ESPFNet: An Edge-Aware Spatial Pyramid Fusion Network for Salient Shadow Detection in Aerial Remote Sensing Images

Shuang Luo, Huifang Li, *Member, IEEE*, Ruzhao Zhu, Yuting Gong, and Huanfeng Shen, *Senior Member, IEEE*

***Abstract*—Shadows can hinder image interpretation in aerial remote sensing images. The existing shadow detection methods focus on all shadow regions and detect the shadow regions directly, but they ignore the fact that salient shadows have a more significant effect. In this work, a novel edge-aware spatial pyramid fusion network (ESPFNet) under a multitask learning framework is proposed for salient shadow detection in aerial remote sensing images. ESPFNet has three components: a parallel spatial pyramid (PSP) structure; an edge detection module (EDM); and an edge-aware multibranch integration (EMI). The PSP structure is constructed to extract multiscale features from the input image and fuse them gradually. The EDM then integrates the shallow features and deep features to detect the shadow edges. Finally, the EMI incorporates the edge features with multibranch features, and then concatenates them with the shallow features to generate the salient shadow detection result. The experimental analyses confirm the effectiveness of the ESPFNet method in both the qualitative and quantitative performance, compared to the existing methods, with the F-score reaching 92.04% in the salient shadow test set.**

***Index Terms*—Aerial remote sensing images, convolutional neural network, multitask learning, salient shadow detection.**

I. Introduction

SHADOW is a widespread phenomenon in high-resolution aerial images, especially in urban regions, due to the numerous high-altitude land covers, such as buildings, bridges and trees [1], [2]. Shadows can provide additional geometric information for object location and altitude estimation, but lead to radiometric information loss, making the image interpretation more difficult [3], [4], [5]. Therefore, for both the geometric and radiometric applications of remote sensing images, shadow detection is of great importance.

Any obstacle, such as a car, tree, or building, leads to shadow when direct light exists, and these shadows can be cast on any surface, such as a road, roof, or another shadow. Thus, shadows show three distinctive spatial features in high-resolution aerial images: shadows are usually widely dispersed; the shape and size of shadows are varied; and shadows are usually mutually connected, as shown in Fig. 1. From the remote sensing image application viewpoint, not all the shadows are of concern. In fact, only those salient shadows with a relatively large area, a medium length-width ratio, or shadows cast by a main target need to be treated to benefit geometric location or image interpretation. Therefore, we focus on salient shadow detection in this article.

The existing shadow detection methods, i.e., the geometrical methods and property-based methods, usually detect all the shadows, and they do not distinguish between salient and non-salient shadows in remote sensing images [6]–[10]. The geometrical methods determine the location of shadows through modeling the geometrical optics based on the surface altitude and the sensor position information, which can be difficult to acquire for high-resolution images [11]–[13]. The property-based methods take the shadow properties into account, and they directly recognize the shadow regions from the image [3], [6], [7], [14]–[18]. The property-based methods can be categorized into two main groups: thresholding-based methods; and machine learning based methods. For the thresholding-based methods, the image is usually transformed into a special feature space and then a shadow index is constructed to highlight the shadows, based on which a threshold can be set to separate shadow from nonshadow [3], [8], [9], [14], [16]–[18]. This kind of method is simple and efficient, but the use of only a shadow index is not sufficient to support accurate shadow detection. While for the machine learning based methods, a lot of samples are selected to train the classifier [19]. The traditional machine learning methods are mainly based on pixel-wise training samples, such as support vector machine [7], [20], and they do not consider the spatial correlations among adjacent pixels, which can lead to noisy results. In recent years, deep learning based methods have been introduced to detect the shadows in remote sensing images and have shown promising results, due to the powerful ability of data learning [6], [21]. As the accuracy of the deep learning based methods is heavily dependent on the training data, the aerial imagery dataset for shadow detection (AISD)

Manuscript received January 24, 2021; revised March 1, 2021; accepted March 12, 2021. Date of publication March 17, 2021; date of current version May 20, 2021. This work was supported in part by the National Natural Science Foundation of China under Grant 41631180, Grant 41871246, and Grant 41971303. *(Corresponding author: Huifang Li.)*

Shuang Luo, Huifang Li, and Yuting Gong are with the School of Resource and Environmental Sciences, Wuhan University, Wuhan 430079, China (e-mail: sluo@whu.edu.cn; huifangli@whu.edu.cn; yutinggong@whu.edu.cn).

Ruzhao Zhu is with the KylinSoft Company, Ltd., Changsha 410073, China (e-mail: zhuruzhao@kylinos.cn).

Huanfeng Shen is with the School of Resource and Environmental Sciences and the Collaborative Innovation Center of Geospatial Technology, Wuhan University, Wuhan 430079, China (e-mail: shenhf@whu.edu.cn).

