Shadow Removal Refinement via Material-Consistent Shadow Edges

Shilin Hu[1], Hieu Le[2], ShahRukh Athar[1,3], Sagnik Das[1], Dimitris Samaras[1]

[1]Stony Brook University   [2]EPFL   [3]Captions

Abstract

Shadow boundaries can be confused with material boundaries as both exhibit sharp changes in luminance or contrast within a scene. However, shadows do not modify the intrinsic color or texture of surfaces. Therefore, on both sides of shadow edges traversing regions with the same material, the original color and texture should be the same if the shadow is removed properly. These shadow/shadow-free pairs are very useful but difficult-to-collect supervision signals. The crucial contribution of this paper is to learn how to identify those shadow edges that traverse material-consistent regions and how to use them as self-supervision for shadow removal refinement during test time. To achieve this, we fine-tune SAM, an image segmentation foundation model, to produce a shadow-invariant segmentation and then extract material-consistent shadow edges by comparing the SAM segmentation with the shadow mask. Utilizing these shadow edges, we introduce color- and texture-consistency losses to enhance the shadow removal process. We demonstrate the effectiveness of our method in improving shadow removal results on more challenging, in-the-wild images, outperforming the state-of-the-art shadow removal methods. Additionally, we propose a new metric and an annotated dataset for evaluating the performance of shadow removal methods without the need for paired shadow/shadow-free data. Our code and dataset are available at: https://github.com/cvlab-stonybrook/ShadowRemovalRefine

1. Introduction

Shadow edges were extensively studied in early shadow removal work [12,34,53]. These edges delineate the boundary between shadow and non-shadow areas, providing a clear indication of the transition in pixel colors and intensity that characterize the shadow, serving as useful supervision signals for shadow removal [24, 27, 46]. However, shadow edges can also coincide with object boundaries. In these cases, the color and texture differences between the two sides of the shadow edges result from both the shadow effect and the different materials, making them difficult to

disentangle. Therefore, identifying shadow edges that cross areas of consistent material is beneficial for shadow removal. However, this is challenging since object boundaries can be confused with shadow boundaries as both exhibit sharp changes in luminance or contrast within a scene. In this work, we propose a novel method to identify material-consistent shadow edges and use them as supervision signals for shadow removal.

To identify material-consistent shadow edges, we propose to fine-tune SAM [25], an image segmentation foundation model, to produce a shadow-invariant segmentation. The main idea is to force SAM to output the same segmentation mask, with or without shadows, thereby guiding the model to be less responsive to the presence of shadows. We then compare the output shadow-invariant segmentation with the shadow mask to obtain material-consistent shadow edges. These are shadow edges that do not coincide with any material edges from SAM.

These material-consistent shadow edges can be extracted for any image. We show that at test time, we can extract these edges and use them as supervision to refine shadow-removal results. To do so, we sample pixels and patches alongside the shadow edges to form shadow/shadow-free pairs. We then introduce two losses: an RGB distance loss and an RGB distribution loss. The RGB distance loss calculates the minimum distance between each sampled shadow and non-shadow pixel along the selected edges to recover the correct color of the shadow region. The RGB distribution loss computes the Earth Mover's Distance (EMD) between the color distributions of the sampled pixels to ensure texture consistency alongside the shadow edges. We further use the Learned Perceptual Image Patch Similarity (LPIPS) [56] loss on sampled patches within the same material to constrain the texture consistency between the non-edge shadow/non-shadow region. By imposing these constraints, we refine the pre-trained model in a *self-supervised* manner, enabling adaptation to more complex shadow images (see Fig. 1).

Further, we propose a novel evaluation metric based on material-consistent shadow edges, namely Color Distribution Difference (CDD). In essence, we can evaluate shadow removal performance by measuring the disparity in pixel

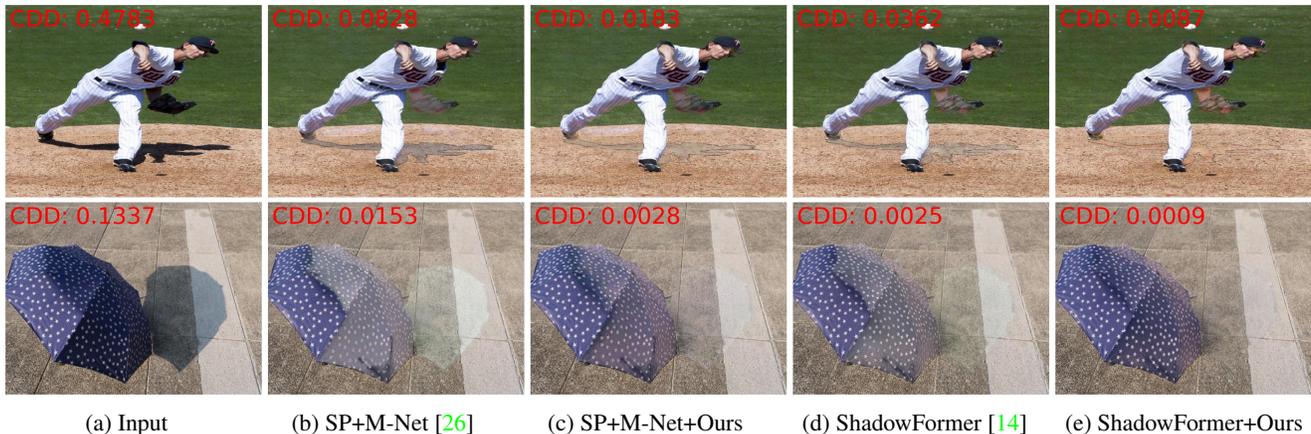| (a) Input | (b) SP+M-Net [26] | (c) SP+M-Net+Ours | (d) ShadowFormer [14] | (e) ShadowFormer+Ours |

Figure 1. Examples from our proposed SBU-S (*top*) and CUHK-S (*bottom*) testing sets. We show the shadow removal results of two existing state-of-the-art methods, SP+M-Net [26] and ShadowFormer [14], in columns (b) and (d), respectively, for two challenging cases. The results of both methods are significantly improved when used jointly with our refinement method, as shown in columns (c) and (e). Our proposed Color Distribution Difference (CDD) metric for each image is shown in red, which can measure shadow removal performance without the need for shadow-free images.

color distribution on both sides of the material-consistent shadow edges. These edges can be easily annotated, even in cases where shadow-free images are hard to obtain. A lower CDD value corresponds to a more effective shadow removal performance, signifying a closer alignment between the textures on either side of the shadow edge, as shown in Fig. 1. This new evaluation scheme enables benchmarking shadow removal methods on challenging, in-the-wild shadow images. We curate a test set sourced from existing shadow detection datasets, namely SBU [42] and CUHK [21]. These datasets contain images under complex shadow scenarios, providing a more comprehensive representation of shadow images in general. We annotate pixels on both sides of the shadow edge that belong to the same background material for each image within the proposed dataset. This benchmark test set will serve as a valuable resource for evaluating the generalizability of shadow removal methods on complex shadow images.

