Single Image Shadow Detection via Complementary Mechanism

Yurui Zhu
zyr@mail.ustc.edu.cn
University of Science and Technology of China
Hefei, Anhui, China

Xueyang Fu*
xyfu@ustc.edu.cn
University of Science and Technology of China
Hefei, Anhui, China

Chengzhi Cao
chengzhicao@mail.ustc.edu.cn
University of Science and Technology of China
Hefei, Anhui, China

Xi Wang
wangxxi@mail.ustc.edu.cn
University of Science and Technology of China
Hefei, Anhui, China

Qibin Sun
qibinsun@ustc.edu.cn
University of Science and Technology of China
Hefei, Anhui, China

Zheng-Jun Zha
zhazj@ustc.edu.cn
University of Science and Technology of China
Hefei, Anhui, China

ABSTRACT

In this paper, we present a novel shadow detection framework by investigating the mutual complementary mechanisms contained in this specific task. Our method is based on a key observation: in a single shadow image, shadow regions and non-shadow counterparts are complementary to each other in nature, thus a better estimation on one side leads to an improved estimation on the other, and vice versa. Motivated by this observation, we first leverage two parallel interactive branches to jointly produce shadow and non-shadow masks. The interaction between two parallel branches is to retain the deactivated intermediate features of one branch by introducing the negative activation technique, which could serve as complementary features to the other branch. Besides, we also apply identity reconstruction loss as complementary training guidance at the image level. Finally, we design two discriminative losses to satisfy the complementary requirements of shadow detection, *i.e.*, neither missing any shadow regions nor falsely detecting non-shadow regions. By fully exploring and exploiting the complementary mechanism of shadow detection, our method can confidently predict more accurate shadow detection results. Extensive experiments on the three widely-used benchmarks demonstrate our proposed method achieves superior shadow detection performance against state-of-the-art methods with a relatively low computational cost. Our source code is available at this repository.

CCS CONCEPTS

• **Computing methodologies** → *Scene understanding*.

KEYWORDS

Shadow Detection, Neural Networks, Complementary Mechanisms

*Xueyang Fu is the corresponding author (xyfu@ustc.edu.cn).

MM '22, October 10–14, 2022, Lisboa, Portugal

ACM ISBN 978-1-4503-9203-7/22/10…$15.00
https://doi.org/10.1145/3503161.3547904

ACM Reference Format:
Yurui Zhu, Xueyang Fu, Chengzhi Cao, Xi Wang, Qibin Sun, and Zheng-Jun Zha. 2022. Single Image Shadow Detection via Complementary Mechanism. In *Proceedings of the 30th ACM International Conference on Multimedia (MM '22), October 10–14, 2022, Lisboa, Portugal.* ACM, Lisbon, Portugal, 10 pages. https://doi.org/10.1145/3503161.3547904

1 INTRODUCTION

As a common natural phenomenon, shadows are usually cast when associated objects completely or partially block the light sources from a specific direction. Hence, aware of the shadow location could offer valuable visual hints for perceiving the scene geometry [18, 19, 29], the light sources position [23, 30], camera parameters [43], *etc.* This will largely facilitate various scene understanding related works, *e.g.*, image segmentation [8], 3D scene reconstruction [42] and rough geometry estimation [31]. Therefore, detecting shadows from a single image is crucial for computer vision tasks.

Currently, a series of methods have been proposed for this specific task, which could be roughly divided into two flavors: traditional methods and convolutional neural networks (CNNs) methods. The former mainly focuses on designing hand-crafted priors and assumptions, *e.g.*, color chromaticity [4, 5], textures and illumination cues [16, 34, 50]. Since the hand-crafted priors are designed for specific shadow scenes, when the shadow scenarios deviate from the predefined assumptions, the detection results of these methods in complex conditions are often unsatisfactory.

Recently, deep CNNs-based approaches merge as a promising solution for shadow detection and dominate this field. However, how to accurately detect shadows still remains challenging, *e.g.*, existing methods are easily mistaking the ambiguous areas. To address the above drawbacks, researchers attempt to [2, 13, 13, 15, 40, 51] explore effective contextual information or multi-scale techniques to boost the detection performance. There also exist methods to explore the potential hints to resolve these ambiguous areas. For example, FDRNet [52] suggests the intensity-bias cue and devises a novel decomposition and re-weighting strategy to mitigate the intensity-bias for shadow detection. However, such external cues also bring the side effect, *e.g.*, due to the gradual changes of the shadow effects near the shadow boundary [24, 25], the intensity bias will affect the shadow intensity across the shadow boundary, resulting in coarse boundary results. However, there is potential complementary information internally within this specific task that

has not yet been utilized. For example, the above methods usually employ a single-branch framework for shadow detection, which fails to utilize the mutual complementary contextual information from their non-shadow counterparts, limiting their further shadow detection performance.

Instead, in this paper, we explore and exploit the mutual complementary mechanisms contained in this specific task for shadow detection. We first notice the essential complementarity between the shadow regions and their non-shadow counterparts. Therefore, we propose a novel shadow detection framework, which consists of two interactive branches to concurrently estimate the shadow and non-shadow masks. In order to collaborate and boost these two complementary parts, we introduce the negative activation technique and the identity reconstruction constraint to fully mine the complementarity at the feature and image level. For the features interaction, we retain the deactivated intermediate features of one branch and offer these features as additional supplementary information to the other branch, and vice versa. At the image level, the results of two branches should always sum to a constant at each corresponding position, and we impose this identity constraint as the guidance to optimize our framework. With the complementary auxiliary information from each other's branches, our network could embrace interesting merits: while improving the performance of shadow detection, it also obviously improves the confidence of predicted results.

In addition, we further explore the complementary requirements of shadow detection: not to miss all shadow areas, as well as not to mistake any non-shadow counterparts. Moreover, we transfer these complementary requirements as two complementary discriminative losses for our framework. To be specific, we introduce the inner and outer discriminators to encourage the shadow detection branch to predict the accurate shadow locations in an adversarial manner. Moreover, we further employ the dilation and erosion operation on the ground truth masks to generate pseudo masks for benefiting the discriminating capability of two discriminators. Finally, our framework could confidently predict more accurate shadow detection results by fully exploring and exploiting the complementary mechanism of shadow detection.

In summary, the contributions of this paper are as follows:

- We propose a novel shadow detection framework, in which the mutual complementary information contained in this specific task is explored and exploited, *e.g.*, the complementarity between the shadow regions and non-shadow counterparts, and the complementary requirements of shadow detection.
- We observe that the shadow and surrounding non-shadow regions are interrelated and complementary. Hence, we elaborately design dual interactive branches to cooperatively offer complementary auxiliary information and supervision for each other at the feature and image-level via introducing the negative activation technique and identical reconstruction loss.
- We investigate the complementary requirements of shadow detection: neither missing any shadow regions nor falsely detecting non-shadow regions. Moreover, we further devise the complementary discriminative constraints derived from the above complementary requirements to boost the performance of shadow detection.
- Extensive experiments indicate that our proposed method could achieve superior shadow detection performance both quantitatively and qualitatively with relatively smaller parameters and faster inference speed.