Digital Object Identifier 10.1109/JSTARS.2021.3066791

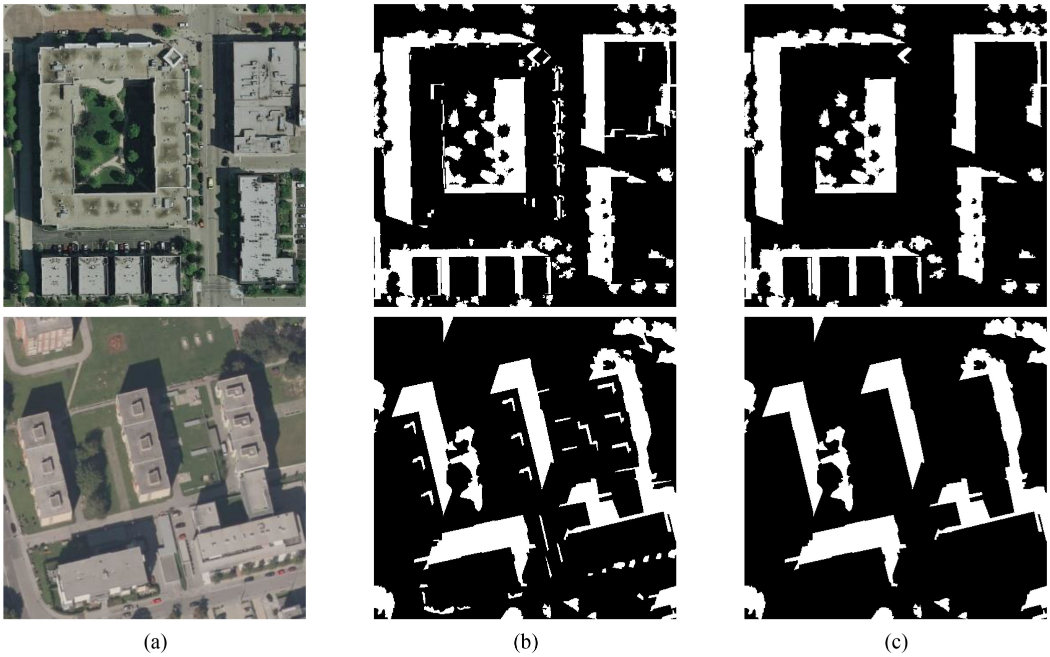


Fig. 1. Representative examples of salient shadow masks. (a) Shadow images. (b) Original detailed shadow masks. (c) Salient shadow masks.

have been released publicly, which is the first aerial shadow imagery dataset [6]. A deeply supervised convolutional neural network for shadow detection (DSSDNet) was also proposed at the same time. However, these methods above consider all the shadows equally, and they do not rank the shadows. As a result, these methods cannot identify the salient shadows.

In contrast, numerous salient object detection (SOD) methods have been developed in computer vision (CV) [22]–[28]. SOD is a task based on a visual attention mechanism, which aims to detect objects more attractive than the surrounding areas in a scene or an image [29], [30]. SOD occupies an important position in many CV applications, such as object recognition [31], [32], visual tracking [33], and image retrieval [34]. Recent deep learning based methods have shown a superior performance by learning global contextual features along with the local context [27], or by adding extra edge information for SOD [35], [36].

In this article, inspired by the above observations, a novel edge-aware spatial pyramid fusion network (ESPFNet) is proposed to detect the salient shadows in aerial remote sensing images. The major contributions of the proposed method are as follows. First, salient shadows are considered for the first time in remote sensing images, and some discriminative rules are defined. Second, a multitask framework for salient shadow detection is proposed by simultaneously detecting salient shadow regions and shadow edges from a single image. The whole network has three main parts: the parallel spatial pyramid (PSP) structure; the edge detection module (EDM); and the edge-aware multibranch integration (EMI).

The rest of this article is organized as follows. Section II gives a definition of salient shadows in remote sensing images. Section III presents the proposed salient shadow detection method in detail. Section IV shows the qualitative and quantitative analysis for the experimental results. The postprocessing of small shadow regions removal is discussed in Section V. The conclusion of this article is drawn in Section VI.

II. Salient Shadows in Remote Sensing Images

Inspired by the concept of salient objects in close-shot images, the concept of salient shadows is based on a visual attention mechanism which is more attractive than the surrounding areas, and the influence for image interpretation in remote sensing images. Salient shadows have some special features, as follows

1) *Relatively Large Area:* A salient shadow usually covers a relatively large area, and the shadows of small objects can be excluded, such as cars, sparse trees, etc.
2) *Medium Length-Width Ratio:* The slender shadow regions with a large length-width ratio can be regarded as non-salient shadow.
3) *Cast by the Main Targets:* In general, salient shadows are cast by the main targets in remote sensing images, such as buildings, bridges, and large trees.

To explain the concept of salient shadow more clearly, two representative examples of salient shadow masks are shown in Fig. 1. It can be seen, the small shadows on the roof and the slender shadow regions near the boundary of the buildings in Fig. 1(b) have been excluded in Fig. 1(c). The fragmentary shadows of vegetation have also been eliminated in Fig. 1(c).

III. Edge-Aware Spatial Pyramid Fusion Network

In this section, the overall architecture and key components of the proposed ESPFNet are introduced. Specifically, the three components of ESPFNet, i.e., the PSP structure, the EDM, and the EMI, and also the loss functions, are illustrated in detail.

