TO INPAINT, OR NOT TO INPAINT: AUXILIARY-FREE SHADOW REMOVAL VIA INPAINTING AND POST FUSION

Anonymous authors
Paper under double-blind review

ABSTRACT

Shadows degrade downstream vision tasks, yet most shadow removal methods depend on auxiliary inputs such as shadow masks, surface normals, or depth maps, which are often unavailable or inaccurate at deployment. We propose Inpaint2Free (I2F), an auxiliary-free shadow removal framework that exploits the mask conditioning of a pretrained inpainting diffusion model. I2F is fine-tuned in two stages: first with ground-truth shadow masks to learn localized shadow removal, then with Full-one masks so that inference requires only the shadow image. We further propose Post Fusion, which estimates a high-recall pseudo mask from the difference between the removal result and the input, and uses it to restore non-shadow details from the original image. On ADSP, the two-stage fine-tuning substantially improves shadow restoration, while Post Fusion markedly improves non-shadow and full-image fidelity. When compared with mask-dependent baselines using FDRNet-predicted masks on ADSP, auxiliary-free I2F achieves a shadow-region PSNR of 41.178 dB, outperforming all baselines. The framework thus removes the dependence on both auxiliary inputs and upstream prediction models, providing a practical standalone solution for real-world shadow removal.

1 INTRODUCTION

Shadows arise when a light source is partially or fully occluded, leaving regions of reduced illumination that corrupt visual cues such as texture and color. They degrade or even break downstream vision algorithms, including recognition (Zhang et al., 2019), detection (Cucchiara et al., 2001; Sultana et al., 2021; Zhang et al., 2007), segmentation (Le et al., 2019; Stander et al., 1999), tracking (Sanin et al., 2010; Cucchiara et al., 2003; Saravanakumar et al., 2010; KaewTraKulPong & Bowden, 2002; Danelljan et al., 2014), intrinsic image decomposition (Li & Snavely, 2018; Fu et al., 2019), and other tasks (Shin et al., 2022; Zhong et al., 2022; Su et al., 2016). Shadow removal, which recovers the underlying background while keeping the restored and non-shadow regions visually consistent, is therefore a long-standing image restoration task. Ideally, a deployed model should generalize across diverse scenes without relying on externally provided scene priors or auxiliary information.

Early shadow removal methods were prior-based, exploiting gradient (Finlayson et al., 2006; Liu & Gleicher, 2008; Finlayson et al., 2009), illumination (Finlayson et al., 2002; Yang et al., 2012; Zhang et al., 2015), and texture (Vicente & Samaras, 2015; Lalonde et al., 2010) cues. Deep learning then

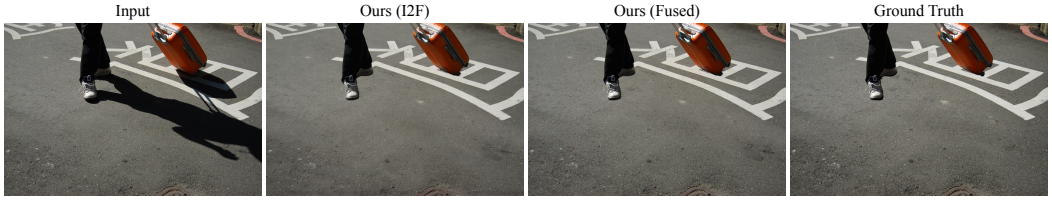

Figure 1: **Illustration of our auxiliary-free shadow removal framework.** Given a shadow image as input, the proposed Inpaint2Free, fine-tuned in two stages, effectively removes shadows without requiring external auxiliary information. Then, the proposed Post Fusion further preserves the original content in non-shadow regions, producing a result closer to the ground truth.

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

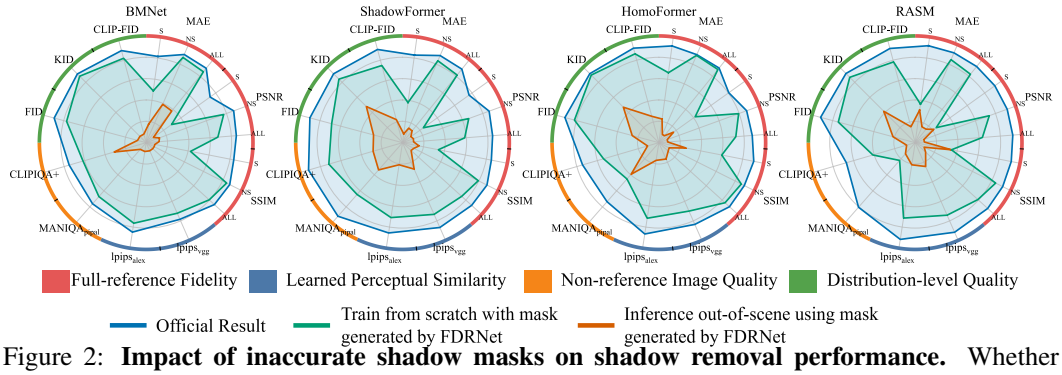

Figure 2: **Impact of inaccurate shadow masks on shadow removal performance.** Whether trained with FDRNet (Zhu et al., 2021) masks from scratch (green polygons) or directly evaluated with them (orange polygons), all models exhibit performance degradation relative to their original results with ground-truth masks (blue polygons). All metrics are normalized such that larger radial values indicate better performance.

shifted the field to CNN-, Transformer-, and diffusion-based models that directly learn the shadow-to-shadow-free mapping. To reduce ambiguity and provide explicit spatial or structural guidance, many of these methods incorporate auxiliary inputs such as shadow masks, depth maps, and surface normals (Guo et al., 2023a; Lee et al., 2026), while others operate on the shadow image alone (Xu et al., 2025b; Jin et al., 2024). Despite substantial quality gains, their varying dependence on auxiliary information remains a key concern for practical deployment.

This dependence introduces three limitations. First, accurate manually annotated auxiliary information is unavailable in real-world scenarios, so results obtained with ideal auxiliary inputs may not reflect deployed performance, as shown in fig. 2; imperfect inputs degrade performance even when models are retrained on predicted masks. Second, estimating auxiliary information on the fly with external models shifts the burden of reliability to upstream estimators, making overall performance dependent on their generalization and susceptible to error propagation. Third, discarding auxiliary information entirely is the most direct remedy, but it forfeits demonstrably useful cues such as the shadow localization provided by masks, which may constrain the achievable removal quality.

We therefore propose a framework that uses auxiliary information in a sufficient yet restrained manner, with two key components, as qualitatively illustrated in fig. 1. First, we adopt a pretrained inpainting model as the backbone and fine-tune it in two stages. Since inpainting models use masks to specify the regions to be processed, they naturally adapt to auxiliary-free shadow removal; we thus name our framework Inpaint2Free (I2F). Stage one trains with the human-annotated ground-truth (GT) masks from triplet datasets, restricting the inpainting objective to shadow regions. Stage two replaces them with all-white masks, which we call Full-one masks, so the model learns to restore the entire image. I2F can hence inpaint with an external shadow mask or remove shadows without one, and our experiments show that the two-stage strategy improves removal performance. Second, to mitigate the global quality degradation caused by auxiliary-free removal, we propose Post Fusion, a post-processing step. It computes a pixel-wise difference map between the removal result and the input shadow image, binarizes it with our Gaussian Mixture Model (GMM)-based thresholding followed by simple refinement to obtain a pseudo mask, and uses this mask to fuse the removal result with the original input: restored content is kept in shadow regions for effective removal and natural transitions, while original content is kept elsewhere to preserve fine details and avoid artifacts. Post Fusion substantially improves non-shadow and full-image quality, and the GMM-based thresholding achieves both high shadow coverage and high non-shadow preservation.

In summary, our contributions are threefold:

- A two-stage training strategy for an inpainting backbone: the model first learns spatially guided shadow restoration with annotated masks, then is fine-tuned with Full-one masks to shed its dependence on external auxiliary information, so that inference requires only a shadow image.
- Post Fusion, which estimates a pseudo shadow mask from the difference between the removal result and the shadow image and uses it to retain restored content in shadow regions while preserving original content elsewhere, balancing effective shadow removal and detail preservation.

- Extensive experiments on multiple benchmarks validate that the training strategy and Post Fusion reduce auxiliary dependence, improve restoration quality, and preserve non-shadow details, achieving competitive performance in an auxiliary-free setting that better reflects practical deployment.

2 Related Work

Shadow Removal. Early methods (Finlayson et al., 2006; 2002; Lalonde et al., 2010; Guo et al., 2013) relied on physical priors and hand-crafted assumptions, and generalized poorly when those assumptions were violated. Learning-based methods first used CNNs or encoder–decoder networks to map shadow images to shadow-free ones (Le & Samaras, 2022; Xu et al., 2024; Liu et al., 2023; Yang et al., 2023; Wang et al., 2024), later adding refinement (Zhang et al., 2020), attention (Sen et al., 2023; Zhu et al., 2025), and adversarial learning (Vasluianu et al., 2020; Dasgupta et al., 2023; Lin et al., 2025a). Transformers have since been adopted to model long-range dependencies (Wan et al., 2022b; Li et al., 2025; Hu et al., 2024; Dong et al., 2025; Chen et al., 2025), and diffusion models to restore shadow-affected regions (Zeng et al., 2024; Mei et al., 2024; Wang et al., 2025; Zeng et al., 2025). These methods differ considerably in their use of shadow localization, scene geometry, and other auxiliary information, motivating the two categories below.