2 RELATED WORK

In this paper, we mainly focus on shadow detection from a single shadow image. The corresponding studies can be roughly divided into two groups: traditional and Convolutional Neural Networks (CNNs)-based methods.

2.1 Traditional Methods

Early methods focus on exploring various hints for shadow detection, *e.g.*, illumination models or color priors, hand-crafted image features, edges, and textures. Zhu *et al.*[50] attempt to detect shadows in real-world scenes where color information is unavailable based on the usual characteristics of shadows. Vicente *et al.*[37] identify shadows via incorporating the learned appearance and contextual cues of shadows. However, These strategies could produce relatively accurate detection results, but their performances will drop significantly when dealing with complex shadow scenes where the handcrafted features are far from enough to discern the shadows regions.

2.2 CNNs-based Methods

Thanks to the remarkable success of CNNs in various computer vision tasks, CNNs-based shadow detection methods are capable of easily identifying shadow context from the public shadow datasets. Therefore, CNNs-based shadow detection methods have been far exceeding the performance of previous traditional methods. Khan *et al.*[20] firstly propose a CNNs-based framework to automatically detect shadows by building a 7-layer network structure. Shen *et al.*[32] constructs CNNs to exploit the detected shadow edges and efficient least-square optimization for detecting shadow regions.

Recently, many works tend to design more efficient and effective feature extraction modules to improve the network's ability to understand shadow scenes, thereby further enhancing shadow detection performance. For example, Zhu *et al.*[51] formulate the recurrent attention residual module and construct the bidirectional feature pyramid network to explore the global and local shadow contexts. Hu *et al.*[15] present a direction-aware spatial module to aggregate the contextual features for better detecting shadows. DSD [47] proposes a distraction-aware module to explicitly consider various ambiguous cases for improving the performance of shadow detection. Fang *et al.*[2] explore an effective context augmentation with the parallel multi-scale convolution operations for robust shadow detections.

There also exist studies that attempt to address shadow detection from other perspectives. RCMPNet [27] develops a designed ensemble model to predict corresponding confidence maps of previous methods' results, causing their performance to be greatly dependent on the performance of the previous methods. ADNet [26] employs the shadow attenuation network to produce more adversarial training examples for their shadow detection network.

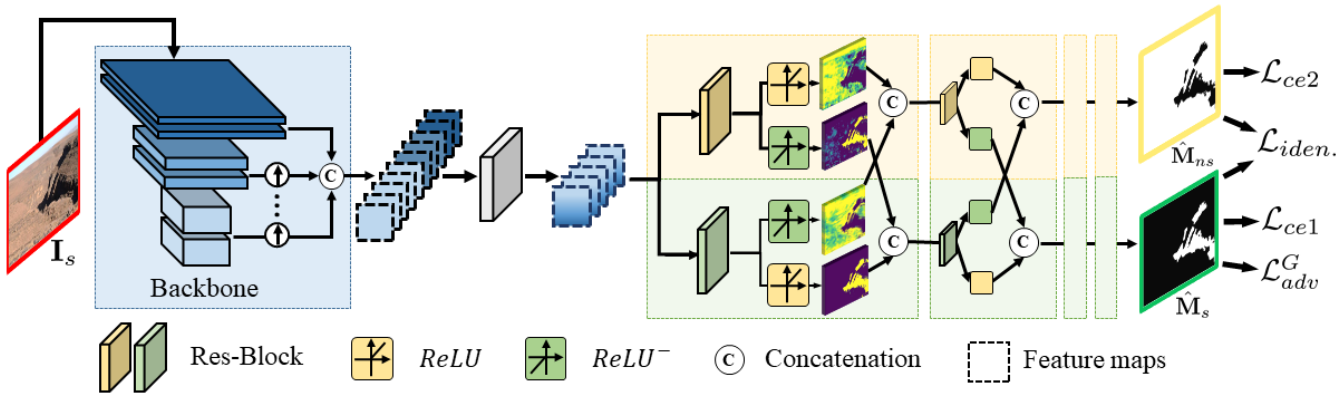

Figure 1: Illustration of generator G of our proposed shadow detection framework, which consists of one parameters-shared encoder for extracting backbone features, two interactive decoders as detection heads for predicting $\hat{\mathbf{M}}_s$ and $\hat{\mathbf{M}}_{ns}$. Moreover, we provide visualizations of interactive features in the above pipeline.

Hu *et al.*[14] tend to resolve the shadow detection for the general real-world shadow scenes.

In the era of deep learning, CNNs-based methods also explore the potential cues for shadow detection, *e.g.*, combining with multi-task learning [1, 38], intensity bias [52], ensemble techniques [27]. However, few methods notice the complementarity between the shadow regions and their non-shadow counterparts for this specific task. To the best of our knowledge, we are the first to explore and exploit this complementary mechanism to boost the performance of shadow detection. Moreover, previous methods also make use of Generative Adversarial Networks (GANs) to address the shadow detection problem. However, the purpose of using GANs is different from ours. scGAN [28] introduces a tunable sensitivity parameter to overcome the inflexibility of GANs in the shadow detection task. ST-CGAN [38] takes advantage of GANs to obtain the high-level semantics and global scene characteristics for shadow detection. Unlike theirs, we utilize the tailored GANs to implement the complementary requirements of shadow detection, *e.g.*, the specific inputs after erosion or dilation operation for discriminators.

3 METHOD

We present our proposed shadow detection framework in Figure 1. In this section, we would like to further describe the motivation for designing the above framework and the details of our framework.

3.1 Motivation

Here we illustrate the motivation behind the effective complementary mechanisms utilized in our framework. The first complementarity is that we observe that the shadow regions and non-shadow counterparts are mutually interrelated and complementary in nature. The shadow mask $\hat{\mathbf{M}}_s$ and non-shadow counterpart $\hat{\mathbf{M}}_{ns}$ themselves should satisfy the identity constraint relationship: $\mathbb{1} = \hat{\mathbf{M}}_s + \hat{\mathbf{M}}_{ns}$. The design of dual detection branches could collaborate these two estimations by reducing the output of one branch from the original input as complementary auxiliary information for the other branch. However, previous methods only predict shadow masks, which naturally ignores such auxiliary information. Furthermore,

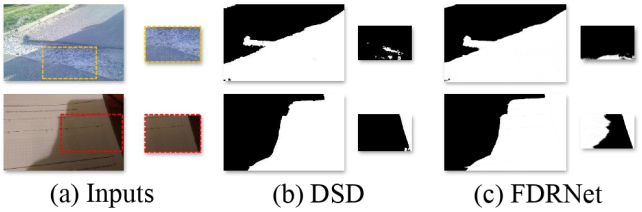

Figure 2: Visual results of DSD [47] and FDRNet [52] with the different inputs. The inputs are the original images and the cropped patches that contain only the shadow regions. When the cropped patches are used as inputs, the shadow detection performance of the previous methods drops drastically, indicating the surrounding non-shadow information is crucial for shadow detection.

inspired by the fact that the information of the surrounding non-shadow regions as references is essential for existing methods to realize the judgment of the shadow location. As shown in Figure 2, although the existing methods could predict the shadow masks well in the images with shadows and non-shadow regions. However, the cropped patches from original inputs only contain shadow regions, and the performance of these methods drops significantly. This also indicates that the assistance of the surrounding non-shadow regions is also required and important for shadow detection.