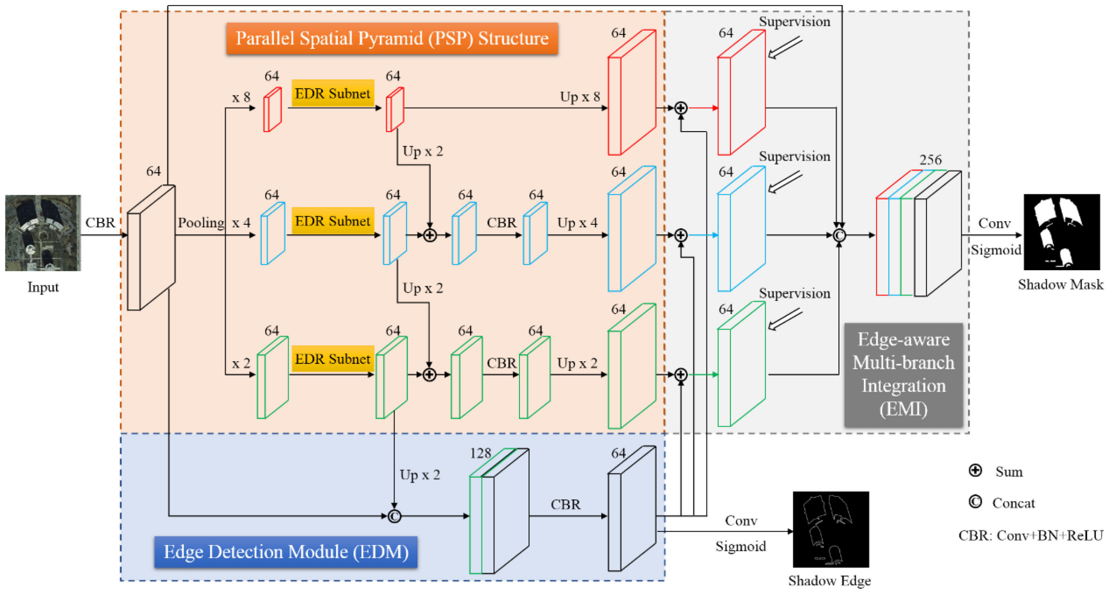


Fig. 2. Architecture of the proposed ESPFNet framework. The ESPFNet has three components. (a) PSP structure for feature extraction. (b) EDM for edge prediction. (c) EMI for feature fusion.

A. *Overall Architecture*

Given a shadow image, the proposed ESPFNet simultaneously detects the salient shadow regions and the corresponding shadow edges in an end-to-end manner in Fig. 2.

The PSP structure is mainly composed of spatial pyramid pooling (SPP) and encoder-decoder residual (EDR) subnets. It extracts multiscale feature representations and eliminates the non-salient information, while fusing the features between different scales gradually. The EDM focuses on the shadow edge detection task by integrating the shallow features containing detail information with the deep features containing abstract information. The EMI not only incorporates the edge features with multibranch features, but also concatenates the edge-aware branches with shallow features to generate the salient shadow detection result.

Theoretically, the PSP structure can extract the salient shadow regions and eliminate the non-salient shadow regions for a shadow dataset with detailed annotation or salient annotation.

B. *PSP Structure for Feature Extraction*

1) *Spatial Pyramid Pooling:* The SPP can collect multi-level information, and is more representative than global pooling [37]–[39]. Therefore, the SPP is introduced to extract multiscale feature representations and eliminate the non-salient information.

Given an input image, as shown in Fig. 2, a convolutional layer is followed by a batch normalization (BN) layer [40] and a rectified linear unit (ReLU) [41], which is referred as a "CONV+BN+ReLU" (CBR) block. SPP is then used to resize the input feature maps into three different scales, with pooling sizes of 2 × 2, 4 × 4, and 8 × 8, respectively, covering half of and small portions of the image. The type of pooling operation is average pooling, following the pooling operation used in PSPNet [39].

2) *Encoder-Decoder Residual Subnet:* Through the SPP processing, features in three different spatial scales can be extracted in the proposed network. In order to extract and aggregate multiscale shadow information, a set of subnets are built independently for each level. All the subnets have the same EDR module in Fig. 3.

For the encoder network, the input feature maps with 64 channels are gradually filtered with convolutional operation and down-sampled with max pooling operation at each stage to extract deep context features. The decoder network is then constructed to up-sample the feature maps progressively and concatenate them with the corresponding encoder feature maps. Furthermore, residual learning [42] is adopted to accelerate the network convergence and improve the performance of salient shadow detection in the proposed method. Specifically, the residual blocks are used in each scale of the encoder and decoder network. Details of the residual block are presented in Fig. 4.

In order to help propagate information between the different branches, each output feature maps of the EDR subnet are up-sampled to the same size of the output feature maps in the next branch, and then fused by the sum operation in the proposed network.

C. *EDM for Edge Prediction*

Although the PSP structure can extract multiscale features effectively, it may also lead to inaccurate shadow pixels around the boundaries. This is because the down-sampling process can extract the high-level abstract features, but it also results in spatial information loss. It has been proved that edge information can provide complementary cues for image segmentation [43]–[46]. Therefore, a multitask learning framework is proposed by

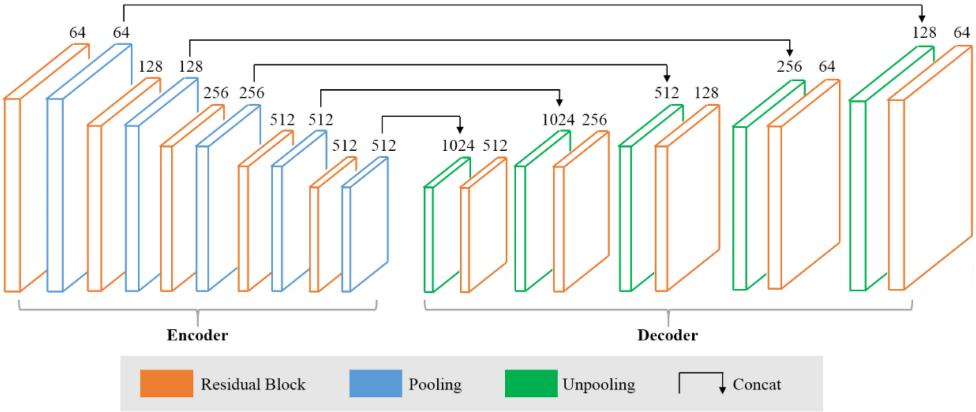


Fig. 3. Illustration of the EDR subnet.

