Distraction-Aware Edge Enhancement for Shadow Detection in Remote Sensing Images

Boyong Du, Li Wang, and Sheng Xu

***Abstract*— Currently, the complex features of remote sensing images complicate shadow detection, such as blurred shadow edges, a large number of fragmented shadows, and varying shadow colors. To address these challenges, a distraction-aware edge enhancement for shadow detection network (DESDNet) is proposed. It utilizes distraction-aware shadow (DS) modules to reduce false positives (FPs) and false negatives (FNs), while employing edge enhancement modules to improve shadow edge detection. In this network, the shallowest convolutional layers capture detailed shadow edge features, including many nonshadow background details, while the deepest convolutional layers, with larger receptive fields, effectively suppressing nonshadow pixels. The edge module integrates the two features to mitigates issues, such as edge blurring, fragmentation, and color inconsistency in shadows. The proposed network has been validated for feasibility on remote sensing image datasets, on which our experiments achieve the accuracies of 98.48% on the ISTD dataset and 96.70% on the SBU dataset, outperforming typical networks recently. The code is available from https://github.com/sfs0/DESDNet/tree/main.**

***Index Terms*— Deep learning, edge enhancement, remote sensing image, shadow detection.**

I. Introduction

SHADOWS are natural occurrences when light is obstructed. While they enrich our understanding of image scenes, they also present technical challenges in image processing. Currently, the primary focus lies in shadow removal, yet accurate shadow detection is an indispensable prerequisite. Due to the intricate interplay of geometry and illumination, shadows introduce a fundamental challenge in computer vision, affecting tasks, such as image segmentation, object detection, and low-light image enhancement [1], [2]. Both classic and deep learning methods have not fully addressed issues in shadow detection, such as blurred shadow boundaries, fragmented shadows, and soft shadows.

Classic Methods: In the early stages, researchers developed methods using physical models to exploit color and illumination for shadow detection. For example, Salvador et al. [3] investigated the spectral and geometrical properties of shadows to segment cast shadows. Panagopoulos et al. [4] designed a high-order Markov random field illumination model incorporating coarse 3-D geometry information. Tian et al. [5] utilized differences in spectral power distributions between daylight and skylight for shadow detection, based on the assumption of illumination invariance. These designs primarily assume color constancy and illumination invariance and rely on manual annotations. However, in real-world scenarios, shadows often exhibit variations in color, making it challenging to annotate subtle or fragmented shadows manually.

Deep Learning Methods: Deep learning has significantly impacted shadow detection in computer vision. Researchers have explored various CNN-based strategies. Hosseinzadeh et al. [6] proposed a swift deep shadow detection network. Hu et al. [7] proposed the direction-aware spatial context features. Zhu et al. [8] designed the recursive attention residual (RAR) module. Zheng et al. [9] introduced a distraction-aware shadow (DS) module for predicting false positives (FPs) and false negatives (FNs). Chen et al. [10] proposed incorporating multitask learning into a self-organizing framework for shadow detection tasks. Zhu et al. [11] introduced a feature decomposition and reweighting scheme to optimize the process. Zhu et al. [12] proposed a complementary mechanism for shadow detection. Sun et al. [13] proposed a multi-to-multi-mapping from high dynamic range original images to sRGB images for multiscale contrast in shadow detection. Feng et al. [14] introduced a material matching-based shadow detection method that does not require training. Liu et al. [15] proposed a shadow detection method based on a multiscale spatial attention mechanism. Chen et al. [16] introduced methods for slice-to-slice context passing and uncertainty region calibration.

While the existing methods have improved detection accuracy to some extent, they perform poorly in detecting shadow edges, fragmented shadows, inconsistent shadow colors, and soft shadows [17], [18].

This letter proposes a distraction-aware edge enhancement network to address the identification of shadow boundaries, fragmentation, and color variations in remote sensing images. The main contributions include the following.

1) By introducing a new edge enhancement module to extract rich edge information, we enhance the model's detection capabilities, addressing the issue of shadow edge blurring overlooked by most existing networks.
2) The improved network effectively identifies edge shadow areas and scenarios with similar features, such as linear, fragmented, and soft shadows, addressing the shortcomings of most existing methods.