To summarize, our contributions are as follows:

- We introduce a novel shadow edge extraction module designed to identify shadow edges that traverse the same material. We achieve this by fine-tuning SAM to generate shadow-invariant segmentation and then comparing this segmentation with the shadow mask to perform the extraction.

- We propose a test-time adaptation method that refines the shadow removal results, which relies on the extracted self-supervision signal to enforce the material consistency between shadow and non-shadow regions.

- We curate a dataset featuring shadow images in general scenes, serving as a benchmark for assessing shadow

removal methods in complex scenarios. We propose a novel evaluation metric, Color Distribution Difference (CDD), to assess the shadow removal performance, even when shadow-free ground truth is not available.

- Experimental results demonstrate that our method can be seamlessly integrated with existing models, significantly enhancing performance on complex shadow images. Specifically, when combined with two state-of-the-art shadow removal approaches, SP+M-Net [26] and ShadowFormer [14], our method outperforms them by at least 30% on the CDD measurement on our proposed test set.

2. Related Works

Early-stage shadow removal research [6, 9–11] was motivated by physical modeling of illumination and color, typically using a light source-occluder system [1, 22, 39]. The aim was to find useful self-supervision signals to fit the model and then remove the shadows. Guo *et al.* [17] proposed identifying pairs of regions under different illuminations within the same material. Several methods [24, 38, 45, 46] looked for cues at shadow edges, typically involving hand-crafted features designed to capture illumination and color changes. However, these methods relied on idealized assumptions from physical modeling, which do not align with real-world settings due to the complexity and variability of shadow appearances.

State-of-the-art shadow removal methods include deep networks trained end-to-end on pairs of shadow/shadow-free images [3, 4, 13, 14, 28, 33, 44, 57], taking advantage of the powerful ability in learning mappings from the train-

ing pairs. Iterative refinement [7, 8] has also been adopted for shadow removal, *e.g*. ARGAN [5], which progressively removes shadows using a multi-step generator, with each step refining the output by removing remaining shadows. Recently, [15, 23, 35, 36] have utilized diffusion models for shadow removal via iterative denoising. A method similar to ours is the work of Guo *et al.* [16] that also explored boundary cues for shadow removal. They found that their illumination model was insufficient to model the boundary region, necessitating additional supervision during training.

A few methods [20, 27, 30] have sought to mitigate the dependence on paired data. Hu *et al.* [20] proposed Mask-ShadowGAN, which utilizes unpaired data to learn the adaptation from the shadow-free domain to the shadow domain and vice versa. Le and Samaras [27] introduced a technique for cropping unpaired patches from the same shadow image to avoid the need for a shadow-free image. However, these un-/semi-supervised methods have not surpassed their fully-supervised counterparts.

Our method is among the first to employ test-time adaptation (TTA) for shadow removal. In general, TTA [41] aims to enhance a pre-trained model's performance on specific test data. Xiao *et al.* [47] and Yuan *et al.* [54] proposed energy-based models to align target samples with the source distribution. MEMO [55] augmented test samples in various ways to encourage consistent and confident predictions for test-time robustness. TENT [43] focused on the fully test-time adaptation setting, similar to ours, by using only test data and a specific test loss for adaptation. In our case, we search for supervision from each testing sample for test-time adaption. By leveraging the learning capacity of deep models and the guidance from material-consistent edges, we propose extracting self-supervision signals for refining shadow removal performance during inference.

3. Method

In this section, we describe our novel self-supervised test-time adaptation method for refining shadow removal given a shadow image and its shadow mask. Our approach introduces a shadow edge extraction module that leverages shadow/non-shadow pairs along shadow edges to provide supervision. We fine-tune the Segment Anything Model (SAM) [25] to segment materials regardless of shadows, enabling the identification of shadow edges crossing the same material.

During inference, shadow edges guide an iterative refinement process that adjusts shadow removal by analyzing color transitions across edges and enforcing texture consistency within material masks. This self-supervised adaptation improves pre-trained shadow removal models, achieving robust results on complex real-world shadow images.

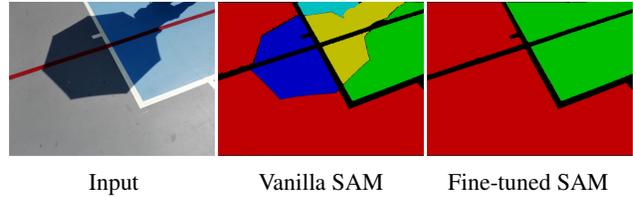|		
Input	Vanilla SAM	Fine-tuned SAM

Figure 2. **Fine-tuning SAM for Material-Consistent Edge Extraction.** Given the input image, we compare the segmentation results of the vanilla SAM [25] with our fine-tuned SAM. Our fine-tuned SAM achieves shadow-invariant segmentation, preserving the material consistency of each mask. In contrast, the vanilla SAM is sensitive to shadow presence, segmenting shadow regions as individual masks.

3.1. Material-Consistent Shadow Edge Extraction

Shadow edges mark the boundaries between shadow and shadow-free areas. The regions on either side of these edges offer potential supervision for shadow removal, as they contain areas with matching colors and textures. However, not all shadow edges reflect consistent underlying materials. For instance, when the shadow edges align with object boundaries, the two sides of the shadow edges display disparate colors and textures, lacking material consistency and rendering them unsuitable for supervising shadow removal. To address this, we propose segmenting materials in the image and selecting only shadow edges that traverse the same material, ensuring reliable and consistent supervision.

We fine-tune the Segment Anything Model (SAM) to produce shadow-invariant material segmentations. While vanilla SAM excels in semantic segmentation, its sensitivity to shadows often causes partially shadowed materials to be split into separate masks (Fig. 2). To address this, we freeze the image and prompt encoders and fine-tune only the mask decoder, following a common strategy [2, 48]. Unlike [2], which adapts prompt encoders, we use the default uniform grid of point prompts [25] on shadow images. SAM is fine-tuned to achieve shadow-invariant segmentation by aligning its outputs with segmentations from shadow-free counterparts. This is enforced using Dice Loss [40], ensuring similarity between masks predicted with and without shadows. The fine-tuned SAM effectively produces shadow-invariant material segmentations, as shown in Fig. 2.