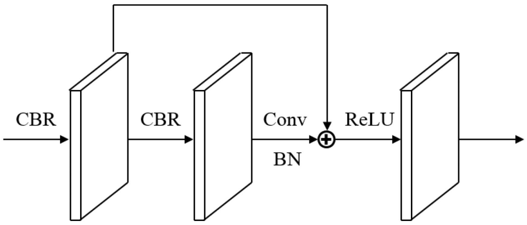


Fig. 4. Details of the residual block.

expanding the EDM, i.e., the edge information is taken as a priori knowledge to refine the shadow detection results.

The EDM reuses features from the PSP structure and incorporates them for edge prediction, effectively reducing the number of parameters and improving the flow of information throughout the network, making the network easy to train. Specifically, the EDM fuses the shallow features containing detail information and the deep features containing abstract information, and then reduces the number of concatenated feature maps with a CBR block. The final shadow edge prediction is generated by a convolutional layer with 1 × 1 kernel size and a sigmoid function, rescaling the values to between 0 and 1.

D. EMI for Feature Fusion

In order to combine multiscale semantics with edge information and local appearance, the EMI is conducted. The edge features are first incorporated with the multibranch features to optimize the shadow edges in the shadow mask. The deeply supervised mechanism [47], [48] is then adopted for each edge-aware branch to guide the network training and predict multibranch output images. Finally, these deeply supervised branches are integrated with the shallow features, and the salient shadow detection result is generated, followed by a convolutional layer with 1 × 1 kernel size and a sigmoid function, rescaling the values to between 0 and 1.

E. Loss Functions

As the multitask learning framework is applied to realize salient shadow region detection and shadow edge detection simultaneously, the total loss of the network contains two parts

$$L_{\text{total}} = L_{\text{region}} + \omega L_{\text{edge}} \tag{1}$$

where L_{total} represents the loss function of the whole network; and L_{region} and L_{edge} are the loss functions of the salient shadow region detection and shadow edge detection, respectively. ω is the weight parameter for balancing these two losses.

For the salient shadow region detection task, three edge-aware branches are introduced and a final fusion layer is also added. Therefore, the loss function of the salient shadow region detection task can be expressed as

$$L_{\text{region}} = L_{\text{region}\_F} + L_{\text{region}\_EB} \tag{2}$$

where L_{region_F} and L_{region_EB} are the loss functions of the final fusion layer and edge-aware branches, respectively. L_{region_EB} can be further expressed as follows:

$$L_{\text{region}\_EB} = \sum_{m=1}^{M} \alpha_m l^m_{\text{region}\_EB} \tag{3}$$

where α_m is the weight of the mth edge-aware branch, and $l^m_{\text{region}\_EB}$ is the loss function of the mth edge-aware branch. M is equal to 3 here. Note that the weights ω and α_m are empirically set to 1.

As the salient shadow region detection task and shadow edge detection task are both binary classification problems, binary cross-entropy loss is utilized for L_{edge}, L_{region_F}, and $l^m_{\text{region}\_EB}$ to guide the network training, and defined as

$$L = -\frac{1}{N}\sum_{i=1}^{N}[y_i log(\hat{y}_i) + (1 - y_i) log(1 - \hat{y}_i)] \tag{4}$$

where N is the total pixel number. y_i and $\hat{y}_i \in [0, 1]$ are the label value and the predicted value of the ith pixel, respectively. $y_i = 1$ if the pixel is in a shadow region or shadow edge, and otherwise, $y_i = 0$.

By minimizing the loss function, the predicted outputs become consistent with the ground-truth labels. In the test stage, the final output of ESPFNet is a probability map ranging from 0 to 1. Otsu's thresholding method [49] is then applied to automatically determine the suitable threshold, and segment the final probability into a binary mask.