Therefore, we explicitly explore and exploit the complementary relationship between the non-shadow and shadow regions to achieve better performance. We develop a novel shadow detection framework, which jointly generates the shadow mask $\hat{\mathbf{M}}_s$ and their non-shadow counterpart $\hat{\mathbf{M}}_{ns}$ via dual interactive branches. Then, we utilize the complementary information through the interaction of intermediate features between two branches and the identity supervision for the final outputs.

Except for the complementary mechanism between the shadow and non-shadow regions, we also explore and exploit the complementary requirements of shadow detection: neither missing any shadow regions nor falsely detecting non-shadow regions. These

two requirements are obviously complementary, and only by meeting these two requirements, accurate shadow detection results could be obtained. Hence, motivated by the above complementary requirements, we devise two complementary discriminative losses imposed on the estimated results of the shadow detection branch and implement them with two complementary discriminators (D_{inner} & D_{outer}). Utilizing D_{inner} imposes the inner discriminative constraint to reduce the detection network to miss any ambiguous shadow regions, *e.g.*, shadow regions like non-shadow patterns. Meanwhile, utilizing D_{outer} imposes the outer discriminative constraint to avoid falsely detecting non-shadow regions like shadow patterns. Note that similar requirements could also be applied to the non-shadow detection branch, but due to the limited memory of our device, we only utilize the complementary requirements for the shadow detection branch. Extensive experiments indicate that this complementarity also brings obvious improvement.

3.2 The Generator G

Our Generator G is in line with the classical encoder-decoder structure. However, unlike the previous methods, our generator G includes one parameters-shared encoder for extracting the backbone features, two interactive decoders as detection heads for predicting $\hat{\mathbf{M}}_s$ and $\hat{\mathbf{M}}_{ns}$ given the input shadow images $\mathbf{I}_s \in [0,1]^{C\times W\times H}$. C, H and W represent the channel, height and width of the original images, respectively. Next, we provide detailed descriptions of the implementation architecture of our framework.

Backbone Encoder. Following [52], we also employ the lightweight EfficientNet-B3 [33] as our backbone network to extract the hierarchical features, namely $\mathbf{F}_i$ (i = 1, 2, $\cdots$, m). After that, these features with different scales are aggregated as encoder features with the up-sampling and concatenation operations. Finally, in order to reduce the dimensions of encoder features $\mathbf{F}_E$ and the computation costs, we apply the 1×1 point-wise convolutional layers to reduce their channels to 32. The aforementioned operations can be defined as

$$\begin{aligned}\mathbf{F}_E &= Concat([\mathbf{F}_0, \mathbf{F}_1^{\uparrow}, ..., \mathbf{F}_m^{\uparrow}]),\\ \mathbf{F}_E &= Convs_{1\times1}(\mathbf{F}_E),\end{aligned} \tag{1}$$

where $^{\uparrow}$ denotes the bilinear up-sampling operation.

Dual Interactive Decoders. Dual decoders share the same architecture, consisting of several Residual Blocks [10] as our basic blocks. Each basic block adopts ReLU [6] as the activation function after the convolution, which is one of the widely-used activation functions in current network architectures. Intuitively, in the decoder for predicting shadow masks, the activation function is devised to highlight (activate) the desired shadow regions under supervised learning, and vice versa. Because of the complementarity between the shadow and non-shadow counterparts, the deactivated features of the shadow detection decoder could be delivered to the non-shadow detection decoder. On the contrary, the deactivated features of the non-shadow detection decoder could be delivered to assist the shadow detection. Hence, we introduce the Negative Activation Technique (NAT) [12] to retain the deactivated features to achieve the interaction between two decoders at the feature level. Assuming that $\mathbf{F}$ is the intermediate feature, then the NAT operation of the ReLU function can be defined as

$$\mathbf{F}^{-} = ReLU^{-}(\mathbf{F}) = \mathbf{F} - ReLU(\mathbf{F}) = min\{\mathbf{F}, \mathbf{0}\}. \tag{2}$$

As shown in Figure 1, we conduct the interaction of the two branches after the basic block of each branch. Here, we utilize $\mathbf{F}_l^i$ ($i \in 1, 2$ for two decoders) to represent outputs of the l-th basic block, and $\mathbf{F}_{l+1}^i$ to represent the inputs of the $(l+1)$-th basic block to illustrate the feature interaction process between two decoders as follows:

$$\begin{aligned}\mathbf{F}_1^{l+1} &= Concat([ReLU(\mathbf{F}_1^l), ReLU^{-}(\mathbf{F}_2^l)]),\\ \mathbf{F}_2^{l+1} &= Concat([ReLU(\mathbf{F}_2^l), ReLU^{-}(\mathbf{F}_1^l)]).\end{aligned} \tag{3}$$

Finally, we optimize our generator G with various objective functions. Following [47, 52], we adopt the weighted binary cross-entropy loss for the results of the shadow detection branch, which is defined as

$$\begin{aligned}\mathcal{L}_{ce1}(\hat{\mathbf{M}}_s, \mathbf{M}_s) = -\sum_j (&\frac{N_n}{N} M_s^{(j)} \log(\hat{\mathbf{M}}_s^{(j)}) +\\ &\frac{N_{ns}}{N}(1 - M_s^{(j)} \log(1 - \hat{\mathbf{M}}_s^{(j)}))),\end{aligned} \tag{4}$$

where $\mathbf{M}_s$ denotes the ground truth shadow masks; j denotes the pixel index along the spatial dimension. N_s, N_{ns}, and N refer to the number of pixels in the shadow regions, the number of pixels in the non-shadow regions, and the number of pixels in the entire image, respectively. Similarly, the loss of the non-shadow detection branch, which is defined as

$$\begin{aligned}\mathcal{L}_{ce2}(\hat{\mathbf{M}}_{ns}, \mathbf{M}_{ns}) = -\sum_j [&\frac{N_{ns}}{N} M_{ns}^{(j)} \log(\hat{\mathbf{M}}_{ns}^{(j)}) +\\ &\frac{N_s}{N}(1 - M_{ns}^{(j)} \log(1 - \hat{\mathbf{M}}_{ns}^{(j)}))],\end{aligned} \tag{5}$$

where $\mathbf{M}_{ns}$ denotes the ground truth non-shadow masks; j denotes the pixel index along the spatial dimension. Moreover, the estimated results $\hat{\mathbf{M}}_s$ and $\hat{\mathbf{M}}_{ns}$ should satisfy the identity constraint as follows:

$$\mathcal{L}_{iden.}(\hat{\mathbf{M}}_s, \hat{\mathbf{M}}_{ns}, \mathbb{1}) = \|\hat{\mathbf{M}}_s + \hat{\mathbf{M}}_{ns} - \mathbb{1}\|_1, \tag{6}$$

where $\mathbb{1}$ denotes an all-ones matrix of the same size as $\hat{\mathbf{M}}_s$ and $\hat{\mathbf{M}}_{ns}$. Besides, we also impose adversarial losses, which are defined as