For a given test image, we compare the shadow mask with the fine-tuned SAM segments, which highlight same-material boundaries. Shadow edges intersecting SAM segments are identified as material-consistent shadow edges. This allows us to extract edges crossing the same material and sample pixels near these edges, as well as patches from shadow and non-shadow regions, enabling effective supervision for shadow removal.

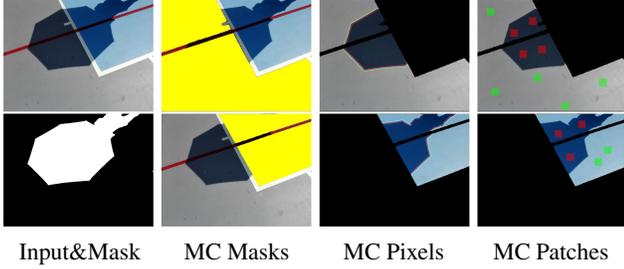| Input&Mask | MC Masks | MC Pixels | MC Patches |

Figure 3. **Supervision for the adaptation.** Given the input image and the shadow mask, we first use the fine-tuned SAM to produce material-consistent (MC) masks. We sample pixels on both sides alongside the MC shadow edge, denoted as S_{in} (shown in red) and S_{out} (shown in green), and patches within the same material, denoted as P_{in} (shown in red) and P_{out} (shown in green). Pixels and patches on both sides of the shadow edge provide supervision for the subsequent adaptation process.

3.2. Test-time Adaptation

Due to the complexity and variability in shadow appearances, existing methods often struggle when confronted with out-of-distribution shadow images, leading to residual shadow effects. To address this challenge, we propose an iterative adaptation approach for pre-trained models based on supervision derived from shadow edges and shadow-invariant segmentation masks.

To ensure color consistency between the shadow and non-shadow regions, we sample pixels along the extracted shadow edges, denoted as S_{in} and S_{out} (see Fig. 3). $S_{in} = \{u_1, \ldots, u_i, \ldots, u_M\}, i \in [1, M]$ represents the pixels inside the shadow boundary, obtained by subtracting an eroded shadow mask from the original shadow mask. Similarly, $S_{out} = \{v_1, \ldots, v_j, \ldots, v_N\}, j \in [1, N]$ is obtained by subtracting the shadow mask from its dilated version. To smooth out the boundaries of the binary shadow mask, we apply dilation and erosion operations to the shadow mask before subtraction. We then introduce two novel losses to constrain color consistency: 1) The RGB distance loss calculates the mean of the minimum distance between each pixel in S_{in} and all pixels in S_{out}, enforcing the restoration of the correct color of the shadow region; 2) The RGB distribution loss computes the Earth Mover's Distance (EMD) between the color distributions of the two sets of sampled pixels, ensuring consistency in color distributions within the sampled pixel sets. The losses are formulated as:

$$\mathcal{L}_{distance} = \frac{\sum_{i=1}^{M} \min_{j \in [1,N]} d(u_i, v_j)}{M}, \qquad (1)$$

$$\mathcal{L}_{distribution} = \text{EMD}([Hist(S_{in}), Hist(S_{out})]), \qquad (2)$$

where $d(u, v)$ denotes the Euclidean distance between two pixels in RGB color space, and $Hist(S)$ is the histogram of sampled pixel colors across RGB channels.

To ensure texture consistency between the shadow and non-shadow regions, we randomly sample 16×16 patches within the segmentation mask of the same material, denoted as P_{in} and P_{out} (see Fig. 3). $P_{in} = \{p_1, \ldots, p_i, \ldots, p_M\}, i \in [1, M]$ represents patches in the shadow region, $P_{out} = \{q_1, \ldots, q_j, \ldots, q_N\}, j \in [1, N]$ represents patches in the non-shadow region. Note that, to ensure the patches are within the material masks, we apply an erosion operation to the material mask before sampling. We then compute the mean of the minimum of the Learned Perceptual Image Patch Similarity (LPIPS) [56] loss between each patch in the shadow region and all patches in the non-shadow region. The loss is formulated as follows:

$$\mathcal{L}_{per} = \frac{\sum_{i=1}^{M} \min_{j \in [1,N]} \text{LPIPS}(p_i, q_j)}{M}, \qquad (3)$$

We compute the average across all material masks for $\mathcal{L}_{distance}, \mathcal{L}_{distribution},$ and $\mathcal{L}_{per}$. Finally, the pre-trained shadow removal model is updated iteratively to adapt to each testing case, using the weighted sum of all three losses:

$$\mathcal{L}_{total} = \lambda_1 \cdot \mathcal{L}_{distance} + \lambda_2 \cdot \mathcal{L}_{distribution} + \lambda_3 \cdot \mathcal{L}_{per}, \quad (4)$$

where $(\lambda_1, \lambda_2, \lambda_3)$ are controlling parameters.

Our method can be seamlessly integrated into existing pre-trained models. The shadow edge extraction serves as a standalone module while minimizing the proposed losses is the training objective of the pre-trained models.

4. New Benchmark Test set and Evaluation Metric

The most commonly used datasets for shadow removal training and evaluation are ISTD [44], ISTD+ [26], and SRD [37], which provide triplets of shadow image, shadow mask, and shadow-free ground truth. However, these datasets contain only simple shadows as they lack occluders within the image. We argue that the current evaluation data is insufficient to assess the adaptability of current shadow removal methods due to limitations in data diversity. Collecting paired data is laborious [18, 19, 29, 31, 32, 49–52], with inherent limitations in diversity.

To address this gap, we curate a benchmark test set of shadow images in the wild. These images are sourced from existing shadow detection datasets, namely SBU [42] and CUHK [21], exhibiting more complex scenes and larger variations in illumination, thus better representing shadow images in real-world scenarios. The proposed dataset contains 400 images in total. Among them, 210 shadow images are from SBU (SBU-S), and 190 images from CUHK (CUHK-S). To enable assessment of shadow removal performance on the proposed benchmark test set, we annotate pixels alongside the shadow edges that traverse the same material for each image in the dataset, examples of images

(a) Shadow Images from SBU-S   (b) Shadow Images from CUHK-S

Figure 4. Example images from our proposed benchmark test set for shadow removal performance evaluation. Annotated pixels alongside the shadow edges are also shown for each image: S_{in} in red and S_{out} in green. (a) shows examples from SBU-S, and (b) shows images from CUHK-S.

and corresponding annotations are shown in Fig. 4. We believe that this dataset can serve as a valuable resource for evaluating shadow removal performance on general shadow images in real-world scenarios.