Manuscript received 11 April 2024; revised 10 June 2024; accepted 13 June 2024. Date of publication 17 June 2024; date of current version 26 June 2024. This work was supported in part by the Scientific and Technological Innovation 2030-Major Projects under Grant 2023ZD0405605 and in part by the Practice Innovation Training Program Projects for Jiangsu College Students under Grant 202310298040Z. *(Corresponding author: Sheng Xu.)*

The authors are with the College of Information Science and Technology and Artificial Intelligence, Nanjing Forestry University, Nanjing 210037, China (e-mail: dby@njfu.edu.cn; xusheng@njfu.edu.cn).

Digital Object Identifier 10.1109/LGRS.2024.3415637

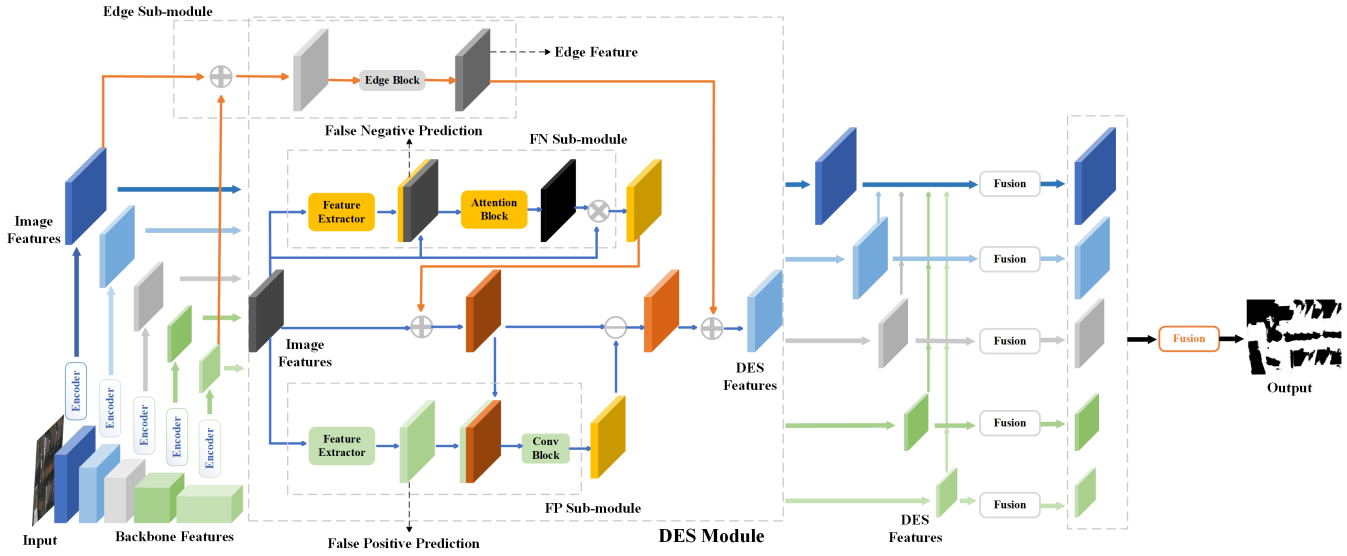

Fig. 1. DESD network architecture: DESDNet takes an image and outputs a shadow map end-to-end. Initially, the image undergoes feature extraction at different scales. Subsequently, image features enter the DES module to generate DES features for shadow detection. Finally, all DES features are upsampled, connected, and fused to produce a shadow score map, which is then merged to generate the final predicted shadow map.

3) The improved network can handle more complex images. It not only demonstrates superior performance on public datasets but also transfers effectively to complex remote sensing images, yielding significant results.