$$\begin{aligned}\mathcal{L}_{adv}^{G} = &\mathbb{E}_{(\mathbf{I}_s, \hat{\mathbf{M}}_s)\sim P_{data}(\mathbf{I}_s, \hat{\mathbf{M}}_s)}[\log(D_{inner}(\mathbf{I}_s, \hat{\mathbf{M}}_s))] +\\ &\mathbb{E}_{(\mathbf{I}_s, \hat{\mathbf{M}}_s)\sim P_{data}(\mathbf{I}_s, \hat{\mathbf{M}}_s)}[\log(D_{outer}(\mathbf{I}_s, \hat{\mathbf{M}}_s))],\end{aligned} \tag{7}$$

where $\mathbf{I}_s$ is the input shadow image, acting as the conditional input in the discriminators. Therefore, the total loss of our generator G is a weighted sum of the predefined losses:

$$\mathcal{L}_G = \mathcal{L}_{ce1} + \lambda_{ce2}\mathcal{L}_{ce2} + \lambda_{iden.}\mathcal{L}_{iden.} + \lambda_{adv}\mathcal{L}_{adv}^{G}, \tag{8}$$

where λ_{ce2}, $\lambda_{id.}$, and λ_{adv} indicate the weight factors.

3.3 Dual Complementary Discriminators

These two complementary discriminators have the same network architecture, but different functions. The architectures for D_{inner} and D_{outer} refer to [38, 46]. During the training phase, we refer to the Generative Adversarial Networks (GANs) [7] to update the shadow detection network and discriminators in an alternately iterative manner.

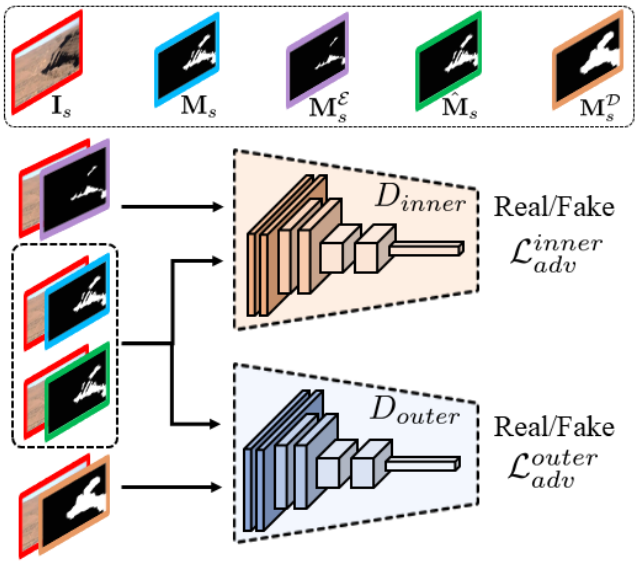

Figure 3: Illustration of our proposed Dual Discriminators, which are designed to achieve the complementary requirements of shadow detection.

Inner Discriminator. The inner discriminator D_{inner} is trained to distinguish whether the detection shadow results miss any shadow regions. Meanwhile, in order to fool D_{inner}, our generator G has to detect the shadow masks $\hat{\mathbf{M}}_s$ covering possible shadow regions. Therefore, we achieve the first requirement in an adversarial fashion. Moreover, we employ the image erosion operation [3, 17] on the ground truth masks to obtain pseudo masks $\mathbf{M}_s^{\mathcal{E}}$ for enhancing the discriminating capability of D_{inner}. Note that dilation and erosion are basic image morphological processing operations. The dilation operation often utilizes a structuring element for probing and expanding the shapes contained in the input image. And the erosion operation often utilizes a structuring element for probing and reducing the shapes contained in the input image. Therefore, the eroded masks naturally miss partial shadow regions, and the dilated masks naturally include some extra non-shadow regions, which could help the inner discriminator to distinguish whether the detection shadow results miss any shadow regions. Therefore, the corresponding adversarial constraint to optimize D_{inner} can be defined as

$$\begin{aligned}\mathcal{L}_{adv}^{inner} = &\mathbb{E}_{(\mathbf{I}_s,\mathbf{M}_s)\sim P_{data}(\mathbf{I}_s,\mathbf{M}_s)}[\log(D_{inner}(\mathbf{I}_s,\mathbf{M}_s))]+\\ &\mathbb{E}_{(\mathbf{I}_s,\mathbf{M}_s^{\mathcal{E}})\sim P_{data}(\mathbf{I}_s,\mathbf{M}_s^{\mathcal{E}})}[\log(1-D_{inner}(\mathbf{I}_s,\mathbf{M}_s^{\mathcal{E}}))]+\\ &\mathbb{E}_{(\mathbf{I}_s,\hat{\mathbf{M}}_s)\sim P_{data}(\mathbf{I}_s,\hat{\mathbf{M}}_s)}[\log(1-D_{inner}(\mathbf{I}_s,\hat{\mathbf{M}}_s))].\end{aligned} \tag{9}$$

Outer Discriminator. The outer discriminator D_{outer} is trained to distinguish whether the detection shadow branch falsely detects any non-shadow regions as shadows. Meanwhile, in order to fool D_{outer}, our generator G has to detect the shadow masks $\hat{\mathbf{M}}_s$ and avoid detecting possible non-shadow regions. Similarly, we employ the image dilation operation on the ground truth masks to obtain pseudo masks $\mathbf{M}_s^{\mathcal{D}}$. The dilated masks naturally include some extra non-shadow regions, which could help the outer discriminator to distinguish whether the shadow results falsely contain any non-shadow regions. The adversarial constraint of D_{outer} can be defined as

$$\begin{aligned}\mathcal{L}_{adv}^{outer} = &\mathbb{E}_{(\mathbf{I}_s,\mathbf{M}_s)\sim P_{data}(\mathbf{I}_s,\mathbf{M}_s)}[\log(D_{outer}(\mathbf{I}_s,\mathbf{M}_s))]+\\ &\mathbb{E}_{(\mathbf{I}_s,\mathbf{M}_s^{\mathcal{D}})\sim P_{data}(\mathbf{I}_s,\mathbf{M}_s^{\mathcal{D}})}[\log(1-D_{outer}(\mathbf{I}_s,\mathbf{M}_s^{\mathcal{D}}))]+\\ &\mathbb{E}_{(\mathbf{I}_s,\hat{\mathbf{M}}_s)\sim P_{data}(\mathbf{I}_s,\hat{\mathbf{M}}_s)}[\log(1-D_{outer}(\mathbf{I}_s,\hat{\mathbf{M}}_s))].\end{aligned} \tag{10}$$

Finally, with the help of D_{inner} and D_{outer}, we could naturally achieve the complementary requirements of shadow detection. Furthermore, the generator G has an incentive to produce a more accurate shadow mask under the designed discriminative constraints.