Another challenge in assessing shadow removal for shadow images in the wild is the lack of an evaluation metric. All current metrics used in shadow removal evaluation, *e.g.* Mean Absolute Error (MAE), Peak Signal-to-Noise Ratio (PSNR), and Structural Similarity Index (SSIM), require comparison between the shadow-removed result and the ground truth shadow-free image. However, due to the complexity of real-world scenarios, obtaining the shadow-free version of the scene is impractical.

To tackle the absence of an effective evaluation metric, we propose a straightforward metric called Color Distribution Difference (CDD). CDD is grounded in the assumption that the area alongside the shadow edge maintains the same underlying texture. The CDD metric calculates the Earth Mover's Distance (EMD) between the color distribution histograms of pixels on both sides of the shadow edges.

$$CDD = \text{EMD}([Hist(S), Hist(NS)]) \qquad (5)$$

where S, NS denotes the pixel sets in the shadow region and the non-shadow region, respectively.

The CDD quantifies the disparity in color distribution and is correlated with the quality of shadow removal performance. A lower mean CDD value indicates a better shadow-removal outcome, reflecting alignment between the colors on both sides of the shadow edge. A lower variance suggests more consistent performance across various shadow appearances, illustrating the method's generalizability.

5. Experiments

Implementation Details. The proposed method is implemented using PyTorch. All experiments are conducted on

Table 1. Comparison with SOTA models. We report the performance of existing state-of-the-art (SOTA) methods on our proposed test set and compare them with our adaptation method. CDD mean and variance values are provided, note that these values are reported as $1000\times$ the original value. $*$ denotes that only the SRD pre-trained model is available for evaluation.

Methods	CUHK-S		SBU-S	
	CDD Mean	CDD Var	CDD Mean	CDD Var
Input	291.6	125.0	279.0	119.0
Inpaint4Shadow [33]	25.3	36.0	28.6	40.7
ShadowDiffusion* [15]	50.2	89.1	82.3	111.9
ShadowDiffusion*+Ours	48.7	77.1	78.2	103.4
SP+M-Net [26]	41.7	52.9	43.0	54.9
SP+M-Net+Ours	22.2	36.5	23.9	46.7
ShadowFormer [14]	23.3	33.6	27.9	45.4
ShadowFormer+Ours	**16.4**	**31.9**	**15.0**	**26.3**

an NVIDIA TITAN RTX GPU. We apply our method to three models, ShadowDiffusion [15], SP+M-Net [26] and ShadowFormer [14]. Hyperparameters are set to be the same as in the training phase, except for the learning rate, which is set to $1e^{-5}$. During the adaptation process, we update the pre-trained model for 20 iterations per image. The refining parameters λ_1, λ_2 and λ_3 are set to $1, 1, 0.1$ in our experiments. More details are in the supplementary.

Datasets and Evaluation Metrics. We evaluate shadow removal performance for images in the wild on our proposed test set, comprising 400 shadow images from SBU [42] and CUHK [21], with material-consistent shadow edge pixels annotated for each image. As shadow-free images are not available in our test set, we use the proposed Color Distribution Difference (CDD) as the quantitative evaluation metric. Additional evaluation results on the ISTD+ [26] dataset are provided in the supplementary material.

5.1. Comparison with SOTA Models

To demonstrate the seamless integration of our proposed adaptation method with existing shadow removal models, we apply it on top of three state-of-the-art (SOTA) models: SP+M-Net [26], a CNN-based model, ShadowFormer [14], a Transformer-based model, and ShadowDiffusion [15], a diffusion-based model.

Quantitative evaluation. For fair comparison, we evaluate the performance of all existing methods using models trained on the ISTD+ dataset (except for ShadowDiffusion, where only the SRD-trained model is available). As shown in Tab. 1, our adaptation method enables all three methods to surpass their original results, with ShadowFormer achieving the best performance when our adaptation is applied. The high errors from existing SOTA methods indicate limited adaptability to various complex shadow scenes in the wild, as they are trained on simple and limited scenes. In

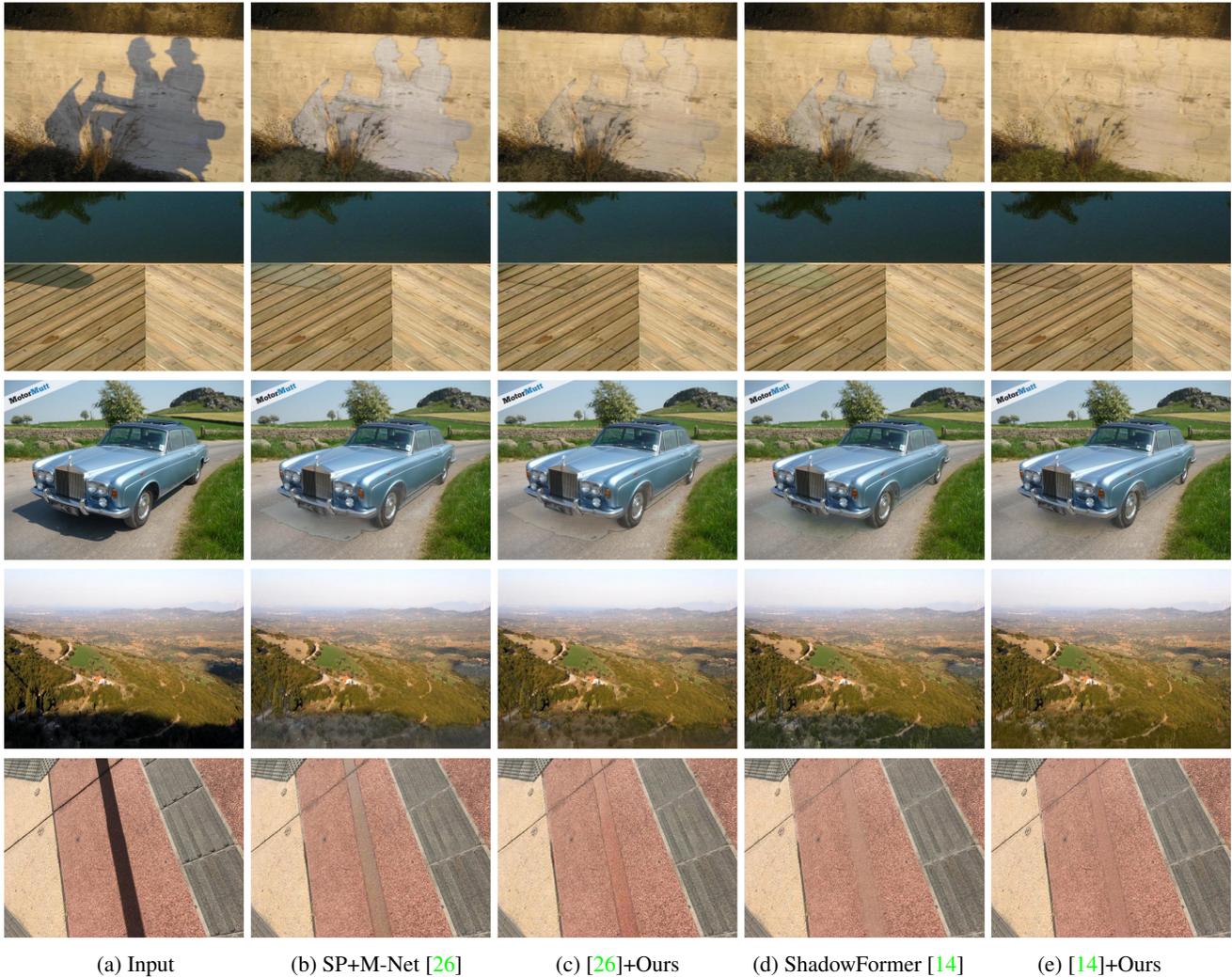|				
(a) Input	(b) SP+M-Net [26]	(c) [26]+Ours	(d) ShadowFormer [14]	(e) [14]+Ours