Auxiliary-Guided Shadow Removal. Since Wang et al. (2018) introduced joint shadow detection and removal with the ISTD triplet dataset, shadow masks have become the dominant auxiliary input. Most methods (Le & Samaras, 2020; Fu et al., 2021; Gao et al., 2022; Yu et al., 2022; Guo et al., 2023b; Li et al., 2023; Jiang et al., 2023; Guo et al., 2023c; Liu et al., 2024b; Luo et al., 2024; Wang et al., 2026b) take the mask as an additional input, making inference depend on externally provided or predicted masks, whereas others (Hu et al., 2019; Vasluianu et al., 2021; Liu et al., 2021; Chen et al., 2021) use masks only for training supervision or guidance. Recent methods (Xu et al., 2025a; Lin et al., 2025b; Lee et al., 2026; Wang et al., 2026a) further incorporate depth or surface normal maps, shifting rather than removing the dependence on external auxiliary information at deployment.

Auxiliary-Free Shadow Removal. Auxiliary-free methods avoid this deployment dependency. Early works (Gong & Cosker, 2016; Ma et al., 2016; Qu et al., 2017b) predate mask-annotated triplet datasets and naturally follow this setting. Jin et al. (2023) guide shadow-aware feature transformation with shadow/shadow-free domain classifiers and estimate a soft mask from image differences for boundary supervision. The NTIRE 2023 (Vasluianu et al., 2023b; Cui et al., 2023; Chang et al., 2023) and 2024 (Vasluianu et al., 2024; Lu et al., 2024; Dong et al., 2024) challenges adopted WSRD (Vasluianu et al., 2023a), an indoor benchmark without mask annotations, encouraging mask-free designs such as a two-stage network that estimates the mask for its refinement stage from the initial result, a pyramid-based ensemble, HirFormer, and ShadowRefine, albeit in a setting different from the outdoor scenarios considered here. Other works reduce mask dependence differently: Kubiak et al. (2024) use estimated masks only for loss computation, Jin et al. (2024) combine adaptive attention with pretrained DINO-ViT similarity guidance during diffusion sampling, and Xu et al. (2025b) inject DINOv2-based shadow-aware features. All obtain shadow-related guidance internally while avoiding externally supplied masks at inference time.

3 Method

3.1 Inpaint2Free (I2F)

Unlike many restoration tasks in which degradation may be spatially distributed across the image, shadows typically exhibit explicit spatial support. This characteristic naturally aligns with the region-conditioned restoration mechanism of inpainting models. We therefore exploit the flexible mask conditioning of an inpainting backbone to progressively transition the model from mask-guided to auxiliary-free removal. Figure 3 provides an overview of the proposed Inpaint2Free (I2F) model. We employ Stable Diffusion 2 Inpainting (Rombach et al., 2022) as the backbone and introduce a ControlNet (Zhang et al., 2023) conditioned on the shadow image to reformulate shadow removal as a conditional generation task. We further insert LoRA (Hu et al., 2021) layers into the attention modules of the latent diffusion UNet, including the query, key, and value projections and the feed-forward networks, for task-specific adaptation. During training, only the ControlNet and LoRA parameters are updated, while the remaining backbone parameters are frozen.

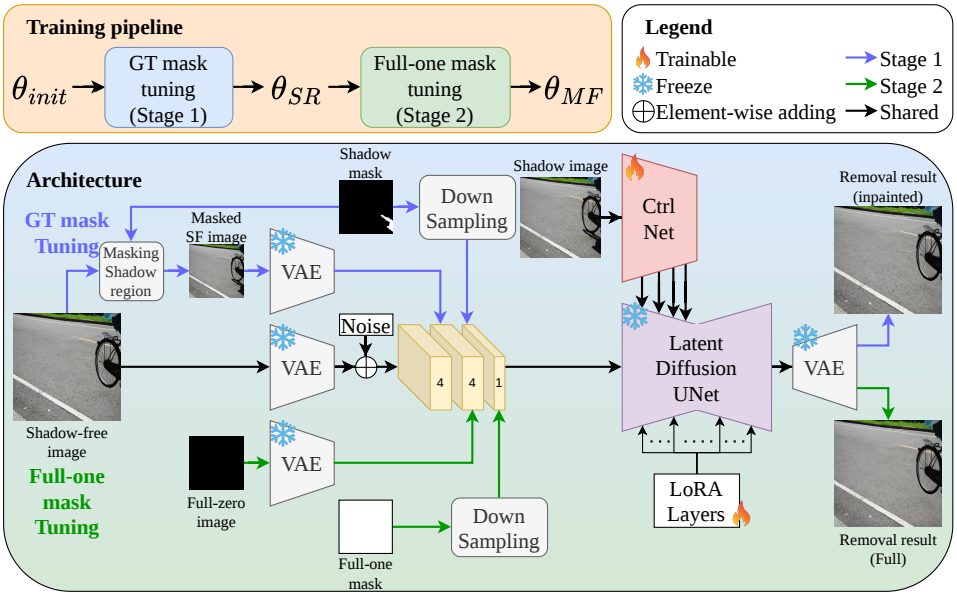

Figure 3: **Overview of the proposed two-stage training framework.** Starting from the initialized model θ_{init}, the first stage performs standard inpainting training with ground-truth shadow masks to obtain the shadow-removal model θ_{SR} (blue). The second stage replaces the ground-truth mask with a Full-one mask for global inpainting, yielding the mask-free model θ_{MF} (green). The two stages therefore differ mainly in the mask used to construct the inpainting input.

We adopt a two-stage fine-tuning strategy. Starting from the initial parameters θ_{init}, the model is first fine-tuned using manually annotated ground-truth (GT) shadow masks, resulting in a shadow-removal model denoted as θ_{SR}. We refer to this stage as GT mask tuning, as illustrated in blue in fig. 3. The GT mask is used to construct a masked target representation that preserves only the non-shadow regions. The mask is separately downsampled to obtain the corresponding mask latent. Subsequently, θ_{SR} is further fine-tuned using a Full-one mask, yielding the mask-free model θ_{MF}, as illustrated in green in fig. 3. In this stage, the masked target latent becomes all-zero and the mask latent becomes all-one. Since the entire image is specified for restoration, the mask no longer provides image-specific shadow-location information, leaving the shadow image introduced through ControlNet as the only image-dependent condition. Consequently, θ_{SR} learns shadow removal under explicit spatial guidance, while θ_{MF} performs removal without requiring external auxiliary information at inference time.

3.2 Post Fusion

θ_{MF} satisfies the requirement of mask-free shadow removal, yet some image details are inevitably lost. This degradation mainly stems from two factors: (1) θ_{MF} lacks a GT mask for preserving the original content in non-shadow regions, and (2) latent diffusion models suffer from VAE reconstruction errors introduced in exchange for computational efficiency. Therefore, following the same principle as θ_{SR}, we propose Post Fusion to further improve fidelity. Figure 4 presents an overview of Post Fusion and its core component, pseudo-mask generation. We begin by computing the pixel-wise difference between the shadow image and the removal result, yielding a continuous-valued difference map that reveals shadow-related spatial responses. We convert the difference map into a binarized map by thresholding, and refine it into the final pseudo mask. The resulting pseudo mask is used for content fusion, preserving the removal result in shadow regions while restoring the original content in non-shadow regions. The key to Post Fusion lies in accurately estimating the shadow mask, which relies on the low-frequency recovery capability of I2F.

We propose a thresholding strategy based on Gaussian Mixture Model (GMM) analysis, as illustrated by the yellow region in fig. 4. Pixels in shadow regions generally exhibit larger differences after shadow removal. We refer to this difference as the shadow response. We then model the distribution using a one-component and a two-component GMM, denoted by G^1 and G^2, respectively. We use $\Delta\mathrm{BIC} = \mathrm{BIC}_1 - \mathrm{BIC}_2$ to determine which model better explains the distribution.

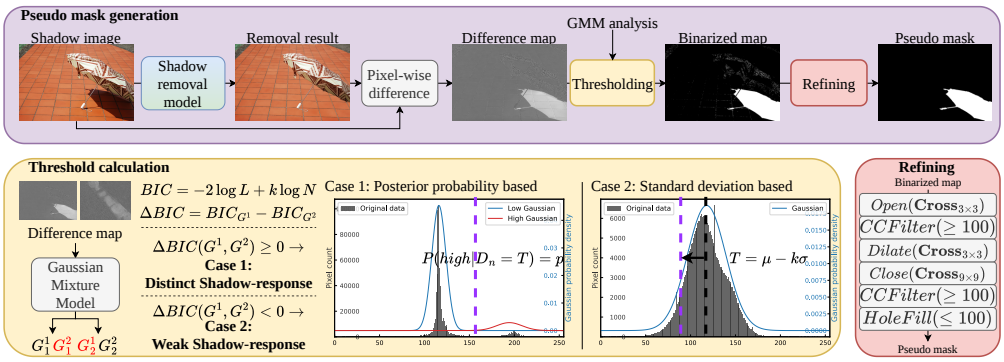

Figure 4: **Overview of the pseudo-mask generation pipeline.** A difference map is computed from the shadow input and its removal result, from which an adaptive threshold is determined through GMM analysis using one of two thresholding strategies. The resulting binary map is then refined to obtain the final pseudo mask. Here, G_i denotes a GMM with i Gaussian components; BIC denotes Bayesian Information Criterion (Schwarz, 1978); $Open$, $Close$, and $Dilate$ denote morphological operations; $Cross_{j\times j}$ denotes a cross-shaped kernel of size $j \times j$; $CCFilter$ denotes connected components filter; $HoleFill$ denotes hole filling.