3.4 Experiments

Implementation Details. We implement our proposed shadow detection framework via the PyTorch platform on the PC with the RTX 1080Ti GPU. For training, the training images are resized to the fixed resolution of 416×416 and applied random flipping as the augmentation strategy. The proposed framework is optimized for 30 epochs by the Adamax [21] optimizer with a fixed learning rate of $1e-3$. The minimum training batch size is 4. The whole framework takes about 2 hours on the SBU and 1 hour on the ISTD dataset. For the hyperparameters, the weight factors (λ_{ce2}, $\lambda_{iden.}$, and λ_{adv}) in Equation (8) are empirically set as 1, $1e-4$ and $1e-2$. Following [1, 26, 27, 27, 52], we also apply the fully connected CRF operation [22] to refine the estimated results in the inference phase.

Dataset. We conduct shadow detection experiments on the three representative benchmark datasets, *i.e.*, SBU [36], UCF [50], and ISTD [38] dataset. SBU dataset includes 4089 and 638 pairs of images for training and testing. ISTD dataset includes 1870 image triples (shadow images, shadow-free images, and shadow masks). This dataset has been separated into 1330 and 540 triplets for the training and testing. Following the same experiment setting with previous methods [1, 26, 27, 27, 52], we train our framework on the SBU training dataset, and test on the SBU and UCF testing dataset. The testing for the ISTD dataset is utilizing the model trained on the corresponding training dataset.

Evaluation Metric. Following previous methods [1, 27, 52], we adopt the widely-used metric, balance error rate (BER) [35], to evaluate the quantitative performance. BER is defined as

$$BER = (1 - \frac{1}{2}(\frac{TP}{TP+FN}) + \frac{TN}{TN+FP})) \times 100, \tag{11}$$

where TP, TN, FP, and FN stand for the number of True Positives, True Negatives, False Positives, and False Negatives, respectively. Lower BER values denote better results.

3.5 Detection Evaluation on Benchmarks

In Table 1, we report the comparison results with recent state-of-the-art (SOTA) methods on the three benchmarks, including one traditional method: Unary-Pariwise [9] and 13 CNNs-based methods: FDRNet [52], RCMPNet [27], ECA [2], MTMT [1], DSD [47], DC-DSPF [43], ADNet [26], DSC [15], BDRAR [51], ST-CGAN [38], patched-CNN [11], scGAN [28], and stacked-CNN [36]. For fair comparisons with these SOTA methods, we utilize the provided pre-trained model or results from authors to obtain the quantitative

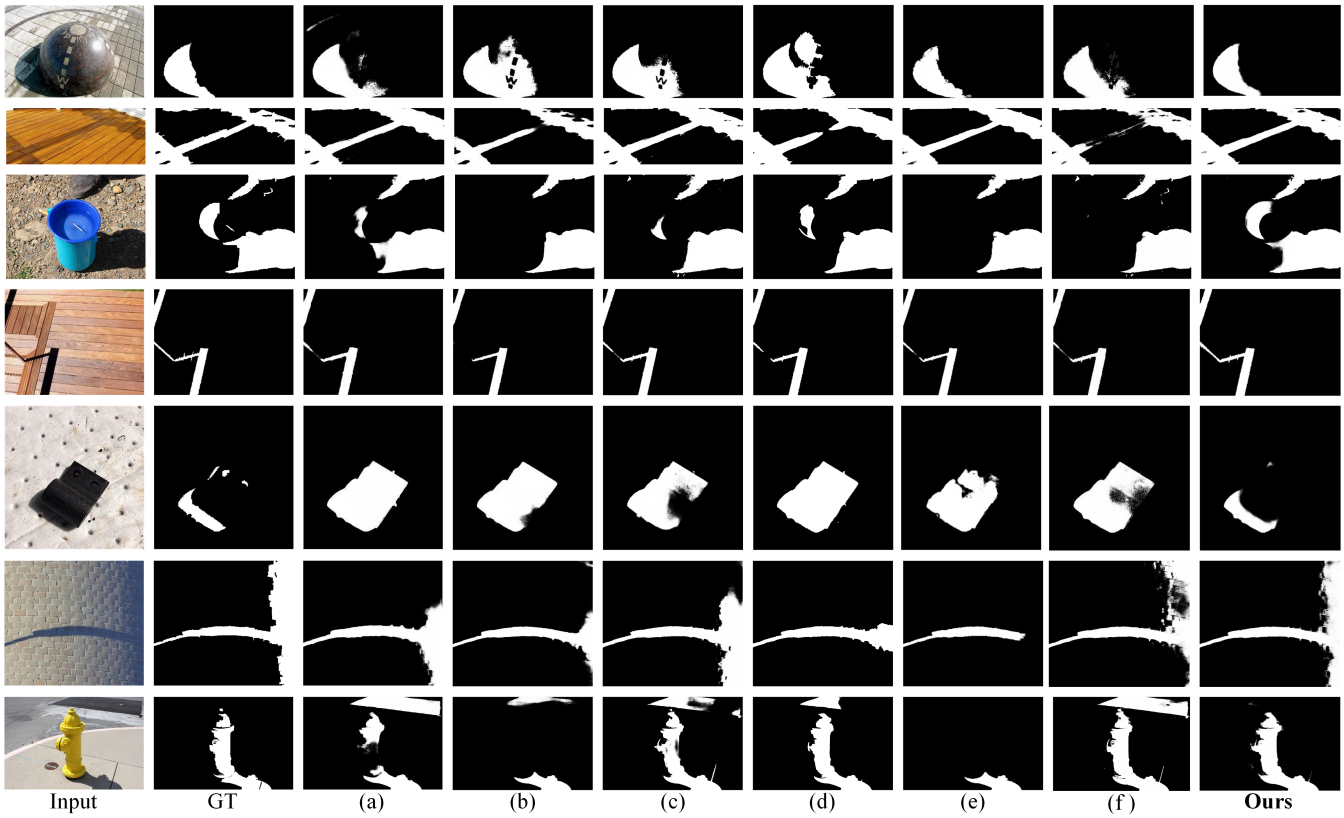

Figure 4: The visual comparison results of different methods on the real shadow scenarios. (a) to (f) are the detection results from state-of-the-art methods: BDRAR [51], DSC [13], DSD [47], MTMT [1], ECA [2], and FDRNet [52], respectively.

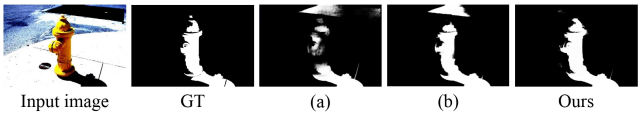

Figure 5: Visual comparison results of models with different configurations. (a) the result of Model-1; (b) the result of Model-5.

results. For example, we employ the publicly available detection results provided by authors of ECA [2], not the metric values of their paper. Besides, we also compare our results with four SOTA saliency object detection methods referring to FDRNet [52]. They were retrained and tested on the shadow datasets, and the corresponding evaluation results of these methods can be found in [52]. Obviously, our method performs the best BER scores over SOTA methods on the three benchmark datasets. Compared to RCMPNet, which is the second best-performing method, our method successfully reduces the BER score by 1.3% on the SBU testing dataset. Note that RCMPNet is an ensemble model to predict shadow masks based on the previous three SOTA methods, which largely demonstrates the effectiveness of our proposed shadow detection framework.