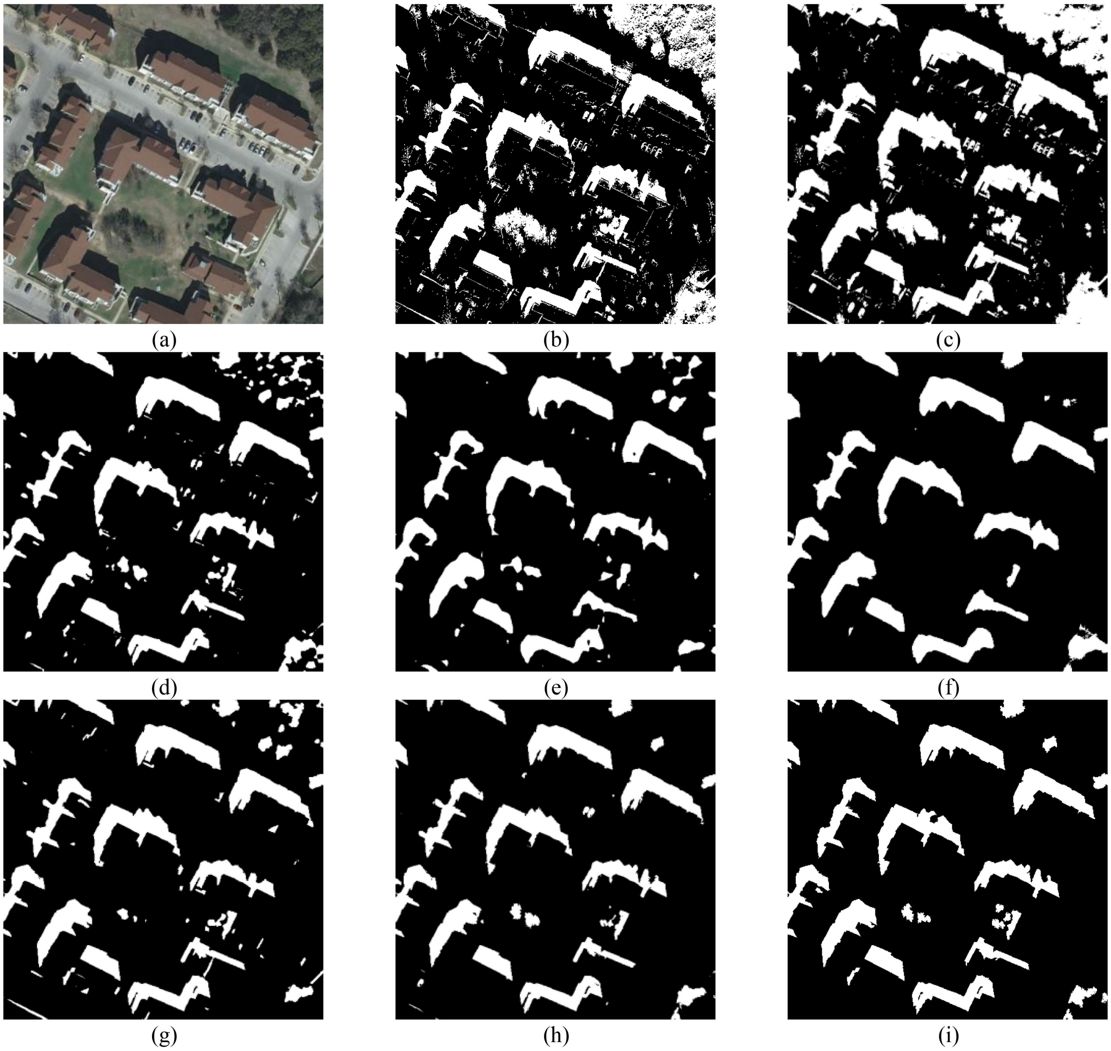


Fig. 5. Comparison of shadow detection results by the different methods in image #1. (a) Aerial image. (b) SRS. (c) ERWSD. (d) U-Net. (e) PSPNet. (f) BDRAR. (g) DSSDNet. (h) ESPFNet. (i) Salient shadow ground truth.

TABLE I
AVERAGE SALIENT SHADOW DETECTION ACCURACY COMPARISONS FOR THE SALIENT TEST IMAGES

Index	SRS	ERWSD	U-Net	PSPNet	BDRAR	DSSDNet	ESPFNet
F-score	71.30%	75.32%	87.98%	88.32%	88.28%	91.70%	**92.04%**

IV. EXPERIMENTS

In this section, the dataset, implementation details, compared methods and evaluation metrics used in this study are first introduced. Qualitative and quantitative comparisons are then presented. Finally, the effectiveness of the EDM is verified by the ablation analysis.

A. *Dataset Description and Implementation Details*

The performance of the proposed network is evaluated on the AISD [6], which can be freely downloaded.[1] This dataset consists of 514 aerial images and the associated shadow mask at a 30-cm spatial resolution. Specifically, 80% of the AISD dataset is used as the training data, and the validation data and test data each count for 10%. The ground-truth shadow edge maps are generated by Canny operator [50] applied on the shadow mask. Data augmentation is adopted for the training data to prevent network overfitting and improve the effectiveness. The training data are cropped to generate 11836 patches in total with a fixed size of 256 × 256. It should be noted that the shadow masks of the test set have been carefully adjusted into salient shadow masks, according to the rules defined in Section II.