Figure 5. Examples from test cases in our proposed dataset, (a) shows input images from our curated dataset; the first three rows are from SBU-S, and the last two rows are from CUHK-S. (b) and (c) present results from the pre-trained and refined SP+M-Net [26]. (d) and (e) show results from the pre-trained and refined ShadowFormer [14].

contrast, our adaptation method leverages self-supervision from shadow edges in each image, refining each case during test time. Notably, the SRD-trained ShadowDiffusion performs the worst among the SOTA methods due to the intensity differences between the training shadow and shadow-free pairs, which we discuss further in Sec. 5.4.

Qualitative evaluation. As shown in Fig. 5, we demonstrate the improvement with our proposed adaptation method on both base models. Column (a) shows the input shadow images. In columns (b) and (d), we can see that the pre-trained models struggle with complex shadows of various shapes, near the dark materials, and with visible occluders in the scene, resulting in obvious artifacts in the shadow region. With our adaptation method, as shown in columns (c) and (e), we fix the color shift in the shadow

region. This is because our proposed adaptation constrains the color and texture consistency between the shadow and non-shadow regions.

5.2. Usefulness of CDD

We propose the Color Distribution Difference (CDD) metric to address the lack of evaluation metrics for shadow removal in real-world scenarios, where existing metrics rely on ground truth shadow-free images. To validate the usefulness of our CDD metric, we plot the Mean Absolute Error (MAE) values alongside the CDD values for each shadow image in the ISTD+ test set [26] in Fig. 6. The plot demonstrates a strong correlation between the two metrics, highlighting that CDD serves as a reliable alternative for evaluating shadow removal methods in the absence of shadow-free
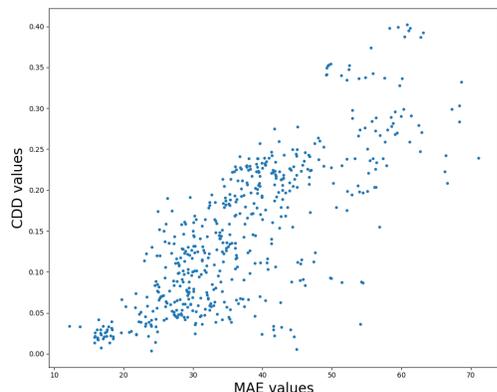
Figure 6. Per-image CDD and MAE errors for shadow images in the ISTD+ [26] test set. Each point represents the CDD and MAE errors computed for a single image. The two metrics are highly correlated. Note that measuring MAE requires a paired shadow-free image while measuring the proposed CDD only requires annotating MC-shadow edges in the input shadow image.

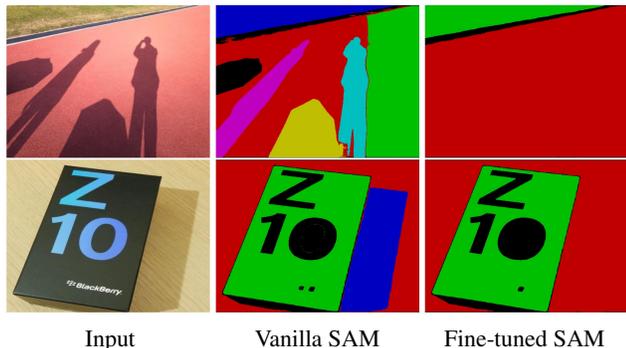Input   Vanilla SAM   Fine-tuned SAM

Figure 7. Comparison of vanilla SAM and fine-tuned SAM on test images. Our fine-tuned SAM predicts shadow-invariant masks while vanilla SAM is sensitive to shadow regions.

ground truth images.

5.3. Ablation Studies

The effect of Fine-tuned SAM. We fine-tune the Segment Anything Model [25] to extract shadow edges that traverse the same material. In Fig. 7, we demonstrate that the fine-tuned SAM produces shadow-invariant masks. Additionally, Tab. 2 compares the performance of using vanilla SAM and our fine-tuned SAM for adaptation. The latter surpasses the former due to its reduced sensitivity to shadow presence. **The effectiveness of the Proposed Adaptation Method.** Tab. 2 shows evaluations of different configurations for the adaptation method. We find that the proposed approach, which combines fine-tuned SAM edge extraction with pixel and patch sampling for supervision, yields the best overall results. "Per Mask" refines the edge pixels and patches on each SAM-detected material mask that intersects with the shadow mask, and "Pixels" refines only the pixels on

Table 2. Quantitative results for different configurations of the adaptation method and with different loss settings. We report the CDD mean and variance values on our proposed dataset. Note that the CDD values are reported as $1000\times$ the original value.