II. Method

Fig. 1 illustrates our distraction-aware edge enhancement for shadow detection network (DESDNet) architecture, an optimized version of DSDNet, using ResNeXt-101 [19] as the backbone network. Initially, the image undergoes feature extraction at five different scales. Then, the DES module takes these features, combines them with features from deep and shallow CNN layers, and generates DES features. After that, bilinear interpolation is used for upsampling, followed by the dense connections. Fusion is applied using two convolutional layers. Specifically, let F_k denote the upsampled features at scale k, and the merged features at the current scale can be obtained as $F_k^m = \text{Conv}(\text{Concat}(F_k, \ldots, F_1))$. Finally, all shadow score maps are merged using a 1×1 convolutional layer, and a sigmoid activation function is applied to output the soft binary shadow map.

A. Module

As shown in Fig. 1, the DES module takes image features at different scales as input and outputs DES features. Its main function is to learn various features to enhance image characteristics, ultimately outputting edge-enhanced distraction-aware features. In the following, we will discuss FN, FP, and edge submodules in detail.

FN Submodule: It is designed to learn FN features and FN mask features to enhance input image features. Initially, FN features are extracted from the image features for FN prediction. Then, a soft binary map is estimated to indicate potential FN locations, capturing the required semantic information. FN features are combined with image features and input into an attention block to generate a soft mask. After elementwise multiplication with the image features, FN region features are activated, resulting in enhanced FN image features. The attention mechanism assists the network in swiftly focusing on features surrounding potential FN regions.

FP Submodule: Similar to the FN submodule, this submodule extracts FP features from image features, predicts a soft binary map for FP, and captures useful semantic information for potential FP regions. Then, FP features are combined with enhanced FN image features and processed through a convolutional block to generate FP-aware image features. These features are subtracted from the original image features to reduce FP interference. The convolutional block, consisting of multiple convolutional layers, captures larger contextual information to effectively distinguish between FP regions and true shadows.

Edge Submodule: The deepest convolutional layer has a larger receptive field, effectively suppressing nonshadow pixels, while the shallowest convolutional layer captures rich shadow edge features, but also includes many nonshadow background details. By combining the features from these two layers, complementing each other's strengths and weaknesses, the merged result is input into the edge block. The obtained shadow edge features are then added to the distractedly processed image features to output the final DES feature. The edge block consists of multiple convolutional layers, effectively enhancing shadow edge information and suppressing nonshadow information. In addition, experiments show notable results in detecting soft shadows.

B. Train

We employed the optimization approach of the base network DSD for training our network, jointly optimizing the predictions of shadows, FN, and FP maps across all scales, alongside the final predictions of shadows, FN, and FP maps

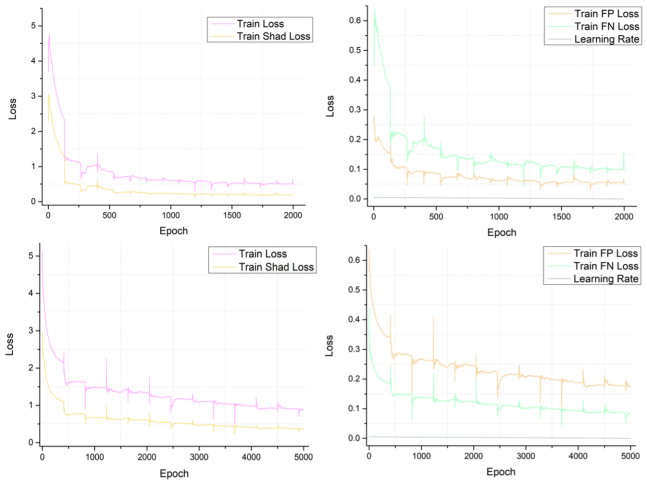

Fig. 2. Loss function trained on the ISTD and SBU datasets. On the left are the train loss and shadow loss, while on the right are the FP loss and FN loss.