A nonnegative ΔBIC indicates that G_2 provides a better fit, suggesting a clearer separation between low- and high-response populations. We refer to this as a distinct shadow response (Case 1). The desired threshold is expected to lie between two Gaussian components associated with shadow and non-shadow regions. Thus, we adopt posterior-probability-based thresholding, locating the response value where the posterior probability of belonging to the high-response Gaussian reaches a predefined value p. Conversely, when $\Delta\text{BIC} < 0$, G^1 provides a better fit, indicating a weak shadow response (Case 2). Since G^1 is largely dominated by non-shadow pixels, we adopt standard-deviation-based thresholding and place the threshold on the left side of the G^1 distribution, at $\mu - k\sigma$. We empirically set $p = 0.15$ and $k = 0.3$. Both p and k are selected with a recall-oriented objective, prioritizing the inclusion of potentially shadowed pixels over strict precision.

Finally, the thresholded mask is refined using morphological operations, connected-component analysis, and hole filling, as illustrated by the red region in fig. 4. Following the recall-oriented design, region-removing operations are applied conservatively, whereas region-expanding operations are more aggressive to reduce false negatives and improve mask continuity.

4 Experiments

4.1 Experimental Setup

We evaluate our method on ADSP (Hsieh & Ding, 2025), SRD (Qu et al., 2017a), and ISTD+ (Le & Samaras, 2019) using their official train/test splits. We use ground-truth shadow masks for ADSP and ISTD+, while adopting the masks provided by Cun et al. (2019) for SRD. For deployment-like evaluation, we further use masks predicted by FDRNet (Zhu et al., 2021) and BDRAR (Zhu et al., 2018). Following prior work, we report MAE, PSNR, and SSIM for shadow, non-shadow, and full-image regions using the MATLAB evaluation scripts provided by Zhu et al. (2022b). We additionally evaluate perceptual and distribution-level quality using LPIPS (Zhang et al., 2018), MUSIQ (Ke et al., 2021), MANIQA (Yang et al., 2022), CLIPIQA (Wang et al., 2022), FID (Heusel et al., 2018), KID (Bińkowski et al., 2021), and CLIP-FID (Kynkäänniemi et al., 2023).

We compare with six representative SOTA methods: SG-ShadowNet (Wan et al., 2022a), BMNet (Zhu et al., 2022a), ShadowFormer (Guo et al., 2023a), HomoFormer (Xiao et al., 2024), and RASM (Liu et al., 2024a), which use shadow masks as auxiliary inputs, and PhaSR (Lee et al., 2026), which uses surface normals and depth maps. Reported results or official implementations with recommended settings are used for evaluation, and mask-based models in the predicted-mask setting are retrained using their official configurations. Due to space limitations, the main paper focuses on ADSP results with MAE, PSNR, and SSIM under the FDRNet-based deployment setting, while results on SRD and ISTD+, additional evaluation metrics, and BDRAR-based experiments are provided in the supplementary material.

Table 1: **Quantitative comparison of shadow-region restoration performance under ideal auxiliary information.** The Stage-1 model uses the ground-truth shadow mask to explicitly specify the restoration region and is compared with existing SOTA methods using shadow-region metrics only. The best, second-best, and third-best results are highlighted with three different background shades, respectively.

Model	SG-ShadowNet	BMNet	ShadowFormer	HomoFormer	RASM	PhaSR	Ours (Stage 1)
Aux. input			GT mask			Normal/Depth	GT mask
MAE@s ↓	10.793	7.161	6.989	5.815	5.788	8.506	5.128
PSNR@s ↑	35.735	41.514	41.915	42.423	43.889	41.150	44.347
SSIM@s ↑	0.9937	0.9972	0.9973	0.9978	0.9979	0.9967	0.9973

Table 2: **Quantitative comparison between the proposed method and existing SOTA models under realistic deployment settings.** After Full-one mask tuning, the Stage-2 model θ_{MF} becomes an auxiliary-free shadow removal model. Combined with Post Fusion, the proposed method substantially improves non-shadow and full-image fidelity over I2F while maintaining strong shadow-region restoration. M_{GT} and M_{F} denote ground-truth masks and FDRNet-predicted masks, respectively.

Model	Train on.	Eval on.	MAE ↓			PSNR ↑			SSIM ↑		
			s	ns	all	s	ns	all	s	ns	all
SG-ShadowNet	M_{GT}	M_F	17.160	6.753	7.122	32.692	21.841	20.821	0.9891	0.9258	0.9115
BMNet	M_{GT}	M_F	15.810	7.823	8.034	32.976	20.406	19.705	0.9905	0.9168	0.9045
ShadowFormer	M_{GT}	M_F	17.067	10.471	10.606	33.150	19.851	19.355	0.9911	0.9199	0.9070
HomoFormer	M_{GT}	M_F	15.166	11.181	11.239	34.095	19.277	18.921	0.9920	0.9189	0.9080
RASM	M_{GT}	M_F	14.031	10.334	10.388	34.696	19.372	18.967	0.9928	0.9177	0.9093
SG-ShadowNet	M_F	M_F	8.827	2.133	2.408	37.814	35.108	32.323	0.9953	0.9846	0.9787
BMNet	M_F	M_F	11.562	2.615	2.960	35.714	34.960	31.346	0.9935	0.9840	0.9755
ShadowFormer	M_F	M_F	13.059	2.987	3.376	35.030	32.399	29.461	0.9928	0.9797	0.9706
HomoFormer	M_F	M_F	9.245	2.392	2.663	37.353	35.473	32.388	0.9951	0.9858	0.9795
RASM	M_F	M_F	13.256	2.864	3.272	35.055	34.179	30.459	0.9928	0.9840	0.9751
Ours (I2F)	-	-	6.321	3.969	4.066	41.178	32.383	31.510	0.9967	0.9513	0.9465
Ours (Fused)	-	-	6.326	3.310	3.429	40.961	34.236	32.690	0.9966	0.9846	0.9800
Ours (Oracle)	-	-	**6.118**	**0.047**	**0.318**	**41.306**	**50.942**	**39.832**	**0.9967**	**0.9989**	**0.9948**

4.2 Implementation Details

We freeze the pretrained Stable Diffusion backbone and update only the ControlNet and LoRA parameters, with the LoRA rank set to 8. During training, images are randomly cropped to 256×256 patches and optimized with AdamW using a batch size of 4. We adopt a v-prediction-based masked diffusion denoising objective. Training follows the 1,000-step DDPM (Ho et al., 2020) noise scheduler, while inference is performed with a 40-step DDIM (Song et al., 2022) sampler. For ADSP, Stage 1 and Stage 2 are trained for 1,122K and 1,973K iterations, respectively, using a fixed learning rate of 1e-5. In both stages, a Charbonnier loss (Charbonnier et al., 1994) with a weight of 0.1 is combined with the diffusion denoising objective, while an LPIPS loss with a weight of 1.0 is additionally introduced in Stage 2. Further implementation details, together with the complete training configurations for SRD and ISTD+, are provided in the supplementary material.

4.3 Comparison with State-of-the-art methods

In this subsection, we compare the proposed method with state-of-the-art shadow removal models from both quantitative and qualitative perspectives. Our comparison is designed to demonstrate two points. First, we highlight the importance of eliminating dependence on auxiliary information and developing shadow removal models that can operate independently at inference time. Second, we show that the proposed model and Post Fusion provide advantages under realistic deployment settings where shadow removal must be performed without access to ideal auxiliary inputs.

Table 1 compares the shadow-region restoration performance of I2F after GT mask tuning with existing SOTA methods on ADSP. With accurate shadow localization provided by the ground-truth mask, I2F achieves competitive or superior performance across the evaluated shadow-region metrics, demonstrating that the inpainting-based backbone can effectively adapt to the shadow removal task. More importantly, these results validate the role of Stage 1 in our two-stage training strategy: before removing the dependence on auxiliary information, the model first acquires strong shadow restoration capability under explicit spatial guidance, providing a favorable initialization for the subsequent auxiliary-free tuning stage.

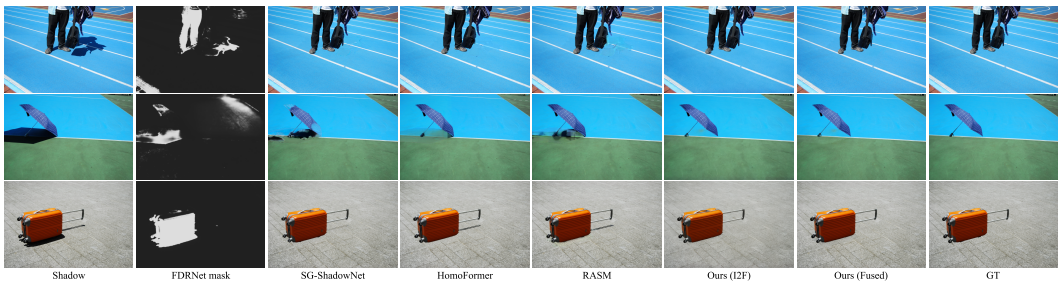

Figure 5: **Qualitative results on ADSP.** For SG-ShadowNet, HomoFormer, and RASM, we use results from models retrained from scratch with FDRNet masks to represent performance under realistic deployment settings.