TABLE II
AVERAGE SHADOW DETECTION ACCURACY FOR THE SALIENT SHADOW TEST IMAGES

Index	ESPFNet-WoEDM	ESPFNet-WoEFF	ESPFNet
F-score	90.37%	90.68%	**92.04%**

[1] [Online]. Available: https://github.com/RSrscoder/AISD

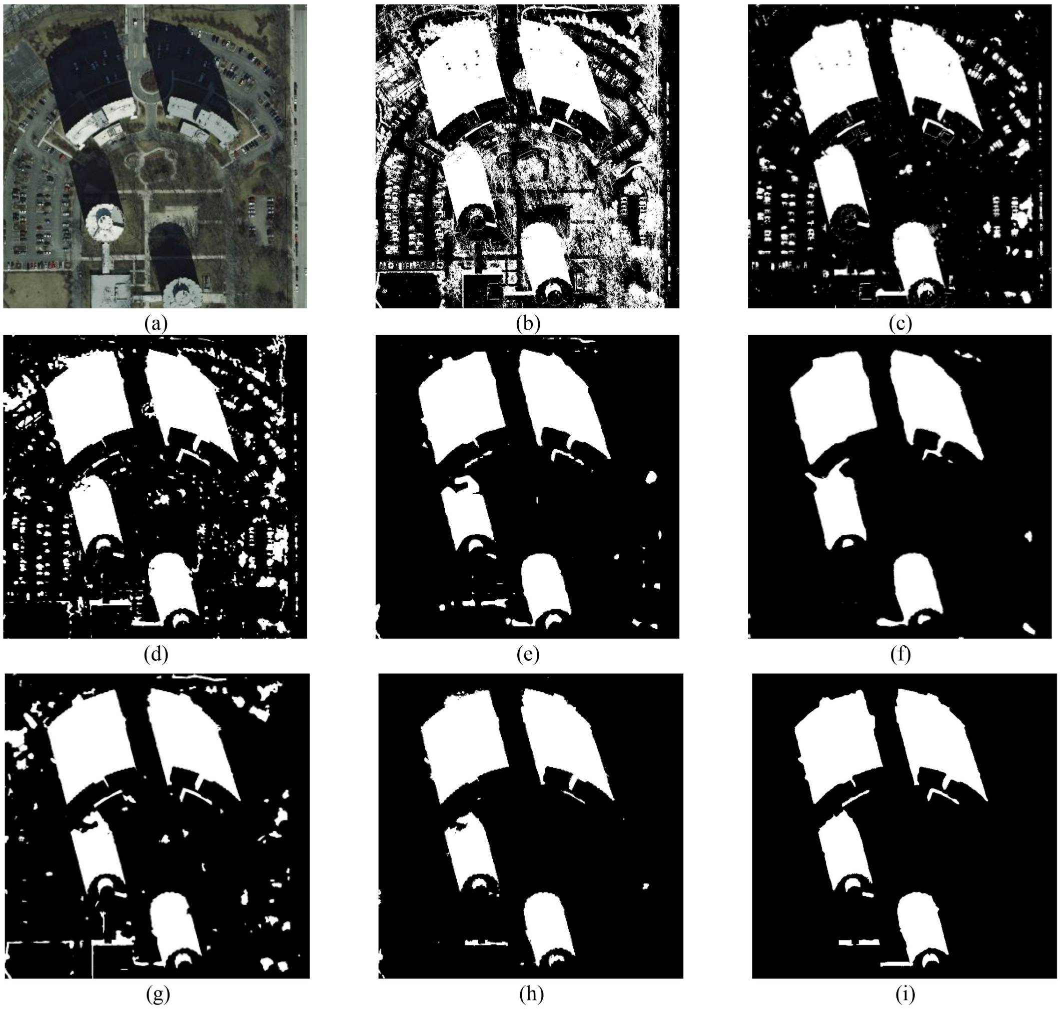


Fig. 6. Comparison of shadow detection results by the different methods in image #2. (a) Aerial image. (b) SRS. (c) ERWSD. (d) U-Net. (e) PSPNet. (f) BDRAR. (g) DSSDNet. (h) ESPFNet. (i) Salient shadow ground truth.

For the implementation details of the proposed ESPFNet, the kernel size of the convolutional layers is 3 × 3 except for the deeply supervised layers and prediction layers with 1 × 1. The stochastic gradient descent is used to optimize the network with the weight decay of 0.0005 and the momentum of 0.9. Besides, the training epochs are 100, the batch size is 10 and the learning rate is initialized to 10^{-3}, with a “poly” policy used for the learning rate decay. The proposed network is implemented using PyTorch framework [51].

B. Compared Methods and Evaluation Metrics

To evaluate the superiority of the proposed ESPFNet, six representative methods are utilized for comparison, including two traditional shadow detection methods, i.e., spectral ratioing segmentation (SRS) [3] and extended random walker based shadow detection (ERWSD) [7], and four deep learning based methods, U-Net [52], the pyramid scene parsing network (PSPNet) [39], the bidirectional feature pyramid network with recurrent attention residual (BDRAR) modules, [53] and the DSSDNet [6]. U-Net and PSPNet are two widely used networks for semantic segmentation and the shadow detection can also be regarded as a semantic segmentation problem. BDRAR is designed for shadow detection in close-shot images, and DSSDNet is proposed specifically for shadow detection in aerial images. And all the deep learning based methods were trained on the same training data with the same parameter settings.

To analyze the compared methods quantitatively, a widely used quantitative metric, i.e., the F-score, is adopted with the value ranging from 0 to 1 [1], [8]. The closer the F-score is to 1, the better performance of the methods. Besides, box plot is introduced, which is an excellent way to provide a visual representation of data distributions. And the receiver operating characteristic (ROC) curves and precision-recall (PR) curves are also plotted to analyze the binary classification accuracies of the compared methods.

C. Qualitative Comparison

Four representative shadow images are presented to compare the performance of compared methods in Figs. 5–8. These images contain various land covers, shadow sizes and shadow shapes. The experiments results show that the SRS can detect