by minimizing the objective

$$L = \alpha \sum_i L^i_{\text{shadow}} + \beta \sum_i L^i_{\text{fn}} + \lambda \sum_i L^i_{\text{fp}} + \alpha L^F_{\text{shadow}} + \beta L^F_{\text{fn}} + \lambda L^F_{\text{fp}} \tag{1}$$

where L^i_{shadow}, L^i_{fn}, and L^i_{fp} are the losses of shadow, FN, and FP map predictions at the ith scale, respectively. L^F_{shadow}, L^F_{fn}, and L^F_{fp} are the losses of the final predictions of shadow, FN, and FP maps. Here, the probability of the jth pixel is represented as p_j, and the corresponding ground truth is represented as y_j. In addition, we formulated the shadow loss at scale i as $L^i_{\text{shadow}} = l_1 + l_2$, where l_1 [see (2)] is the weighted cross-entropy loss and l_2 [see (3)] is the attention-dispersed cross-entropy loss. The former handles cases where nonshadow pixels outnumber shadow pixels, while the latter addresses regions prone to misidentification

$$l_1 = \sum_j \left(-\frac{N_n}{N_n + N_p} y_i \log(p_j) - \frac{N_p}{N_n + N_p} (1 - y_j) \log(1 - p_j) \right) \tag{2}$$

where the index j iterates over all pixels in the image. N_n and N_p represent the numbers of FN and FP pixels, respectively,

$$l_2 = \sum_j \left(-\frac{N_n}{N_n + N_p} y^{\text{fnd}}_j y_i \log(p_j) - \frac{N_p}{N_n + N_p} y^{\text{fnd}}_j (1 - y_j) \log(1 - p_j) \right) \tag{3}$$

where y^j_{fnd} is the ground-truth value for FN pixels and y^j_{fpd} is the ground-truth value for FP pixels.

As shown in Fig. 2, during training on both datasets, we monitored the loss over 6000 iterations. By progressively adjusting the number of iterations, we found optimal performance at around 2000 iterations for ISTD and around 5000 iterations for SBU, with training losses converging to approximately 0.48 and 0.87, respectively.

III. Experiments and Evaluations

Network Details: Our experiments used PyTorch version 1.10.0, Python version 3.8 on Ubuntu 20.04, and an RTX 3080 Ti with 12-GB VRAM. ResNeXt-101 served as the backbone network for a fair comparison. Convolutional layers in our network follow a uniform structure with batch normalization and ReLU activation, unless stated otherwise. The encoder in Fig. 1 has two 3×3 convolutional layers with 32 kernels. Both FN submodule and FP submodule feature extractors include two 3×3 convolutional layers with 32 kernels. The FN submodule's attention block comprises one 3×3 convolutional layer with 64 kernels, followed by a sigmoid activation. The FP submodule's Conv block has one residual block with three convolutional layers of varying filter and kernel sizes. The edge submodule uses three 3×3 convolutional layers with 32 input and output channels. It also includes a 1×1 convolutional layer with 32 input channels, 1 output channel, and a sigmoid activation layer.

Training Details: The ResNeXt-101 undergoes pretraining on ImageNet, with other parameters initialized randomly. We fine-tune using SGD optimizer with a momentum of 0.9, a weight decay of 10^{-3}, and a batch size of 10. The initial learning rate is set to 5×10^{-3}, employing a polynomial strategy with a power of 0.9 for learning rate decay. Data augmentation includes random horizontal flipping, and images are resized to 320×320. The model is trained for 6000 iterations. Loss weights are set to $\alpha = 1$, $\beta = 2$, and $\lambda = 2$. During inference, input images are resized to 320×320, and CRF is applied for postprocessing.

We used four datasets, among which the ISTD and SBU public datasets were used for model training and testing. In addition, the AISD dataset and our own remote sensing dataset were used for model testing to validate generalization.

A. Quantitative Analysis

In the quantitative analysis section, we employ accuracy as the metric for evaluating the accuracy of shadow detection, where a higher score indicates better performance. As shown in (4), N_{tp}, N_{tn}, N_{fp}, and N_{fn} represent the numbers of true positives, true negatives, FPs, and FNs, respectively. Using precision, recall, and F_1-score to assist in evaluating the performance of the model

$$\text{Accuracy} = \frac{N_{\text{tp}} + N_{\text{tn}}}{N_{\text{tn}} + N_{\text{fn}} + N_{\text{tp}} + N_{\text{fp}}}. \tag{4}$$

