Zheng, Xiaotian Qiao, Ying Cao, and Rynson W.H. Lau. Distraction-aware shadow detection. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 5167–5176, 2019. 2

[63] Kai Zhou, Wen Wu, Yan-Li Shao, Jing-Long Fang, Xing-Qi Wang, and Dan Wei. Shadow detection via multi-scale feature fusion and unsupervised domain adaptation. *Journal of Visual Communication and Image Representation*, 88:103596, 2022.

[64] Lei Zhu, Zijun Deng, Xiaowei Hu, Chi-Wing Fu, Xuemiao Xu, Jing Qin, and Pheng-Ann Heng. Bidirectional feature pyramid network with recurrent attention residual modules for shadow detection. In *European Conference on Computer Vision*, pages 121–136, 2018.

[65] Lei Zhu, Ke Xu, Zhanghan Ke, and Rynson W.H. Lau. Mitigating intensity bias in shadow detection via feature decomposition and reweighting. In *IEEE International Conference on Computer Vision*, pages 4702–4711, 2021. 1

[66] Yurui Zhu, Xueyang Fu, Chengzhi Cao, Xi Wang, Qibin Sun, and Zheng-Jun Zha. Single image shadow detection via complementary mechanism. In *Proceedings of the 30th ACM International Conference on Multimedia*, pages 6717–6726, 2022. 1, 2

[67] Yurui Zhu, Jie Huang, Xueyang Fu, Feng Zhao, Qibin Sun, and Zheng-Jun Zha. Bijective mapping network for shadow removal. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022. 2, 3

[68] Yurui Zhu, Zeyu Xiao, Yanchi Fang, Xueyang Fu, Zhiwei Xiong, and Zheng-Jun Zha. Efficient model-driven network for shadow removal. In *AAAI Conference on Artificial Intelligence*, 2022. 2, 3