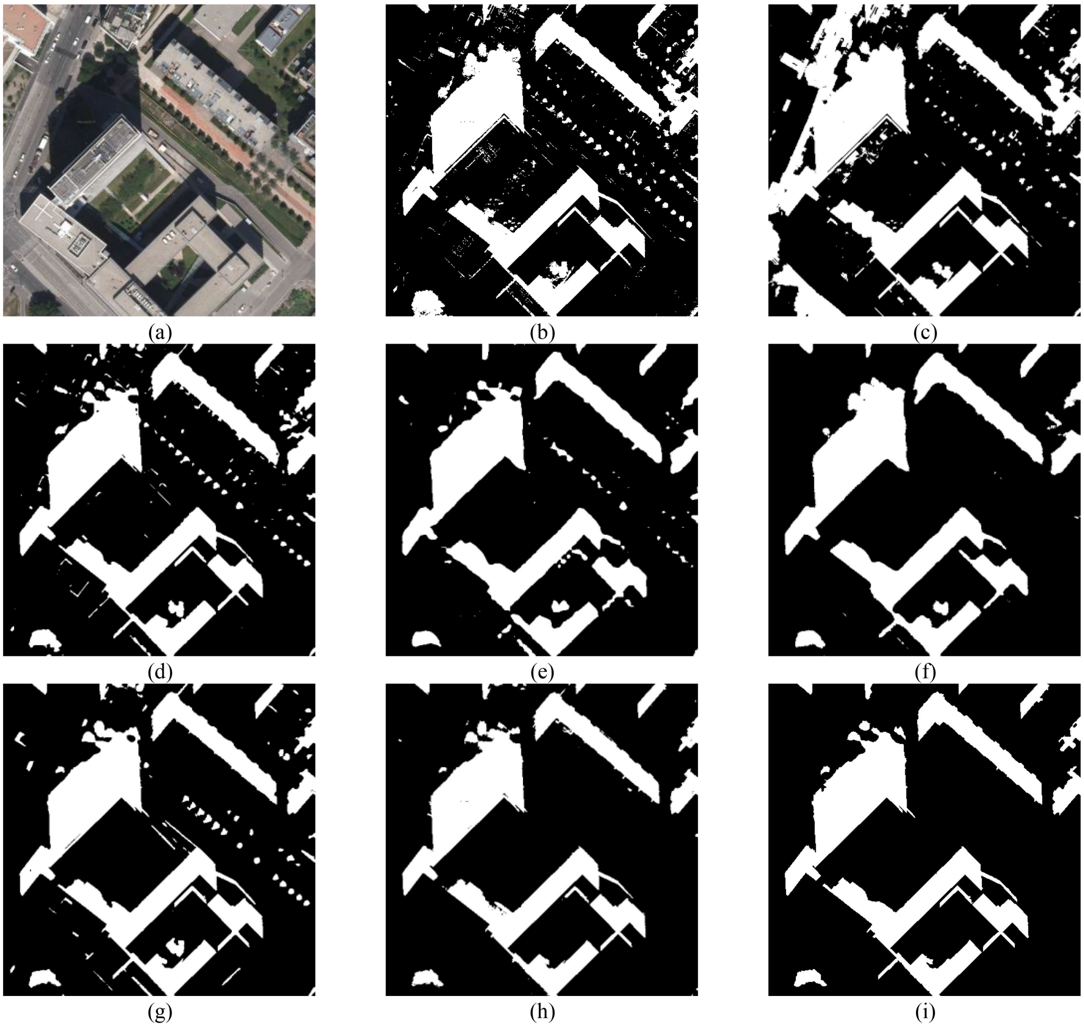


Fig. 7. Comparison of shadow detection results by the different methods in image #3. (a) Aerial image. (b) SRS. (c) ERWSD. (d) U-Net. (e) PSPNet. (f) BDRAR. (g) DSSDNet. (h) ESPFNet. (i) Salient shadow ground truth.

all shadow regions, but also falsely recognize the vegetation as shadows in Figs. 5–8(b). Comparing to the SRS, the ERWSD can exclude some disturbance of vegetation, but the shadow detection results are still noisy in Figs. 5–8(c). Because these two methods are mainly based on the thresholding segmentation, which may result in sunlit dark objects being falsely recognized as shadow, and shadowed bright objects being falsely recognized as nonshadow. The U-Net can generate more accurate results than the above two traditional methods, but the detailed shadow regions have been detected which are not salient, and a part of dark non-shadow regions are mistakenly identified as shadows in Figs. 5–8(d). For the results in Figs. 5–8(e) and (f), the PSPNet and BDRAR can detect the salient shadow regions effectively, but the detected shadow boundaries of the buildings are smooth and irregular, and are not consistent with the real scene. The results of the DSSDNet are accurate and regular, but some detailed shadow regions are also detected in Figs. 5–8(g). The proposed ESPFNet can not only detect the salient shadow regions accurately, but it also maintains regular shadow boundaries, as shown in Figs. 5–8(h).

D. Quantitative Comparison

A quantitative comparison was also carried out. The average salient shadow detection accuracy comparisons for the salient test images are given in Table I, and the proposed ESPFNet achieves highest F-score value, i.e., 92.04%, among all the compared methods. The F-score values of the SRS and ERWSD are 71.30% and 75.32%, respectively, lower than 80%, while the U-Net, PSPNet, BDRAR, and DSSDNet are between 87%–92%. Therefore, the ESPFNet has significant accuracy improvements for the traditional methods, and shows accuracy advantages for the compared deep learning based methods as well.

Besides, the box plots of the F-score distributions for the compared methods on the salient test images are shown in Fig. 9. The box sizes of SRS and ERWSD are large, which means the detection results are unstable, while that of the deep learning based methods are smaller, which are more stable than the traditional methods. And the box location and box size of ESPFNet are both highest and smallest, which verifies the effectiveness of ESPFNet.