We compared our method with several typical shadow detection networks (DSC, DSD, MTMT, FDR, SDCM, and MSASD) on the ISTD and SBU datasets. For DSD and SDCM, we used the results provided by the authors. For the other four networks, we reproduced the results as closely as possible based on their code. Table I demonstrates that our method outperforms the other networks. Specifically, it surpasses the best-performing network by 0.59% and 0.50% points on ISTD and SBU, respectively. Additionally, there are some limitations in precision, which may be attributed

TABLE I
PERFORMANCE COMPARISON ON TWO PUBLIC DATASETS

methods	ISTD Accuracy	ISTD Precision	ISTD Recall	ISTD F_1-score	SBU Accuracy	SBU Precision	SBU Recall	SBU F_1-score
DESDNet(**Ours**)	**98.48%**	**99.54%**	**91.88%**	**95.55%**	**96.70%**	96.81%	**87.74%**	**92.06%**
DSCNet [7]	97.33%	99.10%	90.84%	94.79%	95.27%	96.22%	86.12%	90.89%
DSDNet [9]	97.54%	99.31%	86.96%	92.73%	96.20%	97.25%	85.28%	90.87%
MTMTNet [10]	97.70%	99.25%	88.43%	93.53%	95.46%	95.98%	87.70%	91.65%
FDRNet [11]	97.83%	98.95%	89.58%	94.03%	96.02%	**97.26%**	87.21%	91.96%
SDCMNet [12]	97.89%	98.80%	90.64%	94.54%	95.53%	96.25%	87.36%	91.59%
MSASDNet [15]	97.87%	98.82%	90.59%	94.07%	95.62%	96.48%	87.35%	91.69%

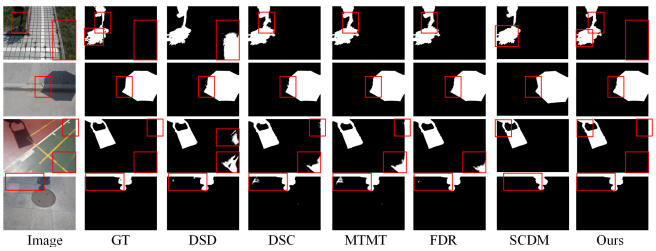

Fig. 3. Results of the comparison on the ISTD dataset.

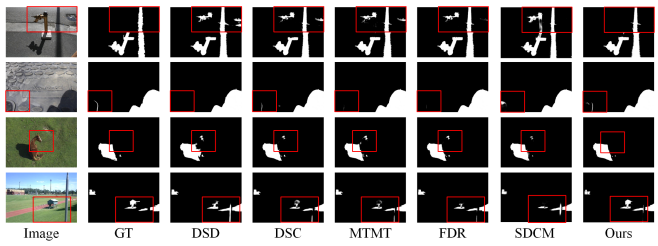

Fig. 4. Results of the comparison on the SBU dataset.

to the similarity in color between objects in the samples and shadows. Overall, our network exhibits superior generalization ability. By enforcing a balance between shadow and nonshadow regions with the attentive model, it further addresses issues, such as blurry shadow edges, difficulty in distinguishing fragmented shadows, inconsistent shadow colors, and unclear identification of linear soft shadows.

B. Qualitative Analysis

Next, we present some visual results, first comparing them on two public datasets and then applying them to remote sensing image datasets.

The red boxes highlight our improvements. From the results, it can be seen that in ISTD (Fig. 3), our detection results have smoother edges (second row), eliminating false detections (third and fourth rows), and better detecting scattered shadow points (first row).

In SBU (Fig. 4), our method better distinguishes cases where object colors are similar to shadow colors (first row), can identify linear shadows under soft shadows, and better fit the ground truth (second, third, and fourth rows). Overall, our approach overcomes the complex issues of unclear shadow edges, difficult differentiation of fragmented shadows, and false detections caused by inconsistent shadow colors.