Configurations	CUHK-S		SBU-S	
	CDD Mean	CDD Var	CDD Mean	CDD Var
VanillaSAM+Pixels&Patches	19.6	34.4	21.3	44.9
FinetunedSAM+Per Mask	16.9	**29.3**	16.8	33.6
FinetunedSAM+Pixels	16.9	32.1	15.5	27.0
FinetunedSAM+Pixels&Patches	**16.4**	31.9	**15.0**	**26.3**
Losses	CDD Mean	CDD Var	CDD Mean	CDD Var
All losses	**16.4**	**31.9**	**15.0**	26.3
- $\mathcal{L}_{distance}$	16.8	32.7	15.0	**25.8**
- $\mathcal{L}_{distribution}$	19.2	30.9	21.4	43.6
- $\mathcal{L}_{per}$	16.9	32.1	15.5	27.0

Input   Per Mask   Pixels   Pixels&Patches

Figure 8. Comparison of different refinement configurations. Refining pixels on all extracted shadow edges and patches per material mask yields the best performance.

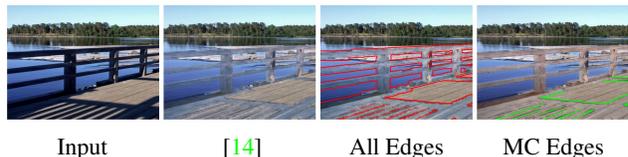Input   [14]   All Edges   MC Edges

Figure 9. Example of the proposed edge extraction. We show the refined outputs using all shadow edges (colored in red), and the refined outputs using only our extracted material-consistent shadow edges (colored in green). Using the extracted MC edges improves the performance from 0.0372 to 0.0039.

all extracted edges at once. We can see that the "Pixels&Patches" configuration achieves the best overall results (shown in Fig. 8), as it maintains texture consistency within each material while ensuring global consistency across all the extracted shadow edges. When no edge extraction is performed, the entire shadow edge is used, including those edges coinciding with material boundaries. The adaptation process is misguided because these edges separate different materials on either side, as depicted in Fig. 9.

Proposed Losses. Our adaptation method comprises three loss functions. In Tab. 2, we present experimental results from the ablation of each of these losses. $\mathcal{L}_{distance}$ and $\mathcal{L}_{distribution}$ enforce color matching and distribution alignment along the shadow edges, while the $\mathcal{L}_{per}$ ensures texture consistency within the same material. The combination of these three losses yields the best performance.

Table 3. Quantitative results of cross dataset testing. ISTD pre-trained ShadowFormer and SRD pre-trained ShadowFormer are tested on the ISTD+ test set.

Trained On	Tested On	Methods	MAE			CDD	
			S	NS	A	Mean	Var
SRD	ISTD+	w.o. Ours	13.7	3.4	5.1	55.0	43.3
		w. Ours	6.2	2.4	3.0	8.0	9.4
ISTD	ISTD+	w.o. Ours	10.6	6.3	7.0	11.8	17.7
		w. Ours	6.3	2.7	3.4	1.0	3.1

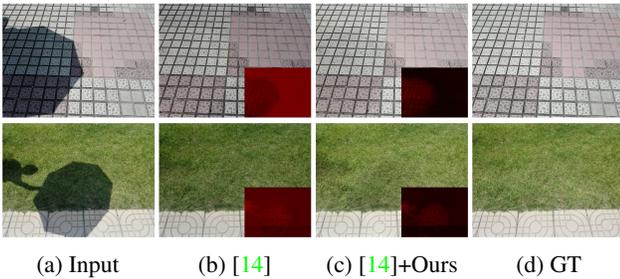(a) Input   (b) [14]   (c) [14]+Ours   (d) GT

Figure 10. Qualitative comparison in cross dataset testing. We use ISTD pre-trained ShadowFormer and test it on the ISTD+ test set. (a) shows input image, (b) shows ShadowFormer [14] result, (c) presents the results with refinement, and (d) presents the ground truth. Error maps are also plotted in the corner.

5.4. Cross-Dataset Testing

To further demonstrate the performance of our adaptation method on out-of-distribution images, we apply our method to ShadowFormer [14] pre-trained on the SRD dataset [37] (shadow masks provided by DHAN [4]) and the ISTD dataset [44], testing on the ISTD+ test set. Note that images from the SRD and ISTD datasets exhibit different light intensities between training shadow and shadow-free image pairs due to the image acquisition process. As a result, models pre-trained on these datasets often alter the overall color intensity of the whole input image.

To address this, we further calculate the Mean Squared Error (MSE) loss of the non-shadow region pixels, denoted as $\mathcal{L}_{nonshadow}$. Tab. 3 presents the ISTD+ test results for these models with and without our proposed adaptation method. With our adaptation, the output images exhibit correct colors for both shadow and non-shadow regions, outperforming pre-trained models in both MAE and CDD measurements. As illustrated in Fig. 10, the ISTD pre-trained model alters the overall light intensity of the image, resulting in subpar performance on ISTD+ test cases. In contrast, our adaptation method effectively corrects this error by incorporating $\mathcal{L}_{nonshadow}$, which enforces color consistency in the non-shadow regions.

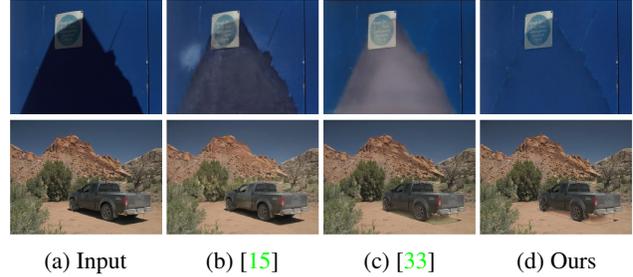(a) Input   (b) [15]   (c) [33]   (d) Ours

Figure 11. **Limitation on shadow edges.** Results of SOTA methods on our test images show visible edge artifacts caused by the limited adaptability of pre-trained models. Our refinement mainly focuses on addressing color discrepancies in the shadow regions.