Table 3: **Properties of the estimated pseudo mask for ADSP.**

Statistic\Metrics	IoU	GT coverage	Area ratio
Mean	0.867012	0.987465	1.472642
P10	0.731276	0.963631	1.011770
P25	0.859407	0.987757	1.035285
P50	0.908900	0.996538	1.076681
P75	0.949127	0.998734	1.146834
P90	0.973390	0.999652	1.312271

We compare the proposed framework with existing SOTA methods, as shown in table 2. Existing methods trained with ideal auxiliary information suffer substantial performance degradation when GT shadow masks are replaced by FDRNet-predicted masks at test time, mainly due to the distribution mismatch between training and inference. Moreover, this gap cannot be fully eliminated by retraining with predicted masks. In contrast, I2F achieves the best restoration performance in shadow regions without external auxiliary information. Meanwhile, the degradation of non-shadow details caused by diffusion-based global restoration can be substantially mitigated by Post Fusion. Since deployable Post Fusion directly preserves non-shadow pixels from the input shadow image, its non-shadow and full-image metrics remain affected by acquisition discrepancies between paired shadow and shadow-free images. To isolate this factor, we additionally report an Oracle setting, where the union of the predicted pseudo mask and the GT shadow mask is used to fuse the removal result with the GT shadow-free image. This prevents GT shadow-region pixels from being directly copied into the fused result while retaining the effect of pseudo mask expansion, since non-shadow pixels covered by the pseudo mask are not replaced by their GT shadow-free counterparts. Therefore, the Oracle setting serves as an idealized diagnostic reference that removes acquisition discrepancies in preserved non-shadow regions without eliminating errors from the estimated pseudo mask.

Figure 5 presents qualitative comparisons among three SOTA methods, I2F, and the fused results. The masks predicted by FDRNet may suffer from errors such as misclassification of dark objects, spatially inconsistent responses, and incomplete shadow coverage, leading to residual shadows or visible artifacts in auxiliary-dependent SOTA methods. In contrast, I2F produces restoration results that are more consistent with the surrounding background and exhibit less noticeable region boundaries. However, due to VAE reconstruction errors inherent in the diffusion backbone, some high-frequency details in non-shadow regions are degraded. Post Fusion substantially alleviates this issue by restoring the original non-shadow content while preserving the shadow-removal results.

4.4 Pseudo-Mask Analysis

We analyze the estimated pseudo masks to evaluate whether they provide reliable shadow localization for Post Fusion. Table 3 reports IoU, GT coverage, and area ratio. Since Post Fusion is more sensitive to false negatives than false positives, GT coverage is our primary concern: shadow pixels corrected by I2F may be copied back from the input image during fusion, whereas moderate over-coverage only causes additional non-shadow regions to retain the removal result. Area ratio is therefore used together with IoU to examine whether high coverage is achieved through excessive mask expansion. The proposed pseudo masks achieve a mean IoU of 0.867 and a mean GT coverage of 0.987, with the 10th-percentile coverage still reaching 0.963. Meanwhile, the mean area ratio remains approximately 1.47, indicating that high shadow coverage is achieved with a moderate degree of mask expansion. As shown in fig. 6, the estimated pseudo masks are also generally more spatially coherent and provide broader coverage of the true shadow regions than masks predicted by

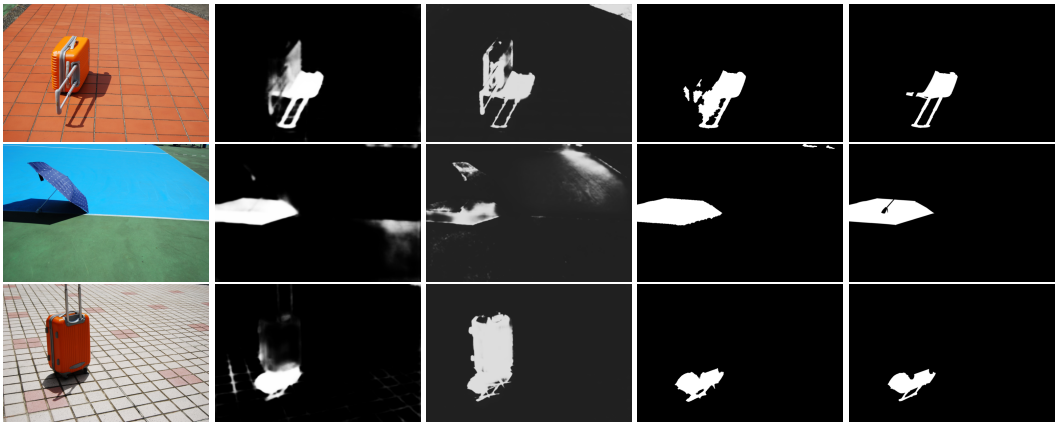

Figure 6: **Qualitative comparison of pseudo masks.** From left to right: shadow input, predictions from BDRAR and FDRNet, the pseudo masks estimated by our method, and the corresponding ground-truth (GT) masks. Our pseudo masks are generated from the shadow removal results and input images for subsequent Post Fusion, rather than by a dedicated shadow detection model.

Table 4: **Quantitative ablation study of GT mask tuning and Post Fusion on the ADSP dataset.** The GT mask tuning improves shadow-region restoration, while Post Fusion substantially enhances non-shadow and full-image fidelity.

	GT mask tuning	Post Fusion	MAE			PSNR			SSIM		
			s	ns	all	s	ns	all	s	ns	all
(1)			8.132	4.761	4.916	39.50	31.75	30.63	0.9956	0.9460	0.9398
(2)		✓	8.083	3.311	3.522	39.33	34.03	32.03	0.9955	0.9845	0.9787
(3)	✓		6.321	3.969	4.066	41.18	32.38	31.51	0.9967	0.9513	0.9465
(4)	✓	✓	6.326	3.310	3.429	40.96	34.24	32.69	0.9966	0.9846	0.9800

external shadow detectors. Together, these quantitative and qualitative results support the suitability of the estimated masks for Post Fusion.

4.5 Ablation Study

We conduct ablation studies to analyze the contributions of the proposed framework. We first examine the effects of GT mask tuning and Post Fusion on the final shadow removal performance. We then focus on the pseudo mask generation process and compare the proposed GMM-based thresholding strategy with Otsu (Otsu, 1979), MET (Kittler & Illingworth, 1986), and GHT (Barron, 2020).

Table 4 analyzes the effects of GT mask tuning and Post Fusion on the final performance. The complete model in setting (4) achieves the best overall performance in non-shadow regions and for the full image, while maintaining shadow-region restoration quality close to that of setting (3) without Post Fusion. This suggests that the two components provide a favorable balance between shadow removal and non-shadow detail preservation. Comparing (1) with (3), as well as (2) with (4), shows that GT mask tuning consistently improves shadow-region performance, supporting the design motivation in section 3: adapting the model to the shadow removal objective under accurate spatial guidance provides a better initialization before transitioning to the auxiliary-free setting. Meanwhile, comparisons between (1) and (2), and between (3) and (4), show that Post Fusion substantially improves non-shadow and full-image performance, with only minor degradation in shadow regions. This indicates that the recall-oriented pseudo-mask design effectively limits the risk of residual shadows introduced by fusion while recovering non-shadow details from the original input.

Figure 7 provides qualitative examples corresponding to table 4. Although the visual differences among the four settings are relatively subtle, configurations without GT mask tuning tend to retain slightly more visible shadow traces, while Post Fusion better preserves fine details in non-shadow regions. These observations are consistent with the quantitative trends in table 4.

Table 5 compares different thresholding strategies for pseudo-mask estimation. Since shadow responses in the difference map can vary from distinct to weak, our goal is not to maximize generic segmentation accuracy but to maintain reliable shadow coverage across different response distributions. Among the non-degenerate alternatives, our GMM-based strategy achieves the highest mean

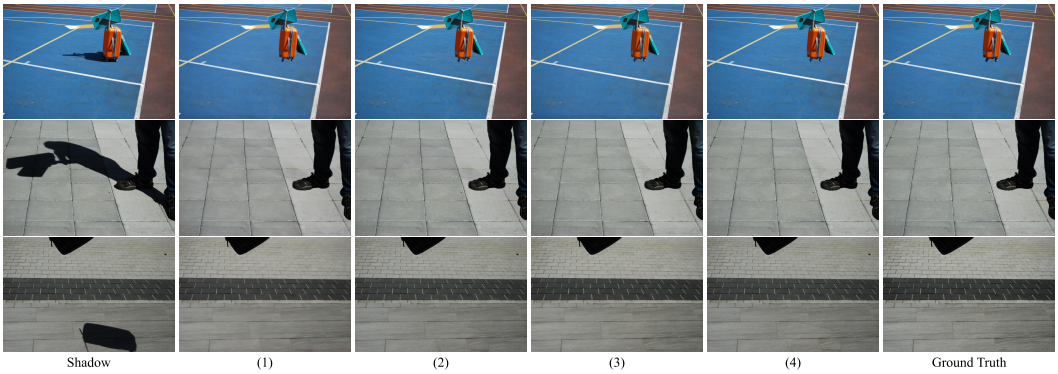

Figure 7: **Qualitative ablation study of GT mask tuning and Post Fusion on the ADSP dataset.**