In addition, we also provide the visual comparison results with different SOTA methods in Figure 4. Obviously, our method performs better than the existing shadow detection results. The shadows in the image in the fifth and last row are very similar to the appearance of the black object, only our method successfully detects the shadows without misjudging. FDRNet still misses partial shadow regions (*e.g.*, first three rows), and there also exist cases where the non-shadow regions are mistaken. In contrast, our method could successfully detect more accurate shadow masks with more fine-grained boundaries. This also indicates the effectiveness of complementary mechanisms adopted in our method.

3.6 Ablation Study

Analysis of the effectiveness of the complementarity between the shadow and non-shadow regions. We compare five models with the different configurations: (1) Model-1: only utilizing a single decoder to predict shadow masks; (2) Model-2: expanding the number of channels in the decoder of Model-1 to twice the original setting; (3) Model-3: utilizing two non-interactive decoders to predict both shadow and non-shadow masks; (4) Model-5: utilizing two interactive decoders with the negative activation technique to predict both shadow and non-shadow masks; (5) Model-6: $\mathcal{L}_{iden.}$ is added based on the Model-3. In the absence of branch interactions, the performance of Model-1 and Model-3 is comparable. Moreover,

Table 1: Quantitative comparison of our method with the SOTA methods on the three public benchmark datasets for shadow detection. For each testing dataset, we provide the BER metric values. The best results are in bold, and ↓ indicates lower is better.

		SBU	UCF	ISTD
Method	Year	BER↓	BER↓	BER↓
Ours	-	**2.94**	6.73	1.44
Ours-w/o-CRF	-	3.02	**6.69**	**1.41**
FDRNet [52]	2021	3.04	7.28	1.55
RCMPNet [27]	2021	2.98	6.75	1.61
ECA [2]	2021	5.93	10.71	2.03
MTMT [1]	2020	3.15	7.47	1.72
DSD [47]	2019	3.45	7.59	2.17
DC-DSPF [41]	2018	4.90	7.90	-
ADNet [26]	2018	5.37	9.25	-
DSC [15]	2018	5.59	10.54	3.42
BDRAR [51]	2018	3.64	7.81	2.69
ST-CGAN [38]	2018	8.14	11.23	3.85
patched-CNN [11]	2018	11.56	-	-
scGAN [28]	2017	9.10	11.50	4.70
stacked-CNN [36]	2016	11.00	13.00	8.60
Unary-Pariwise [9]	2011	25.03	-	-
ITSD [49]	2020	5.00	10.16	2.73
EGNet [45]	2019	4.49	9.20	1.85
SRM [39]	2017	6.57	12.51	7.92
Amulet [44]	2017	6.57	12.51	7.92

with the complementary interaction between two decoders, the detection performance obviously successfully improves 5.1% on the SBU testing dataset, as reported in Table 2. The detection performance obviously successfully improves by 0.11 on the SBU testing dataset, which indicates that only increasing the network parameters is not as effective as interactive strategies. Compared with Model-3 and Model-4, $\mathcal{L}_{iden.}$ also brings a certain improvement in shadow detection performance. In Figure 5, we also provide the visual results of different models to verify the effectiveness of the proposed complementarity.

Analysis of the Confidence of Predictions. We employ entropy to measure the confidence of predicted results. A lower entropy value denotes higher confidence. However, utilizing the entropy is to measure the confidence of the prediction and not to measure the accuracy of the prediction. We only compared the entropy performance with FDRNet [52], which has the closest comparable performance to our method. Note that codes of RCMPNet [27] are not publicly available. As shown in Figure 6, our method could confidently predict more accurate results. Moreover, we conduct the statistics of predictions on the SBU testing dataset among three models: FDRNet [52], our framework with only a decoder for detecting shadow regions, and our default framework. Obviously,

Table 2: Ablation study of the components used in our framework on the SBU and ISTD testing dataset. SD: Single Decoder; DD: Dual Decoder; DC: Direct Concatenation interaction strategy; NAT: Negative Activation Technique interaction strategy.

Models	Setting	Metric (BER ↓)	
		SBU	ISTD
Model-1	SD	3.33	1.72
Model-2	SD w double channels	3.25	1.64
Model-3	DD w/o interaction	3.31	1.69
Model-4	DD + DC	3.19	1.65
Model-5	DD + NAT	3.14	1.61
Model-6	DD + NAT+ $\mathcal{L}_{iden.}$	3.10	1.56
Model-7	DD + NAT+ $\mathcal{L}_{iden.}$ + $\mathcal{L}_{adv}^{inner}$	2.99	1.47
Model-8	DD + NAT+ $\mathcal{L}_{iden.}$ + $\mathcal{L}_{adv}^{outer}$	3.01	1.48
Ours	DD + NAT+ $\mathcal{L}_{iden.}$ + $\mathcal{L}_{adv}^{inner}$ + $\mathcal{L}_{adv}^{outer}$	2.94	1.44

the mean entropy values of our method are less than half that of FDRNet. Meanwhile, the statistical results also demonstrate that the confidence of the prediction results could be improved under the interaction of two decoders.

Analysis of the Interaction Strategy. In our default framework, we introduce the negative activation technique (NAT) as the interaction strategy between two decoders. Here, we further investigate another straightforward manner to achieve the interaction between features of two decoders, which is to concatenate features along the channel dimension. From the results reported in Table 2, we find that adopting the NAT performs better than the straightforward concatenation manner. In our framework, NAT retains the deactivated features of one branch and delivers them to another complementary branch, and vice versa. This interaction strategy allows one branch to utilize the deactivated features of the other branch, thereby achieving complementarity at the feature level. On the contrary, these deactivated features are abandoned in the direct concatenation strategy, leading to suboptimal performance.

Analysis of the effectiveness of the complementary requirements of shadow detection. The complementary requirements of shadow detection are implemented by two complementary discriminators. Therefore, we conduct experiments to verify the effects of these two discriminators D_{inner} and D_{outer}. In Table 2, we provide the shadow detection performance when we remove the D_{inner} and D_{outer} and corresponding discriminative losses. With the help of the complementary discriminators, the BER value is almost reduced by 0.16 on the SBU dataset.

Furthermore, we take the inner discriminator D_{inner} to investigate its corresponding discriminative capability. D_{inner} is designed to distinguish whether the detection shadow results miss any shadow regions to meet the requirement of shadow detection. As shown in Figure 7, we feed D_{inner} three types of shadow masks: $\mathbf{M}_s$, $\mathbf{M}_s^{\mathcal{D}}$, and $\mathbf{M}_s^{\mathcal{E}}$. We find D_{inner} could distinguish the result misses partial shadow regions based on the visualizations of CAM [48]. Regardless of whether the input masks miss the shadow area or not, the feature responses of D_{inner} appear to be quite different.