In the AISD dataset, besides basic shadows of houses and trees, there are also a small number of broken and linear shadows. In addition to containing these elements, our dataset also includes many houses with colors similar to shadows and a large number of linear shadows, making the dataset more diverse and better testing the model's generalization capability. As shown in Figs. 5 and 6, when applied to more complex remote sensing images, the network demonstrates effective shadow detection capabilities, accurately identifying the edges

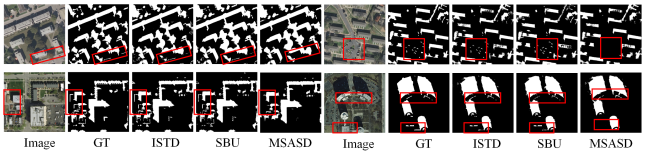

Fig. 5. Visual effects of training our network models on ISTD and SBU and applying them to AISD.

TABLE II
ABLATION STUDY RESULTS CONDUCTED SEPARATELY ON TWO DATASETS

Network	FP	FN	Edge	ISTD Accuracy	SBU Accuracy
Baseline+FP+FN	✓	✓	×	97.54%	96.20%
Baseline+FP+Edge	✓	×	✓	96.70%	95.50%
Baseline+FN+Edge	×	✓	✓	96.33%	95.55%
Full model(**Ours**)	✓	✓	✓	**98.48%**	**96.70%**

of shadows from trees and buildings, and effectively detecting scattered, fragmented, and soft shadows.

C. Ablation Study

To evaluate the design choices of our proposed distraction-aware edge enhancement shadow module, we conducted comparative experiments using accuracy as the evaluation metric, as shown in Table II.

The table shows that removing the FP module causes the most notable performance drop on the ISTD dataset, while omitting the FN module results in a significant decrease in performance on the SBU dataset. However, having all three modules present achieves optimal performance. Thus, the FN, FP, and edge modules are essential, as the absence of any one of them reduces performance. The first two modules have a significant impact, while adding the third module improves the overall network. This highlights the importance of all three

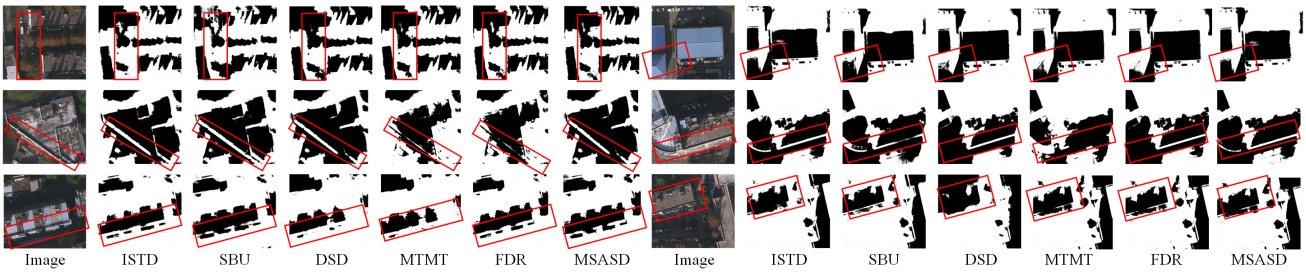

Fig. 6. Application and comparative effects on remote sensing images. ISTD and SBU represent our network's results, while results from other networks were obtained using models trained on SBU.

modules in adapting to complex shadows in remote sensing images.

IV. Conclusion

This letter introduces a shadow detection network based on distraction-aware edge enhancement. When dealing with complex shadow scenarios in remote sensing images, we enhance shadow edge detection by incorporating synthesized edge image features into the existing distraction-aware networks. Experimental results demonstrate that our network achieves an improvement of 0.59% and 0.5% in accuracy on the ISTD and SBU datasets, respectively, when compared with the best-performing networks. In remote sensing images, it effectively addresses issues, such as edge blurring and erroneous detection of linear shadows, and distinguishes fragmented shadows and shadows with inconsistent colors. Our approach achieves a balance between core shadow areas and edge shadow areas.

Due to the dependency of two distraction submodules on the predicted soft binary maps, we adopted a transfer learning approach, directly applying the model trained on public datasets to the remote sensing dataset for prediction. In the future, our research will aim to eliminate this dependency and enhance the applicability of the network.