Table 5: **Quantitative ablation study of GMM-based thresholding on ADSP.**

Algorithm\Metrics	IoU				GT Coverage				Area Ratio			
	Mean	P10	P50	P90	Mean	P10	P50	P90	Mean	P10	P50	P90
Otsu	0.903	0.847	0.944	0.982	0.969	0.932	0.986	0.997	1.802	0.964	1.009	1.056
MET	0.912	0.835	0.936	0.975	0.974	0.943	0.992	1.000	1.046	0.993	1.035	1.129
GHT	0.051	0.007	0.026	0.116	1.000	1.000	1.000	1.000	62.911	8.613	37.778	144.126
Ours (GMM)	0.867	0.731	0.909	0.973	0.987	0.964	0.997	0.999	1.473	1.012	1.077	1.312

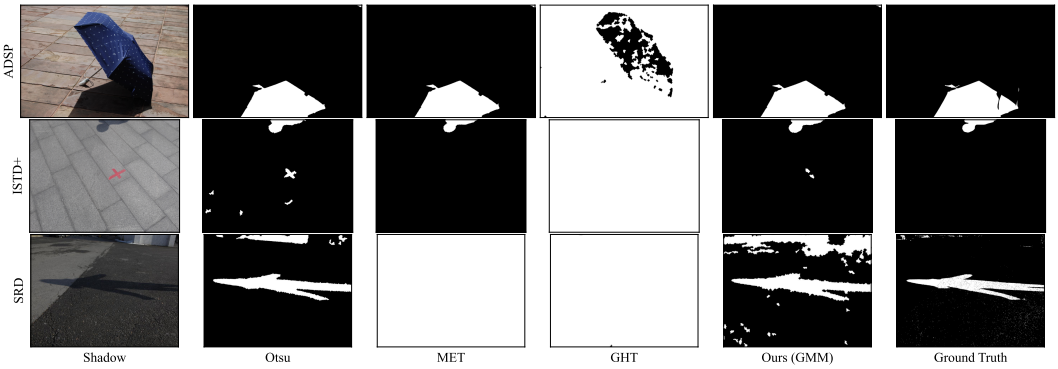

Figure 8: **Qualitative ablation study of thresholding on ADSP, ISTD+, and SRD.**

and P10 GT coverage while avoiding the extreme mask over-expansion exhibited by GHT. Although Otsu and MET obtain higher IoU, the proposed method accepts this moderate reduction in exchange for fewer missed shadow regions, which better matches the false-negative-sensitive objective of Post Fusion. Figure 8 further shows that the proposed strategy maintains stable shadow coverage across cases with different response strengths and dark-region interference. Together, these results indicate that the GMM-based thresholding is better aligned with the adaptive, recall-oriented requirements of Post Fusion rather than serving as a general-purpose replacement for conventional thresholding methods.

5 Conclusion

We revisit shadow removal from the perspective of reducing dependence on external auxiliary information. Inpaint2Free (I2F) fine-tunes an inpainting model in two stages, first with GT masks and then with Full-one masks, retaining the benefits of spatially guided training while enabling auxiliary-free inference. Post Fusion further estimates a recall-oriented pseudo shadow mask from the difference between the shadow input and the removal result to preserve details in non-shadow regions. Experiments on multiple benchmarks validate both components: I2F is competitive with existing SOTA methods without requiring auxiliary inputs at inference and shows clear advantages when compared with auxiliary-dependent methods operating on imperfect predicted auxiliary information, offering a practical step toward bridging laboratory evaluation and real-world deployment.

Limitations A limitation of our method is computational cost due to the use of a diffusion-based backbone. Distillation to a few-step or one-step model remains an important direction for future work.

AI USE STATEMENT

In this work, we used generative AI tools to design or provide feedback on research methodology or experiments, implement methods, and assist with translation. We did not use generative AI tools to clean and reformat datasets, support qualitative and thematic data analysis, or interpret results. The generation of synthetic datasets, development of theoretical models or conceptual frameworks, formulation of mathematical claims, provision of critical ingredients for proving mathematical claims, assistance in the writing of proofs, and proposal or refinement of hypotheses were not applicable to this work.

Additionally, we used generative AI tools to suggest experimental parameters, create or edit software code, draft parts of a research paper, summarize or analyze existing literature, brainstorm, source or search for information, edit the research paper to improve readability, identify relevant literature, and suggest a structure for the research paper. Generative AI tools were also used to assist in implementing visualization and figure-generation code based on figure designs specified by the authors. The scientific figures themselves were not directly generated from AI prompts. For discussions of experimental results, the authors performed the underlying interpretation, while generative AI tools were used only to refine the wording and presentation of these interpretations.

We reviewed all AI-assisted work. AI-assisted revisions of technical writing were manually checked to ensure that the intended technical meaning, experimental settings, and reported results were preserved. For literature-related assistance, technical claims and descriptions were verified against the original papers before being incorporated into the manuscript. AI-assisted suggestions concerning methodology and experimental design were independently assessed by the authors based on experimental evidence and domain knowledge, and only those deemed technically appropriate were adopted. AI-assisted code and implementation suggestions, including code used for visualization and figure generation, were reviewed, executed, and tested by the authors before being used in the work. We take responsibility for the final content of this work, including all text, claims, code, analyses, figures, and other artifacts produced with the aid of generative AI.

REPRODUCIBILITY STATEMENT

To facilitate reproducibility, we provide the experimental protocols and implementation details throughout the main paper and supplementary material. Section 4.1 describes the datasets, evaluation metrics, and comparison methods used in our experiments, while section 4.2 reports the main training configuration on ADSP. Due to space limitations, more comprehensive implementation details are provided in the supplementary material. Specifically, Sec. 2.1 documents the configurations shared across datasets, the dataset-specific settings, and additional details such as data preprocessing and numerical precision. Sec. 2.2 provides the pseudocode and mathematical formulations for the proposed Post Fusion procedure and the adaptive threshold selection for pseudo-mask estimation. We also commit to publicly releasing the complete training, inference, and evaluation code, together with the corresponding configuration files, trained checkpoints, and representative inference samples, to further facilitate reproduction of the reported results.

REFERENCES

Jonathan T. Barron. A generalization of otsu's method and minimum error thresholding, 2020. URL https://arxiv.org/abs/2007.07350.

Mikołaj Bińkowski, Danica J. Sutherland, Michael Arbel, and Arthur Gretton. Demystifying mmd gans, 2021. URL https://arxiv.org/abs/1801.01401.

Hua-En Chang, Chia-Hsuan Hsieh, Hao-Hsiang Yang, I-Hsiang Chen, Yi-Chung Chen, Yuan-Chun Chiang, Zhi-Kai Huang, Wei-Ting Chen, and Sy-Yen Kuo. Tsrformer: Transformer based two-stage refinement for single image shadow removal. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, pp. 1436–1446, June 2023.

P. Charbonnier, L. Blanc-Feraud, G. Aubert, and M. Barlaud. Two deterministic half-quadratic regularization algorithms for computed imaging. In *Proceedings of 1st International Conference on Image Processing*, volume 2, pp. 168–172 vol.2, 1994. doi: 10.1109/ICIP.1994.413553.

Kerui Chen, Zhiliang Wu, Wenjin Hou, Kun Li, Hehe Fan, and Yi Yang. Prompt-aware controllable shadow removal, 2025. URL https://arxiv.org/abs/2501.15043.

Zipei Chen, Chengjiang Long, Ling Zhang, and Chunxia Xiao. Canet: A context-aware network for shadow removal, 2021. URL https://arxiv.org/abs/2108.09894.

R. Cucchiara, C. Grana, M. Piccardi, A. Prati, and S. Sirotti. Improving shadow suppression in moving object detection with hsv color information. In *ITSC 2001. 2001 IEEE Intelligent Transportation Systems. Proceedings (Cat. No.01TH8585)*, pp. 334–339, 2001. doi: 10.1109/ITSC.2001.948679.

R. Cucchiara, C. Grana, M. Piccardi, and A. Prati. Detecting moving objects, ghosts, and shadows in video streams. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 25(10):1337–1342, 2003. doi: 10.1109/TPAMI.2003.1233909.

Shuhao Cui, Junshi Huang, Shuman Tian, Mingyuan Fan, Jiaqi Zhang, Li Zhu, Xiaoming Wei, and Xiaolin Wei. Pyramid ensemble structure for high resolution image shadow removal. In *2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pp. 1311–1319, 2023. doi: 10.1109/CVPRW59228.2023.00138.

Xiaodong Cun, Chi-Man Pun, and Cheng Shi. Towards ghost-free shadow removal via dual hierarchical aggregation network and shadow matting gan, 2019. URL https://arxiv.org/abs/1911.08718.

Martin Danelljan, Fahad Shahbaz Khan, Michael Felsberg, and Joost van de Weijer. Adaptive color attributes for real-time visual tracking. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2014.

Subhrajyoti Dasgupta, Arindam Das, Senthil Yogamani, Sudip Das, Ciarán Eising, Andrei Bursuc, and Ujjwal Bhattacharya. Unshadownet: Illumination critic guided contrastive learning for shadow removal. *IEEE Access*, 11:87760–87774, 2023. ISSN 2169-3536. doi: 10.1109/access.2023.3305576. URL http://dx.doi.org/10.1109/ACCESS.2023.3305576.

Wei Dong, Han Zhou, Yuqiong Tian, Jingke Sun, Xiaohong Liu, Guangtao Zhai, and Jun Chen. Shadowrefiner: Towards mask-free shadow removal via fast fourier transformer, 2024. URL https://arxiv.org/abs/2406.02559.