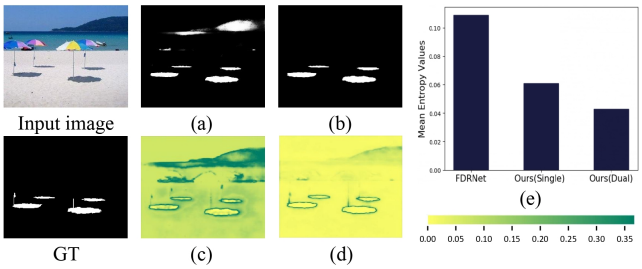

Figure 6: Shadow detection results with the corresponding entropy maps, which demonstrate our method could confidently predict more accurate shadow locations. (a) and (b) are the estimated results from FDRNet [52] and ours; (c) and (d) are the corresponding entropy map of (a) and (b); (e) is the mean entropy map statistics on SBU dataset [36] (from left to right are: mean entropy values of FDRNet, our framework with only single decoder for shadow detection, and our default framework).

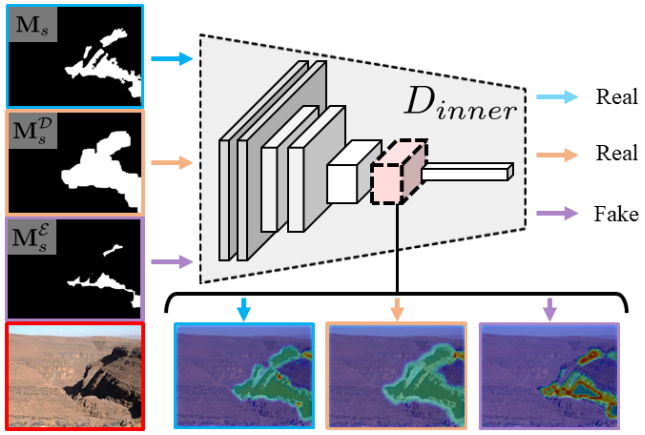

Figure 7: Feature map visualization of the discriminator D_{inner} with different masks as inputs using class activation mapping (CAM) [48]. The different inputs are: the GT shadow mask M_s, the dilated GT shadow mask $M_s^{\mathcal{D}}$, and the eroded GT shadow mask $M_s^{\mathcal{E}}$. Note that $M_s^{\mathcal{D}}$ never participated in the training of D_{inner}, but the visualization result is consistent with the result of M_s, which means that D_{inner} could better distinguish whether missing shadow regions.

Analysis of the Number of Features Interactions, As reported in Table 3, we conduct experiments to verify the effects of the number of interactions between the decoders. Because we conduct the interactions of the two decoders after the basic block of each branch, the number of interactions is also the same as the number of Resblocks [10] adopted in our framework. It reaches the lowest BER value when the number of interactions is 4. Therefore, we empirically set the number of interactions to 4 as the default setting.

Analysis of the Parameters and Inference Time. To verify the effectiveness and efficiency of our method, we also compare the parameters and average inference times. For the inference time, we repeatedly ran different SOTA models on images with a resolution of 416 ×416 100 times to obtain the average inference time. Since the ensemble strategy of RCMPNet needs to be implemented based on the results of the previous shadow detection methods (MTMT, DSDNet, and BARAR), it often requires a large computational cost. However, the source codes of RCMPNet are not publicly available, we provide comparisons of network parameters and the average inference time with these previous methods in Table 4. Obviously, our method is almost 8 times faster in terms of inference time compared with RCMPNet.

Table 3: Ablation study of numbers of features interaction on the SBU dataset.

Interaction Numbers	1	2	3	4	5	6
BER ↓	3.06	3.03	2.98	2.94	2.96	2.99

Table 4: The network parameters and average inference time of different shadow detection methods. The average inference times are obtained with the resolution of 416 × 416 on the 1080Ti GPU device.

Methods	Parameters(M: 10^6)	Average Inference Time (ms)
BADRA [51]	42.46	96.73
DSD [47]	58.16	71.32
MTMT [1]	44.13	57.08
RCMPNet [27]	-	>225.13 (96.73 + 71.32 + 57.08)
ECA [2]	157.76	92.82
FDRNet [52]	10.77	29.74
Ours	10.95	28.31

4 CONCLUSION

In our paper, we develop a novel framework for shadow detection, which investigates and exploits the complementary mechanisms contained in this specific task. Therefore, our framework consists of two key components that correspond to the two investigated complementarities, including the mutual complementarity between the shadow regions and their non-shadow counterparts, and the complementary requirements for shadow detection. Furthermore, we conduct comprehensive experiments and visualizations to demonstrate the effectiveness of two explored complementary mechanisms. Moreover, our method achieves superior performance against the existing state-of-the-art shadow detection methods with faster inference speed and smaller network parameters.

5 ACKNOWLEDGMENTS

This work was supported by the National Key R&D Program of China under Grant 2020AAA0105702, the National Natural Science Foundation of China (NSFC) under Grants U19B2038 and 61901433, the University Synergy Innovation Program of Anhui Province under Grants GXXT-2019-025, the Fundamental Research Funds for the Central Universities under Grant WK2100000024, and the USTC Research Funds of the Double First-Class Initiative under Grant YD2100002003.

REFERENCES

[1] Zhihao Chen, Lei Zhu, Liang Wan, Song Wang, Wei Feng, and Pheng-Ann Heng. 2020. A multi-task mean teacher for semi-supervised shadow detection. In *Proceedings of the IEEE/CVF Conference on computer vision and pattern recognition*. 5611–5620.

[2] Xianyong Fang, Xiaohao He, Linbo Wang, and Jianbing Shen. 2021. Robust Shadow Detection by Exploring Effective Shadow Contexts. In *Proceedings of the 29th ACM International Conference on Multimedia - ACM Multimedia 2021*.

[3] Zunlei Feng, Lechao Cheng, Xinchao Wang, Xiang Wang, Ya Jie Liu, Xiangtong Du, and Mingli Song. 2021. Visual Boundary Knowledge Translation for Foreground Segmentation. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 35. 1334–1342.

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

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

[6] Xavier Glorot, Antoine Bordes, and Yoshua Bengio. 2011. Deep sparse rectifier neural networks. In *Proceedings of the fourteenth international conference on artificial intelligence and statistics*. JMLR Workshop and Conference Proceedings, 315–323.

[7] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. 2014. Generative adversarial nets. *Advances in neural information processing systems* (2014).

[8] Ye-Peng Guan. 2008. Wavelet multi-scale transform based foreground segmentation and shadow elimination. *The Open Signal Processing Journal* 1, 1 (2008).

[9] Ruiqi Guo, Qieyun Dai, and Derek Hoiem. 2011. Single-image shadow detection and removal using paired regions. In *CVPR 2011*. IEEE, 2033–2040.

[10] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. 2016. Deep residual learning for image recognition. In *Proceedings of the IEEE/CVF Conference on computer vision and pattern recognition*.

[11] Sepideh Hosseinzadeh, Moein Shakeri, and Hong Zhang. 2018. Fast shadow detection from a single image using a patched convolutional neural network. In *2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 3124–3129.

[12] Qiming Hu and Xiaojie Guo. 2021. Trash or Treasure? An Interactive Dual-Stream Strategy for Single Image Reflection Separation. *Advances in Neural Information Processing Systems* 34 (2021).

[13] Xiaowei Hu, Chi-Wing Fu, Lei Zhu, Jing Qin, and Pheng-Ann Heng. 2019. Direction-aware spatial context features for shadow detection and removal. *IEEE transactions on pattern analysis and machine intelligence* 42, 11 (2019), 2795–2808.