5.5. Limitations

Our adaptation method has several limitations that can be interesting directions for future work. First, we do not enforce specific constraints on penumbra shadow regions. Handling the smoothly varying shadow effects in those areas is challenging for all shadow removal methods, including ours, which often leaves noticeable shadow edge artifacts (see Fig. 11). While our method focuses on correcting color discrepancies, future research could explore mitigating edge artifacts, particularly for out-of-distribution test cases. Second, our approach involves iterative refinement of the pre-trained model using extracted self-supervision, leading to a computational overhead of approximately 24 seconds per image. Finally, we primarily focus on the supervision obtained from material-consistent edges, but the remaining portions of the shadow edge might also help understand and remove shadows.

6. Conclusion

We introduce a test-time self-supervised adaptation method for deep-learning-based shadow removal. To gather the supervision signal, we fine-tune the image foundation model, SAM, to generate shadow-invariant segmentation masks, effectively extracting shadow edges that traverse the same material. Pixels near these edges and patches within the same material provide valuable supervision. We then propose an iterative adaptation approach for the pre-trained model using the collected supervision to ensure color and texture consistency. We demonstrate that our proposed losses significantly enhance deep shadow removal, both qualitatively and quantitatively, across various challenging testing cases. Furthermore, we introduce a benchmark test set and a metric that enable the evaluation of shadow removal methods on images with complex shadows, even in the absence of shadow-free ground truth images.

References

[1] Harry Barrow, J Tenenbaum, A Hanson, and E Riseman. Recovering intrinsic scene characteristics. *Comput. vis. syst*, 2(3-26):2, 1978. 2

[2] Tianrun Chen, Lanyun Zhu, Chaotao Deng, Runlong Cao, Yan Wang, Shangzhan Zhang, Zejian Li, Lingyun Sun, Ying Zang, and Papa Mao. Sam-adapter: Adapting segment anything in underperformed scenes. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 3367–3375, 2023. 3

[3] Zipei Chen, Chengjiang Long, Ling Zhang, and Chunxia Xiao. Canet: A context-aware network for shadow removal. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 4743–4752, 2021. 2

[4] Xiaodong Cun, Chi-Man Pun, and Cheng Shi. Towards ghost-free shadow removal via dual hierarchical aggregation network and shadow matting gan. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 34, pages 10680–10687, 2020. 2, 8

[5] Bin Ding, Chengjiang Long, Ling Zhang, and Chunxia Xiao. Argan: Attentive recurrent generative adversarial network for shadow detection and removal. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 10213–10222, 2019. 3

[6] Mark S Drew, Graham D Finlayson, and Steven D Hordley. Recovery of chromaticity image free from shadows via illumination invariance. In *IEEE Workshop on Color and Photometric Methods in Computer Vision, ICCV'03*, pages 32–39, 2003. 2

[7] Nikita Durasov, Nik Dorndorf, Hieu Le, and Pascal Fua. Zigzag: Universal sampling-free uncertainty estimation through two-step inference. *Transactions on Machine Learning Research*, 2024. 3

[8] Nikita Durasov, Doruk Oner, Jonathan Donier, Hieu Le, and Pascal Fua. Enabling uncertainty estimation in iterative neural networks. In *Forty-first International Conference on Machine Learning*, 2024. 3

[9] Graham D Finlayson and Mark S Drew. 4-sensor camera calibration for image representation invariant to shading, shadows, lighting, and specularities. In *Proceedings Eighth IEEE International Conference on Computer Vision. ICCV 2001*, volume 2, pages 473–480. IEEE, 2001. 2

[10] Graham D Finlayson, Mark S Drew, and Cheng Lu. Entropy minimization for shadow removal. *International Journal of Computer Vision*, 85(1):35–57, 2009. 2

[11] Graham D Finlayson, Steven D Hordley, and Mark S Drew. Removing shadows from images. In *Computer Vision—ECCV 2002: 7th European Conference on Computer Vision Copenhagen, Denmark, May 28–31, 2002 Proceedings, Part IV 7*, pages 823–836. Springer, 2002. 2

[12] Graham D Finlayson, Steven D Hordley, Cheng Lu, and Mark S Drew. On the removal of shadows from images. *IEEE transactions on pattern analysis and machine intelligence*, 28(1):59–68, 2005. 1

[13] 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. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 10571–10580, 2021. 2

[14] Lanqing Guo, Siyu Huang, Ding Liu, Hao Cheng, and Bihan Wen. Shadowformer: Global context helps image shadow removal. *arXiv preprint arXiv:2302.01650*, 2023. 2, 5, 6, 7, 8

[15] 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 *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 14049–14058, 2023. 3, 5, 8

[16] 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 *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 13045–13054, 2023. 3

[17] 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, 2012. 2

[18] Prantik Howlader, Srijan Das, Hieu Le, and Dimitris Samaras. Beyond pixels: Semi-supervised semantic segmentation with a multi-scale patch-based multi-label classifier. In *ECCV*, 2024. 4

[19] Prantik Howlader, Hieu Le, and Dimitris Samaras. Weighting pseudo-labels via high-activation feature index similarity and object detection for semi-supervised segmentation. In *ECCV*, 2024. 4

[20] Xiaowei Hu, Yitong Jiang, Chi-Wing Fu, and Pheng-Ann Heng. Mask-shadowgan: Learning to remove shadows from unpaired data. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 2472–2481, 2019. 3

[21] 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, 4, 5

[22] Xiang Huang, Gang Hua, Jack Tumblin, and Lance Williams. What characterizes a shadow boundary under the sun and sky? In *2011 international conference on computer vision*, pages 898–905. IEEE, 2011. 2

[23] Yeying Jin, Wei Ye, Wenhan Yang, Yuan Yuan, and Robby T Tan. Des3: Adaptive attention-driven self and soft shadow removal using vit similarity. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 38, number 3, pages 2634–2642, 2024. 3

[24] Salman H Khan, Mohammed Bennamoun, Ferdous Sohel, and Roberto Togneri. Automatic shadow detection and removal from a single image. *IEEE transactions on pattern analysis and machine intelligence*, 38(3):431–446, 2015. 1, 2

[25] Alexander Kirillov, Eric Mintun, Nikhila Ravi, Hanzi Mao, Chloe Rolland, Laura Gustafson, Tete Xiao, Spencer Whitehead, Alexander C Berg, Wan-Yen Lo, et al. Segment anything. In *Proceedings of the IEEE/CVF International Con-*

ference on Computer Vision, pages 4015–4026, 2023. 1, 3, 7

[26] Hieu Le and Dimitris Samaras. Shadow removal via shadow image decomposition. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 8578–8587, 2019. 2, 4, 5, 6, 7