Wei Dong, Han Zhou, Seyed Amirreza Mousavi, and Jun Chen. Retinex-guided histogram transformer for mask-free shadow removal, 2025. URL https://arxiv.org/abs/2504.14092.

G.D. Finlayson, S.D. Hordley, Cheng Lu, and M.S. Drew. On the removal of shadows from images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 28(1):59–68, 2006. doi: 10.1109/TPAMI.2006.18.

Graham D. Finlayson, Steven D. Hordley, and Mark S. Drew. Removing shadows from images. In Anders Heyden, Gunnar Sparr, Mads Nielsen, and Peter Johansen (eds.), *Computer Vision — ECCV 2002*, pp. 823–836, Berlin, Heidelberg, 2002. Springer Berlin Heidelberg. ISBN 978-3-540-47979-6.

Graham D Finlayson, Mark S Drew, and Cheng Lu. Entropy minimization for shadow removal. *International Journal of Computer Vision*, 85(1):35–57, 2009.

Gang Fu, Qing Zhang, and Chunxia Xiao. Towards high-quality intrinsic images in the wild. In *2019 IEEE International Conference on Multimedia and Expo (ICME)*, pp. 175–180, 2019. doi: 10.1109/ICME.2019.00038.

Lan Fu, Changqing Zhou, Qing Guo, Felix Juefei-Xu, Hongkai Yu, Wei Feng, Yang Liu, and Song Wang. Auto-exposure fusion for single-image shadow removal, 2021. URL https://arxiv.org/abs/2103.01255.

Jianhao Gao, Quanlong Zheng, and Yandong Guo. Towards real-world shadow removal with a shadow simulation method and a two-stage framework. In *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pp. 598–607, 2022. doi: 10.1109/CVPRW56347.2022.00075.

Han Gong and Darren Cosker. Interactive removal and ground truth for difficult shadow scenes. *Journal of the Optical Society of America A*, 33(9):1798, August 2016. ISSN 1520-8532. doi: 10.1364/josaa.33.001798. URL http://dx.doi.org/10.1364/JOSAA.33.001798.

Lanqing Guo, Siyu Huang, Ding Liu, Hao Cheng, and Bihan Wen. Shadowformer: Global context helps image shadow removal. *arXiv preprint arXiv:2302.01650*, 2023a.

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 *2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 14049–14058, 2023b. doi: 10.1109/CVPR52729.2023.01350.

Lanqing Guo, Chong Wang, Wenhan Yang, Yufei Wang, and Bihan Wen. Boundary-aware divide and conquer: A diffusion-based solution for unsupervised shadow removal. In *2023 IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 12999–13008, 2023c. doi: 10.1109/ICCV51070.2023.01199.

Ruiqi Guo, Qieyun Dai, and Derek Hoiem. Paired regions for shadow detection and removal. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 35(12):2956–2967, 2013. doi: 10.1109/TPAMI.2012.214.

Martin Heusel, Hubert Ramsauer, Thomas Unterthiner, Bernhard Nessler, and Sepp Hochreiter. Gans trained by a two time-scale update rule converge to a local nash equilibrium, 2018. URL https://arxiv.org/abs/1706.08500.

Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising diffusion probabilistic models, 2020. URL https://arxiv.org/abs/2006.11239.

Chang-Yu Hsieh and Jian-Jiun Ding. Adsp: Advanced dataset for shadow processing, enabling visible occluders via synthesizing strategy. In Minsu Cho, Ivan Laptev, Du Tran, Angela Yao, and Hongbin Zha (eds.), *Computer Vision – ACCV 2024*, pp. 329–347, Singapore, 2025. Springer Nature Singapore. ISBN 978-981-96-0917-8.

Edward J. Hu, Yelong Shen, Phillip Wallis, Zeyuan Allen-Zhu, Yuanzhi Li, Shean Wang, Lu Wang, and Weizhu Chen. Lora: Low-rank adaptation of large language models, 2021. URL https://arxiv.org/abs/2106.09685.

Shilin Hu, Hieu Le, ShahRukh Athar, Sagnik Das, and Dimitris Samaras. Shadow removal refinement via material-consistent shadow edges, 2024. URL https://arxiv.org/abs/2409.06848.

Xiaowei Hu, Yitong Jiang, Chi-Wing Fu, and Pheng-Ann Heng. Mask-shadowgan: Learning to remove shadows from unpaired data. In *2019 IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 2472–2481. IEEE, October 2019. doi: 10.1109/iccv.2019.00256. URL http://dx.doi.org/10.1109/ICCV.2019.00256.

Hao Jiang, Qing Zhang, Yongwei Nie, Lei Zhu, and Wei-Shi Zheng. Learning to remove shadows from a single image. *Int. J. Comput. Vision*, 131(9):2471–2488, June 2023. ISSN 0920-5691. doi: 10.1007/s11263-023-01823-9. URL https://doi.org/10.1007/s11263-023-01823-9.

Yeying Jin, Aashish Sharma, and Robby T. Tan. Dc-shadownet: Single-image hard and soft shadow removal using unsupervised domain-classifier guided network, 2023. URL https://arxiv.org/abs/2207.10434.

Yeying Jin, Wei Ye, Wenhan Yang, Yuan Yuan, and Robby T. Tan. Des3: Adaptive attention-driven self and soft shadow removal using vit similarity, 2024. URL https://arxiv.org/abs/2211.08089.

P. KaewTraKulPong and R. Bowden. *An Improved Adaptive Background Mixture Model for Real-time Tracking with Shadow Detection*, pp. 135–144. Springer US, Boston, MA, 2002. ISBN 978-1-4615-0913-4. doi: 10.1007/978-1-4615-0913-4_11. URL https://doi.org/10.1007/978-1-4615-0913-4_11.

Junjie Ke, Qifei Wang, Yilin Wang, Peyman Milanfar, and Feng Yang. Musiq: Multi-scale image quality transformer, 2021. URL https://arxiv.org/abs/2108.05997.

J. Kittler and J. Illingworth. Minimum error thresholding. *Pattern Recognition*, 19(1):41–47, 1986. ISSN 0031-3203. doi: https://doi.org/10.1016/0031-3203(86)90030-0. URL https://www.sciencedirect.com/science/article/pii/0031320386900300.

Nikolina Kubiak, Armin Mustafa, Graeme Phillipson, Stephen Jolly, and Simon Hadfield. S3r-net: A single-stage approach to self-supervised shadow removal. In *2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pp. 5898–5908, 2024. doi: 10.1109/CVPRW63382.2024.00597.

Tuomas Kynkäänniemi, Tero Karras, Miika Aittala, Timo Aila, and Jaakko Lehtinen. The role of imagenet classes in fréchet inception distance, 2023. URL https://arxiv.org/abs/2203.06026.

Jean-François Lalonde, Alexei A. Efros, and Srinivasa G. Narasimhan. Detecting ground shadows in outdoor consumer photographs. In Kostas Daniilidis, Petros Maragos, and Nikos Paragios (eds.), *Computer Vision – ECCV 2010*, pp. 322–335, Berlin, Heidelberg, 2010. Springer Berlin Heidelberg. ISBN 978-3-642-15552-9.

Hieu Le and Dimitris Samaras. Shadow removal via shadow image decomposition. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, October 2019.

Hieu Le and Dimitris Samaras. From shadow segmentation to shadow removal, 2020. URL https://arxiv.org/abs/2008.00267.

Hieu Le and Dimitris Samaras. Physics-based shadow image decomposition for shadow removal. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 44(12):9088–9101, December 2022. ISSN 1939-3539. doi: 10.1109/tpami.2021.3124934. URL http://dx.doi.org/10.1109/TPAMI.2021.3124934.

Hieu Le, Bento Gonçalves, Dimitris Samaras, and Heather Lynch. Weakly labeling the antarctic: The penguin colony case. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, June 2019.

Chia-Ming Lee, Yu-Fan Lin, Yu-Jou Hsiao, Jing-Hui Jung, Yu-Lun Liu, and Chih-Chung Hsu. Phasr: Generalized image shadow removal with physically aligned priors, 2026. URL https://arxiv.org/abs/2601.17470.

Xiaoguang Li, Qing Guo, Rabab Abdelfattah, Di Lin, Wei Feng, Ivor Tsang, and Song Wang. Leveraging inpainting for single-image shadow removal, 2023. URL https://arxiv.org/abs/2302.05361.

Zhengqi Li and Noah Snavely. Learning intrinsic image decomposition from watching the world. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2018.

Zhuohao Li, Guoyang Xie, Guannan Jiang, and Zhichao Lu. Shadowmaskformer: Mask augmented patch embedding for shadow removal. *IEEE Transactions on Artificial Intelligence*, 6(10):2618–2629, 2025. doi: 10.1109/TAI.2025.3554605.

Tao Lin, Qingwang Wang, Qiwei Liang, Minghua Tang, and Yuxuan Sun. Fasr-net: Unsupervised shadow removal leveraging inherent frequency priors, 2025a. URL https://arxiv.org/abs/2504.05779.

Yu-Fan Lin, Chia-Ming Lee, and Chih-Chung Hsu. Densesr: Image shadow removal as dense prediction, 2025b. URL https://arxiv.org/abs/2507.16472.

Feng Liu and Michael Gleicher. Texture-consistent shadow removal. In *Proceedings of the 10th European Conference on Computer Vision: Part IV*, ECCV '08, pp. 437–450, Berlin, Heidelberg, 2008. Springer-Verlag. ISBN 9783540886921. doi: 10.1007/978-3-540-88693-8_32. URL https://doi.org/10.1007/978-3-540-88693-8_32.