References

[1] L. Xing, H. Qu, S. Xu, and Y. Tian, "CLEGAN: Toward low-light image enhancement for UAVs via self-similarity exploitation," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, 2023, Art. no. 5610714.

[2] H. Zheng, L. Fu, and Q. Ye, "Flexible capped principal component analysis with applications in image recognition," *Inf. Sci.*, vol. 614, pp. 289–310, Oct. 2022.

[3] E. Salvador, A. Cavallaro, and T. Ebrahimi, "Cast shadow segmentation using invariant color features," *Comput. Vis. Image Understand.*, vol. 95, no. 2, pp. 238–259, Aug. 2004.

[4] A. Panagopoulos, C. Wang, D. Samaras, and N. Paragios, "Illumination estimation and cast shadow detection through a higher-order graphical model," in *Proc. CVPR*, Colorado Springs, CO, USA, Jun. 2011, pp. 673–680.

[5] J. Tian, X. Qi, L. Qu, and Y. Tang, "New spectrum ratio properties and features for shadow detection," *Pattern Recognit.*, vol. 51, pp. 85–96, Mar. 2016.

[6] S. Hosseinzadeh, M. Shakeri, and H. Zhang, "Fast shadow detection from a single image using a patched convolutional neural network," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst. (IROS)*, Oct. 2018, pp. 3124–3129.

[7] X. Hu, C. Fu, L. Zhu, J. Qin, and P. Heng, "Direction-aware spatial context features for shadow detection and removal," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 42, no. 11, pp. 2795–2808, Nov. 2020.

[8] L. Zhu et al., "Bidirectional feature pyramid network with recurrent attention residual modules for shadow detection," in *Proc. Eur. Conf. Comput. Vis.*, 2018, pp. 121–136.

[9] Q. Zheng, X. Qiao, Y. Cao, and R. W. H. Lau, "Distraction-aware shadow detection," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 5162–5171.

[10] Z. Chen, L. Zhu, L. Wan, S. Wang, W. Feng, and P. Heng, "A multi-task mean teacher for semi-supervised shadow detection," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 5610–5619.

[11] L. Zhu, K. Xu, Z. Ke, and R. W. H. Lau, "Mitigating intensity bias in shadow detection via feature decomposition and reweighting," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2021, pp. 4682–4691.

[12] Y. Zhu, X. Fu, C. Cao, X. Wang, Q. Sun, and Z.-J. Zha, "Single image shadow detection via complementary mechanism," in *Proc. 30th ACM Int. Conf. Multimedia*, Oct. 2022, pp. 6717–6726.

[13] J. Sun et al., "Adaptive illumination mapping for shadow detection in raw images," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2023, pp. 12663–12672.

[14] J. Feng, Y. K. Kim, and P. Liu, "Image shadow detection and removal based on region matching of intelligent computing," *Comput. Intell. Neurosci.*, vol. 2022, pp. 1–9, Apr. 2022.

[15] D. Liu, J. Zhang, Y. Wu, and Y. Zhang, "A shadow detection algorithm based on multiscale spatial attention mechanism for aerial remote sensing images," *IEEE Geosci. Remote Sens. Lett.*, vol. 19, pp. 1–5, 2022.

[16] H. Chen et al., "Slice-to-slice context transfer and uncertain region calibration network for shadow detection in remote sensing imagery," *ISPRS J. Photogramm. Remote Sens.*, vol. 203, pp. 166–182, Sep. 2023.

[17] L. Shi and Y.-F. Zhao, "Urban feature shadow extraction based on high-resolution satellite remote sensing images," *Alexandria Eng. J.*, vol. 77, pp. 443–460, Aug. 2023.

[18] Z. Wang, Y. Zhou, F. Wang, S. Wang, G. Qin, and J. Zhu, "Shadow detection and reconstruction of high-resolution remote sensing images in mountainous and hilly environments," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 17, pp. 1233–1243, 2024.

[19] S. Xie, R. Girshick, P. Dollár, Z. Tu, and K. He, "Aggregated residual transformations for deep neural networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 5987–5995.