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

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

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

[17] Paul T. Jackway and Mohamed Deriche. 1996. Scale-space properties of the multiscale morphological dilation-erosion. *IEEE transactions on pattern analysis and machine intelligence* 18, 1 (1996), 38–51.

[18] Imran N Junejo and Hassan Foroosh. 2008. Estimating geo-temporal location of stationary cameras using shadow trajectories. In *European conference on computer vision*. Springer, 318–331.

[19] Kevin Karsch, Varsha Hedau, David Forsyth, and Derek Hoiem. 2011. Rendering synthetic objects into legacy photographs. *ACM Transactions on Graphics (TOG)* 30, 6 (2011), 1–12.

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

[21] Diederik P Kingma and Jimmy Ba. 2014. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980* (2014).

[22] Philipp Krähenbühl and Vladlen Koltun. 2011. Efficient inference in fully connected crfs with gaussian edge potentials. *Advances in neural information processing systems* 24 (2011).

[23] Jean-François Lalonde, Alexei A Efros, and Srinivasa G Narasimhan. 2012. Estimating the natural illumination conditions from a single outdoor image. *International Journal of Computer Vision* 98, 2 (2012), 123–145.

[24] Hieu Le and Dimitris Samaras. 2020. From Shadow Segmentation to Shadow Removal. In *The IEEE European Conference on Computer Vision (ECCV)*.

[25] Hieu Le and Dimitris Samaras. 2021. Physics-based shadow image decomposition for shadow removal. *IEEE Transactions on Pattern Analysis & Machine Intelligence* 01 (2021), 1–1.

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

[27] Jingwei Liao, Yanli Liu, Guanyu Xing, Housheng Wei, Jueyu Chen, and Songhua Xu. 2021. Shadow Detection via Predicting the Confidence Maps of Shadow Detection Methods. In *Proceedings of the 29th ACM International Conference on Multimedia*. 704–712.

[28] Vu Nguyen, Tomas F Yago Vicente, Maozheng Zhao, Minh Hoai, and Dimitris Samaras. 2017. Shadow detection with conditional generative adversarial networks. In *Proceedings of the IEEE International Conference on Computer Vision*. 4510–4518.

[29] Takahiro Okabe, Imari Sato, and Yoichi Sato. 2009. Attached shadow coding: Estimating surface normals from shadows under unknown reflectance and lighting conditions. In *2009 IEEE 12th International Conference on Computer Vision*. IEEE, 1693–1700.

[30] Alexandros Panagopoulos, Dimitris Samaras, and Nikos Paragios. 2009. Robust shadow and illumination estimation using a mixture model. In *2009 IEEE Conference on Computer Vision and Pattern Recognition*. IEEE, 651–658.

[31] Alexandros Panagopoulos, Chaohui Wang, Dimitris Samaras, and Nikos Paragios. 2012. Simultaneous cast shadows, illumination and geometry inference using hypergraphs. *IEEE transactions on pattern analysis and machine intelligence* 35, 2 (2012), 437–449.

[32] Li Shen, Teck Wee Chua, and Karianto Leman. 2015. Shadow optimization from structured deep edge detection. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2067–2074.

[33] Mingxing Tan and Quoc Le. 2019. Efficientnet: Rethinking model scaling for convolutional neural networks. In *International conference on machine learning*. PMLR, 6105–6114.

[34] Jiandong Tian, Xiaojun Qi, Liangqiong Qu, and Yandong Tang. 2016. New spectrum ratio properties and features for shadow detection. *Pattern Recognition* 51 (2016), 85–96.

[35] Tomás F. Yago Vicente, Minh Hoai, and Dimitris Samaras. 2018. Leave-One-Out Kernel Optimization for Shadow Detection and Removal. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 40, 3 (2018), 682–695. https://doi.org/10.1109/TPAMI.2017.2691703

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

[37] Tomás F Yago Vicente, Chen-Ping Yu, and Dimitris Samaras. 2013. Single Image Shadow Detection Using Multiple Cues in a Supermodular MRF.. In *BMVC*.

[38] Jifeng Wang, Xiang Li, and Jian Yang. 2018. 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*. 1788–1797.

[39] Tiantian Wang, Ali Borji, Lihe Zhang, Pingping Zhang, and Huchuan Lu. 2017. A stagewise refinement model for detecting salient objects in images. In *Proceedings of the IEEE international conference on computer vision*. 4019–4028.

[40] Tianyu Wang, Xiaowei Hu, Chi-Wing Fu, and Pheng-Ann Heng. 2021. Single-Stage Instance Shadow Detection with Bidirectional Relation Learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 1–11.

[41] Yupei Wang, Xin Zhao, Yin Li, Xuecai Hu, Kaiqi Huang, and NLPR CRIPAC. 2018. Densely Cascaded Shadow Detection Network via Deeply Supervised Parallel Fusion.. In *IJCAI*, Vol. 5. 6.

[42] Scott Wehrwein, Kavita Bala, and Noah Snavely. 2015. Shadow detection and sun direction in photo collections. In *2015 International Conference on 3D Vision*. IEEE, 460–468.

[43] Lin Wu, Xiaochun Cao, and Hassan Foroosh. 2010. Camera calibration and geo-location estimation from two shadow trajectories. *Computer Vision and Image Understanding* 114, 8 (2010), 915–927.

[44] Pingping Zhang, Dong Wang, Huchuan Lu, Hongyu Wang, and Xiang Ruan. 2017. Amulet: Aggregating multi-level convolutional features for salient object detection. In *Proceedings of the IEEE international conference on computer vision*. 202–211.

[45] Jia-Xing Zhao, Jiang-Jiang Liu, Deng-Ping Fan, Yang Cao, Jufeng Yang, and Ming-Ming Cheng. 2019. EGNet: Edge guidance network for salient object detection. In *Proceedings of the IEEE/CVF international conference on computer vision*. 8779–8788.

[46] Wenda Zhao, Cai Shang, and Huchuan Lu. 2021. Self-generated Defocus Blur Detection via Dual Adversarial Discriminators. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 6933–6942.

[47] Quanlong Zheng, Xiaotian Qiao, Ying Cao, and Rynson WH Lau. 2019. Distraction-aware shadow detection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 5167–5176.

[48] Bolei Zhou, Aditya Khosla, Agata Lapedriza, Aude Oliva, and Antonio Torralba. 2016. Learning deep features for discriminative localization. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2921–2929.

[49] Huajun Zhou, Xiaohua Xie, Jian-Huang Lai, Zixuan Chen, and Lingxiao Yang. 2020. Interactive two-stream decoder for accurate and fast saliency detection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*.

9141–9150.

[50] Jiejie Zhu, Kegan GG Samuel, Syed Z Masood, and Marshall F Tappen. 2010. Learning to recognize shadows in monochromatic natural images. In *2010 IEEE Computer Society conference on computer vision and pattern recognition*. IEEE, 223–230.

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

[52] Lei Zhu, Ke Xu, Zhanghan Ke, and Rynson WH Lau. 2021. Mitigating Intensity Bias in Shadow Detection via Feature Decomposition and Reweighting. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 4702–4711.