[27] Hieu Le and Dimitris Samaras. From shadow segmentation to shadow removal. In *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XI 16*, pages 264–281. Springer, 2020. 1, 3

[28] 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, 2021. 2

[29] Hieu Le, Dimitris Samaras, and Heather J Lynch. A convolutional neural network architecture designed for the automated survey of seabird colonies. *Remote Sensing in Ecology and Conservation*, 8(2):251–262, 2022. 4

[30] Hieu Le, Tomas F Yago Vicente, Vu Nguyen, Minh Hoai, and Dimitris Samaras. A+ d net: Training a shadow detector with adversarial shadow attenuation. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pages 662–678, 2018. 3

[31] Hieu Le, Chen-Ping Yu, Gregory Zelinsky, and Dimitris Samaras. Co-localization with category-consistent features and geodesic distance propagation. In *Proceedings of the IEEE International Conference on Computer Vision Workshops*, pages 1103–1112, 2017. 4

[32] Hieu M Le, Bento Collares Gonçalves, Dimitris Samaras, and Heather J Lynch. Weakly labeling the antarctic: The penguin colony case. In *CVPR Workshops*, pages 18–25, 2019. 4

[33] Xiaoguang Li, Qing Guo, Rabab Abdelfattah, Di Lin, Wei Feng, Ivor Tsang, and Song Wang. Leveraging inpainting for single-image shadow removal. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 13055–13064, 2023. 2, 5, 8

[34] Feng Liu and Michael Gleicher. Texture-consistent shadow removal. In *Computer Vision–ECCV 2008: 10th European Conference on Computer Vision, Marseille, France, October 12-18, 2008, Proceedings, Part IV 10*, pages 437–450. Springer, 2008. 1

[35] Yuhao Liu, Zhanghan Ke, Ke Xu, Fang Liu, Zhenwei Wang, and Rynson WH Lau. Recasting regional lighting for shadow removal. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 38, pages 3810–3818, 2024. 3

[36] Kangfu Mei, Luis Figueroa, Zhe Lin, Zhihong Ding, Scott Cohen, and Vishal M Patel. Latent feature-guided diffusion models for shadow removal. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pages 4313–4322, 2024. 3

[37] Liangqiong Qu, Jiandong Tian, Shengfeng He, Yandong Tang, and Rynson WH Lau. Deshadownet: A multi-context embedding deep network for shadow removal. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 4067–4075, 2017. 4, 8

[38] John M Rubin and WA Richards. Color vision and image intensities: When are changes material? *Biological Cybernetics*, 45(3):215–226, 1982. 2

[39] Yael Shor and Dani Lischinski. The shadow meets the mask: Pyramid-based shadow removal. In *Computer Graphics Forum*, volume 27, number 2, pages 577–586. Wiley Online Library, 2008. 2

[40] Carole H Sudre, Wenqi Li, Tom Vercauteren, Sebastien Ourselin, and M Jorge Cardoso. Generalised dice overlap as a deep learning loss function for highly unbalanced segmentations. In *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support: Third International Workshop, DLMIA 2017, and 7th International Workshop, ML-CDS 2017, Held in Conjunction with MICCAI 2017, Québec City, QC, Canada, September 14, Proceedings 3*, pages 240–248. Springer, 2017. 3

[41] Yu Sun, Xiaolong Wang, Zhuang Liu, John Miller, Alexei Efros, and Moritz Hardt. Test-time training with self-supervision for generalization under distribution shifts. In *International conference on machine learning*, pages 9229–9248. PMLR, 2020. 3

[42] 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 *Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11-14, 2016, Proceedings, Part VI 14*, pages 816–832. Springer, 2016. 2, 4, 5

[43] Dequan Wang, Evan Shelhamer, Shaoteng Liu, Bruno Olshausen, and Trevor Darrell. Tent: Fully test-time adaptation by entropy minimization. In *International Conference on Learning Representations*, 2021. 3

[44] 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*, pages 1788–1797, 2018. 2, 4, 8

[45] Andrew P Within. Intensity-based edge classification. In *Proceedings of the Second AAAI Conference on Artificial Intelligence*, pages 36–41, 1982. 2

[46] Tai-Pang Wu, Chi-Keung Tang, Michael S Brown, and Heung-Yeung Shum. Natural shadow matting. *ACM Transactions on Graphics (TOG)*, 26(2):8–es, 2007. 1, 2

[47] Zehao Xiao, Xiantong Zhen, Shengcai Liao, and Cees G. M. Snoek. Energy-based test sample adaptation for domain generalization. In *The Eleventh International Conference on Learning Representations*, 2023. 3

[48] Weiyi Xie, Nathalie Willems, Shubham Patil, Yang Li, and Mayank Kumar. Sam fewshot finetuning for anatomical segmentation in medical images. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pages 3253–3261, 2024. 3

[49] Jingyi Xu and Hieu Le. Generating representative samples for few-shot classification. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 9003–9013, 2022. 4

[50] Jingyi Xu, Hieu Le, Mingzhen Huang, ShahRukh Athar, and Dimitris Samaras. Variational feature disentangling

for fine-grained few-shot classification. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 8812–8821, 2021. 4

[51] Jingyi Xu, Hieu Le, Vu Nguyen, Viresh Ranjan, and Dimitris Samaras. Zero-shot object counting. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 15548–15557, 2023. 4

[52] Jingyi Xu, Hieu Le, and Dimitris Samaras. Generating features with increased crop-related diversity for few-shot object detection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 19713–19722, 2023. 4

[53] Li Xu, Feihu Qi, and Renjie Jiang. Shadow removal from a single image. In *sixth international conference on intelligent systems design and applications*, volume 2, pages 1049–1054. IEEE, 2006. 1

[54] Yige Yuan, Bingbing Xu, Liang Hou, Fei Sun, Huawei Shen, and Xueqi Cheng. Tea: Test-time energy adaptation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 23901–23911, 2024. 3

[55] Marvin Zhang, Sergey Levine, and Chelsea Finn. Memo: Test time robustness via adaptation and augmentation. *Advances in neural information processing systems*, 35:38629–38642, 2022. 3

[56] Richard Zhang, Phillip Isola, Alexei A Efros, Eli Shechtman, and Oliver Wang. The unreasonable effectiveness of deep features as a perceptual metric. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 586–595, 2018. 1, 4

[57] Yurui Zhu, Jie Huang, Xueyang Fu, Feng Zhao, Qibin Sun, and Zheng-Jun Zha. Bijective mapping network for shadow removal. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 5627–5636, 2022. 2