Hengxing Liu, Mingjia Li, and Xiaojie Guo. Regional attention for shadow removal. In *Proceedings of the 32nd ACM International Conference on Multimedia*, MM '24, pp. 5949–5957, New York, NY, USA, 2024a. Association for Computing Machinery. ISBN 9798400706868. doi: 10.1145/3664647.3681126. URL https://doi.org/10.1145/3664647.3681126.

Yuhao Liu, Qing Guo, Lan Fu, Zhanghan Ke, Ke Xu, Wei Feng, Ivor W. Tsang, and Rynson W. H. Lau. Structure-informed shadow removal networks. *IEEE Transactions on Image Processing*, 32: 5823–5836, 2023. doi: 10.1109/TIP.2023.3323814.

Yuhao Liu, Zhanghan Ke, Ke Xu, Fang Liu, Zhenwei Wang, and Rynson W. H. Lau. Recasting regional lighting for shadow removal, 2024b. URL https://arxiv.org/abs/2402.00341.

Zhihao Liu, Hui Yin, Yang Mi, Mengyang Pu, and Song Wang. Shadow removal by a lightness-guided network with training on unpaired data. *IEEE Transactions on Image Processing*, 30: 1853–1865, 2021. ISSN 1941-0042. doi: 10.1109/tip.2020.3048677. URL http://dx.doi.org/10.1109/TIP.2020.3048677.

Xin Lu, Yurui Zhu, Xi Wang, Dong Li, Jie Xiao, Yunpeng Zhang, Xueyang Fu, and Zheng-Jun Zha. Hirformer: Dynamic high resolution transformer for large-scale image shadow removal. In *2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pp. 6513–6523, 2024. doi: 10.1109/CVPRW63382.2024.00651.

Jinting Luo, Ru Li, Chengzhi Jiang, Xiaoming Zhang, Mingyan Han, Ting Jiang, Haoqiang Fan, and Shuaicheng Liu. Diff-shadow: Global-guided diffusion model for shadow removal, 2024. URL https://arxiv.org/abs/2407.16214.

Liqian Ma, Jue Wang, Eli Shechtman, Kalyan Sunkavalli, and Shimin Hu. Appearance harmonization for single image shadow removal, 2016. URL https://arxiv.org/abs/1603.06398.

Kangfu Mei, Luis Figueroa, Zhe Lin, Zhihong Ding, Scott Cohen, and Vishal M. Patel. Latent feature-guided diffusion models for shadow removal. In *2024 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, pp. 4301–4310, 2024. doi: 10.1109/WACV57701.2024.00426.

Nobuyuki Otsu. A threshold selection method from gray-level histograms. *IEEE transactions on systems, man, and cybernetics*, 9(1):62–66, 1979.

Liangqiong Qu, Jiandong Tian, Shengfeng He, Yandong Tang, and Rynson W. H. Lau. Deshadownet: A multi-context embedding deep network for shadow removal. In *2017 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017, Honolulu, HI, USA, July 21-26, 2017*, pp. 2308–2316. IEEE Computer Society, 2017a. doi: 10.1109/CVPR.2017.248. URL https://doi.org/10.1109/CVPR.2017.248.

Liangqiong Qu, Jiandong Tian, Shengfeng He, Yandong Tang, and Rynson W. H. Lau. Deshadownet: A multi-context embedding deep network for shadow removal. In *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 2308–2316, 2017b. doi: 10.1109/CVPR.2017.248.

Robin Rombach, Andreas Blattmann, Dominik Lorenz, Patrick Esser, and Björn Ommer. High-resolution image synthesis with latent diffusion models. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 10684–10695, June 2022.

Andres Sanin, Conrad Sanderson, and Brian C. Lovell. Improved shadow removal for robust person tracking in surveillance scenarios. In *2010 20th International Conference on Pattern Recognition*, pp. 141–144, 2010. doi: 10.1109/ICPR.2010.43.

S. Saravanakumar, A. Vadivel, and C.G Saneem Ahmed. Multiple human object tracking using background subtraction and shadow removal techniques. In *2010 International Conference on Signal and Image Processing*, pp. 79–84, 2010. doi: 10.1109/ICSIP.2010.5697446.

Gideon Schwarz. Estimating the dimension of a model. *The Annals of Statistics*, 6(2):461–464, 1978. ISSN 00905364, 21688966. URL http://www.jstor.org/stable/2958889.

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 *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, pp. 1809–1817, January 2023.

Joongchol Shin, Hasil Park, and Joonki Paik. Region-based dehazing via dual-supervised triple-convolutional network. *IEEE Transactions on Multimedia*, 24:245–260, 2022. doi: 10.1109/TMM.2021.3050053.

Jiaming Song, Chenlin Meng, and Stefano Ermon. Denoising diffusion implicit models, 2022. URL https://arxiv.org/abs/2010.02502.

J. Stander, R. Mech, and J. Ostermann. Detection of moving cast shadows for object segmentation. *IEEE Transactions on Multimedia*, 1(1):65–76, 1999. doi: 10.1109/6046.748172.

Nan Su, Ye Zhang, Shu Tian, Yiming Yan, and Xinyuan Miao. Shadow detection and removal for occluded object information recovery in urban high-resolution panchromatic satellite images. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 9(6):2568–2582, 2016. doi: 10.1109/JSTARS.2016.2570234.

Maryam Sultana, Arif Mahmood, and Soon Ki Jung. Unsupervised moving object detection in complex scenes using adversarial regularizations. *IEEE Transactions on Multimedia*, 23:2005–2018, 2021. doi: 10.1109/TMM.2020.3006419.

Florin-Alexandru Vasluianu, Andres Romero, Luc Van Gool, and Radu Timofte. Self-supervised shadow removal, 2020. URL https://arxiv.org/abs/2010.11619.

Florin-Alexandru Vasluianu, Andrés Romero, Luc Van Gool, and Radu Timofte. Shadow removal with paired and unpaired learning. In *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pp. 826–835, 2021. doi: 10.1109/CVPRW53098.2021.00092.

Florin-Alexandru Vasluianu, Tim Seizinger, and Radu Timofte. Wsrd: A novel benchmark for high resolution image shadow removal. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, pp. 1826–1835, June 2023a.

Florin-Alexandru Vasluianu, Tim Seizinger, Radu Timofte, Shuhao Cui, Junshi Huang, Shuman Tian, Mingyuan Fan, Jiaqi Zhang, Li Zhu, Xiaoming Wei, Xiaolin Wei, Ziwei Luo, Fredrik K. Gustafsson, Zheng Zhao, Jens Sjölund, Thomas B. Schön, Xiaoyi Dong, Xi Sheryl Zhang, Chenghua Li, Cong Leng, Woon-Ha Yeo, Wang-Taek Oh, Yeo-Reum Lee, Han-Cheol Ryu, Jinting Luo, Chengzhi Jiang, Mingyan Han, Qi Wu, Wenjie Lin, Lei Yu, Xinpeng Li, Ting Jiang, Haoqiang Fan, Shuaicheng Liu, Shuning Xu, Binbin Song, Xiangyu Chen, Shile Zhang, Jiantao Zhou, Zhao Zhang, Suiyi Zhao, Huan Zheng, Yangcheng Gao, Yanyan Wei, Bo Wang, Jiahuan Ren, Yan Luo, Yuki Kondo, Riku Miyata, Fuma Yasue, Taito Naruki, Norimichi Ukita, Hua-En Chang, Hao-Hsiang Yang, Yi-Chung Chen, Yuan-Chun Chiang, Zhi-Kai Huang, Wei-Ting Chen, I-Hsiang Chen, Chia-Hsuan Hsieh, Sy-Yen Kuo, Li Xianwei, Huiyuan Fu, Chunlin Liu, Huadong Ma, Binglan Fu, Huiming He, Mengjia Wang, Wenxuan She, Yu Liu, Sabari Nathan, Priya Kansal, Zhongjian Zhang, Huabin Yang, Yan Wang, Yanru Zhang, Shruti S. Phutke, Ashutosh Kulkarni, MD Raqib Khan, Subrahmanyam Murala, Santosh Kumar Vipparthi, Heng Ye, Zixi Liu, Xingyi Yang, Songhua Liu, Yinwei Wu, Yongcheng Jing, Qianhao Yu, Naishan Zheng, Jie Huang, Yuhang Long, Mingde Yao, Feng Zhao, Bowen Zhao, Nan Ye, Ning Shen, Yanpeng Cao, Tong Xiong, Weiran Xia, Dingwen Li, and Shuchen Xia. Ntire 2023 image shadow removal challenge report. In *2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pp. 1788–1807, 2023b. doi: 10.1109/CVPRW59228.2023.00179.

Florin-Alexandru Vasluianu, Tim Seizinger, Zhuyun Zhou, Zongwei Wu, Cailian Chen, Radu Timofte, Wei Dong, Han Zhou, Yuqiong Tian, Jun Chen, Xueyang Fu, Xin Lu, Yurui Zhu, Xi Wang, Dong Li, Jie Xiao, Yunpeng Zhang, Zheng-Jun Zha, Zhao Zhang, Suiyi Zhao, Bo Wang, Yan Luo, Yanyan Wei, Zhihao Zhao, Long Sun, Tingting Yang, Jinshan Pan, Jiangxin Dong, Jinhui Tang, Bilel Benjdira, Mohammed Nassif, Anis Koubaa, Ahmed Elhayek, Anas M. Ali, Kyotaro Tokoro, Kento Kawai, Kaname Yokoyama, Takuya Seno, Yuki Kondo, Norimichi Ukita, Chenghua Li, Bo Yang, Zhiqi Wu, Gao Chen, Yihan Yu, Sixiang Chen, Kai Zhang, Tian Ye, Wenbin Zou, Yunlong Lin, Zhaohu Xing, Jinbin Bai, Wenhao Chai, Lei Zhu, Ritik Maheshwari, Rakshank Verma, Rahul Tekchandani, Praful Hambarde, Satya Narayan Tazi, Santosh Kumar Vipparthi, Subrahmanyam Murala, Jaeho Lee, Seongwan Kim, Sharif S M A, Nodirkhuja Khujaev, Roman Tsoy, Fan Gao, Weidan Yan, Wenze Shao, Dengyin Zhang, Bin Chen, Siqi Zhang, Yanxin Qian, Yuanbin Chen, Yuanbo Zhou, Tong Tong, Rongfeng Wei, Ruiqi Sun, Yue Liu, Nikhil Akalwadi, Amogh Joshi, Sampada Malagi, Chaitra Desai, Ramesh Ashok Tabib, Uma Mudenagudi, Ali Murtaza, Uswah Khairuddin, Ahmad Athif Mohd Faudzi, Adinath Dukre, Vivek Deshmukh, Shruti S. Phutke, Ashutosh Kulkarni, Santosh Kumar Vipparthi, Anil Gonde, Subrahmanyam Murala, Arun Karthik K, Manasa N, Shri Hari Priya, Wei Hao, Xingzhuo Yan, and Minghan Fu. Ntire 2024 image shadow removal challenge report. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, pp. 6547–6570, June 2024.

Tomás F. Yago Vicente and Dimitris Samaras. Single image shadow removal via neighbor-based region relighting. In Lourdes Agapito, Michael M. Bronstein, and Carsten Rother (eds.), *Computer Vision - ECCV 2014 Workshops*, pp. 309–320, Cham, 2015. Springer International Publishing. ISBN 978-3-319-16199-0.

Jin Wan, Hui Yin, Zhenyao Wu, Xinyi Wu, Yanting Liu, and Song Wang. Style-guided shadow removal. In *ECCV*, 2022a.

Jin Wan, Hui Yin, Zhenyao Wu, Xinyi Wu, Zhihao Liu, and Song Wang. Crformer: A cross-region transformer for shadow removal, 2022b. URL https://arxiv.org/abs/2207.01600.

Jianyi Wang, Kelvin C. K. Chan, and Chen Change Loy. Exploring clip for assessing the look and feel of images, 2022. URL https://arxiv.org/abs/2207.12396.

Jifeng Wang, Xiang Li, and Jian Yang. Stacked conditional generative adversarial networks for jointly learning shadow detection and shadow removal. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2018.

Pan Wang, Yihao Hu, Xiujin Liu, and Hang Wang. Cfsr: Geometry-conditioned shadow removal via physical disentanglement. *arXiv preprint arXiv:2604.18032*, 2026a.

Xinrui Wang, Lanqing Guo, Xiyu Wang, Siyu Huang, and Bihan Wen. Softshadow: Leveraging soft masks for penumbra-aware shadow removal, 2025. URL https://arxiv.org/abs/2409.07041.

Yinan Wang, Yan Huang, Yong Xu, and Patrick Le Callet. Freeshadow: Training-free shadow removal via illumination transfer and selective content preservation in diffusion models. *arXiv preprint arXiv:2607.26715*, 2026b.

Yonghui Wang, Wengang Zhou, Hao Feng, Li Li, and Houqiang Li. Progressive recurrent network for shadow removal. *Computer Vision and Image Understanding*, 238:103861, 2024. ISSN 1077-3142. doi: https://doi.org/10.1016/j.cviu.2023.103861. URL https://www.sciencedirect.com/science/article/pii/S1077314223002412.

Jie Xiao, Xueyang Fu, Yurui Zhu, Dong Li, Jie Huang, Kai Zhu, and Zheng-Jun Zha. Homoformer: Homogenized transformer for image shadow removal. In *2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 25617–25626, 2024. doi: 10.1109/CVPR52733.2024.02420.

Jiamin Xu, Zelong Li, Yuxin Zheng, Chenyu Huang, Renshu Gu, Weiwei Xu, and Gang Xu. Omnisr: Shadow removal under direct and indirect lighting, 2025a. URL https://arxiv.org/abs/2410.01719.

Jiamin Xu, Yuxin Zheng, Zelong Li, Chi Wang, Renshu Gu, Weiwei Xu, and Gang Xu. Detail-preserving latent diffusion for stable shadow removal. In *2025 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 7592–7602, 2025b. doi: 10.1109/CVPR52734.2025.00711.

Yimin Xu, Mingbao Lin, Hong Yang, Fei Chao, and Rongrong Ji. Shadow-aware dynamic convolution for shadow removal. *Pattern Recognition*, 146:109969, 2024. ISSN 0031-3203. doi: https://doi.org/10.1016/j.patcog.2023.109969. URL https://www.sciencedirect.com/science/article/pii/S0031320323006672.

Qingxiong Yang, Kar-Han Tan, and Narendra Ahuja. Shadow removal using bilateral filtering. *IEEE Transactions on Image Processing*, 21(10):4361–4368, 2012. doi: 10.1109/TIP.2012.2208976.

Sidi Yang, Tianhe Wu, Shuwei Shi, Shanshan Lao, Yuan Gong, Mingdeng Cao, Jiahao Wang, and Yujiu Yang. Maniqa: Multi-dimension attention network for no-reference image quality assessment, 2022. URL https://arxiv.org/abs/2204.08958.

Ziyi Yang, Yanzhen Chen, Xinyu Gao, Yazhen Yuan, Yu Wu, Xiaowei Zhou, and Xiaogang Jin. Sire-ir: Inverse rendering for brdf reconstruction with shadow and illumination removal in high-illuminance scenes, 2023. URL https://arxiv.org/abs/2310.13030.

Qianhao Yu, Naishan Zheng, Jie Huang, and Feng Zhao. Cnsnet: A cleanness-navigated-shadow network for shadow removal, 2022. URL https://arxiv.org/abs/2209.02174.

Ziqi Zeng, Chen Zhao, Weiling Cai, and Chenyu Dong. Semantic-guided adversarial diffusion model for self-supervised shadow removal, 2024. URL https://arxiv.org/abs/2407.01104.

Ziqi Zeng, Chen Zhao, Weiling Cai, and Yuqing Guo. Semantic-guided coarse-to-fine diffusion model for self-supervised image shadow removal. In Piotr Didyk and Junhui Hou (eds.), *Computational Visual Media*, pp. 271–293, Singapore, 2025. Springer Nature Singapore. ISBN 978-981-96-5812-1.

Edward Zhang, Ricardo Martin-Brualla, Janne Kontkanen, and Brian Curless. No shadow left behind: Removing objects and their shadows using approximate lighting and geometry, 2020. URL https://arxiv.org/abs/2012.10565.

Ling Zhang, Qing Zhang, and Chunxia Xiao. Shadow remover: Image shadow removal based on illumination recovering optimization. *IEEE Transactions on Image Processing*, 24(11):4623–4636, 2015. doi: 10.1109/TIP.2015.2465159.

Lvmin Zhang, Anyi Rao, and Maneesh Agrawala. Adding conditional control to text-to-image diffusion models, 2023. URL https://arxiv.org/abs/2302.05543.

Richard Zhang, Phillip Isola, Alexei A. Efros, Eli Shechtman, and Oliver Wang. The unreasonable effectiveness of deep features as a perceptual metric. In *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 586–595, 2018. doi: 10.1109/CVPR.2018.00068.

Wei Zhang, Xiang Zhong Fang, Xiaokang K. Yang, and Q. M. Jonathan Wu. Moving cast shadows detection using ratio edge. *IEEE Transactions on Multimedia*, 9(6):1202–1214, 2007. doi: 10.1109/TMM.2007.902842.

Wuming Zhang, Xi Zhao, Jean-Marie Morvan, and Liming Chen. Improving shadow suppression for illumination robust face recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 41(3):611–624, 2019. doi: 10.1109/TPAMI.2018.2803179.

Yiqi Zhong, Xianming Liu, Deming Zhai, Junjun Jiang, and Xiangyang Ji. Shadows can be dangerous: Stealthy and effective physical-world adversarial attack by natural phenomenon. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 15345–15354, June 2022.

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 *ECCV*, 2018.

Lei Zhu, Ke Xu, Zhanghan Ke, and Rynson W.H. Lau. Mitigating intensity bias in shadow detection via feature decomposition and reweighting. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 4702–4711, October 2021.

Xiujin Zhu, Chee-Onn Chow, and Joon Huang Chuah. Shadowmamba: State-space model with boundary-region selective scan for shadow removal, 2025. URL https://arxiv.org/abs/2411.03260.

Yurui Zhu, Jie Huang, Xueyang Fu, Feng Zhao, Qibin Sun, and Zheng–Jun Zha. Bijective mapping network for shadow removal. In *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 5617–5626, 2022a. doi: 10.1109/CVPR52688.2022.00554.

Yurui Zhu, Zeyu Xiao, Yanchi Fang, Xueyang Fu, Zhiwei Xiong, and Zheng-Jun Zha. Efficient model-driven network for shadow removal. 2022b.