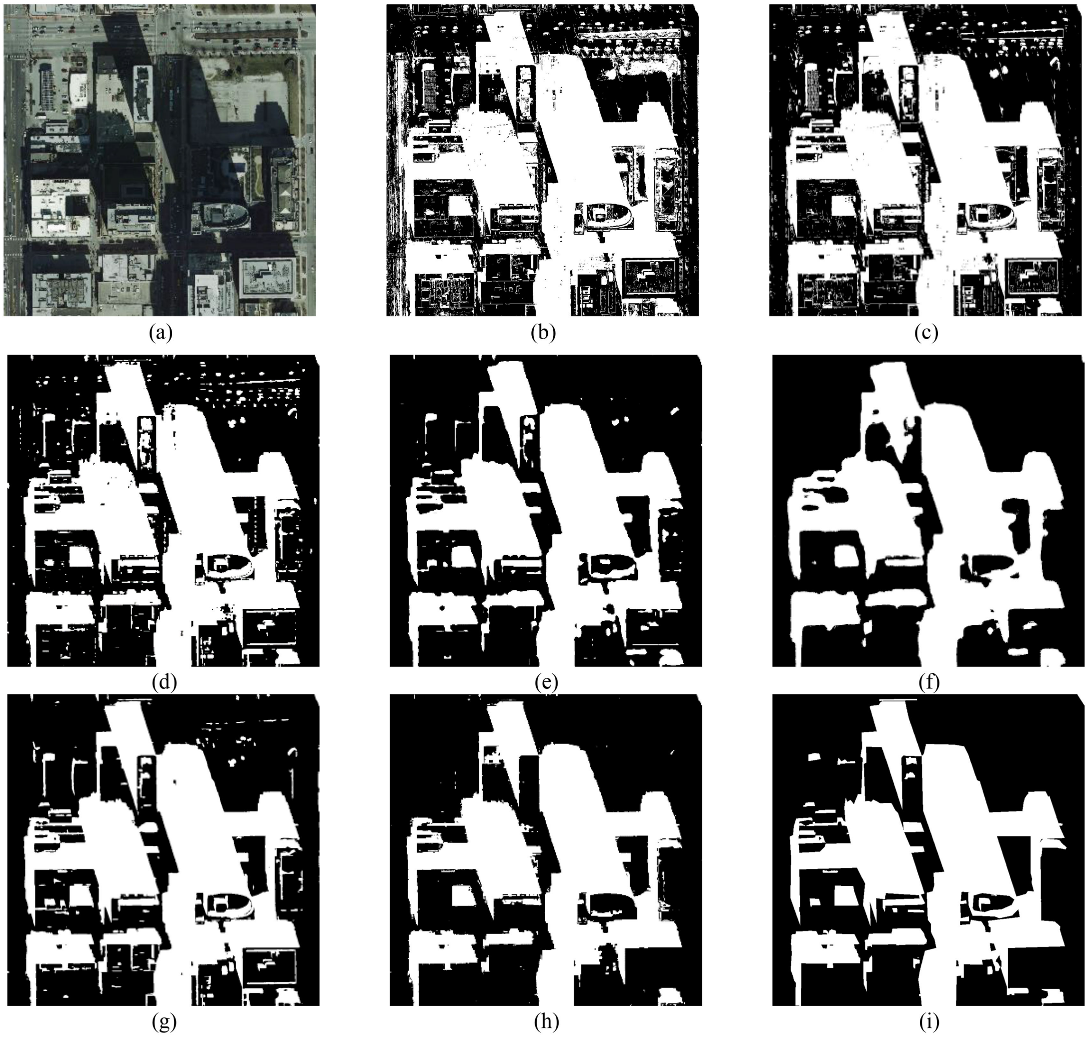


Fig. 8. Comparison of shadow detection results by the different methods in image #4. (a) Aerial image. (b) SRS. (c) ERWSD. (d) U-Net. (e) PSPNet. (f) BDRAR. (g) DSSDNet. (h) ESPFNet. (i) Salient shadow ground truth.

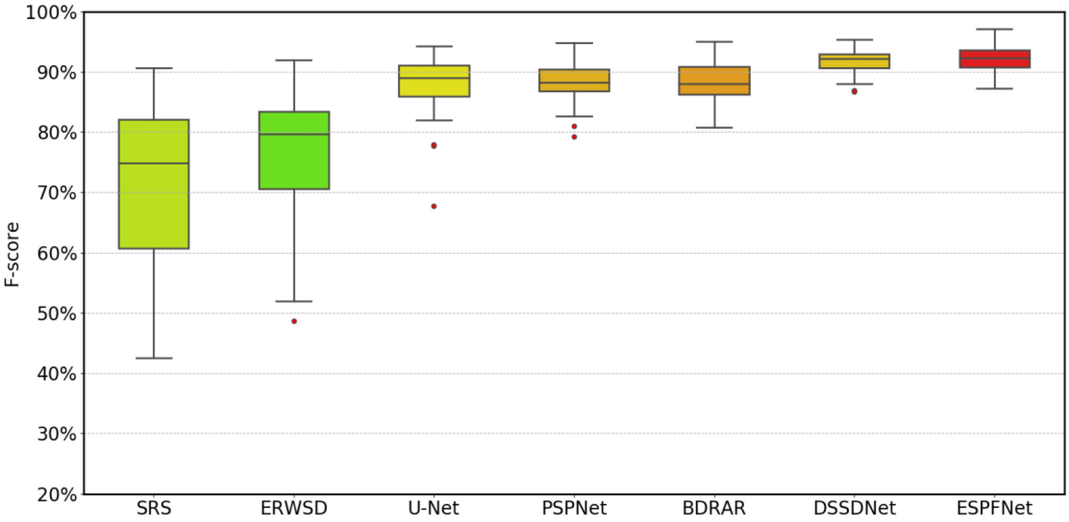
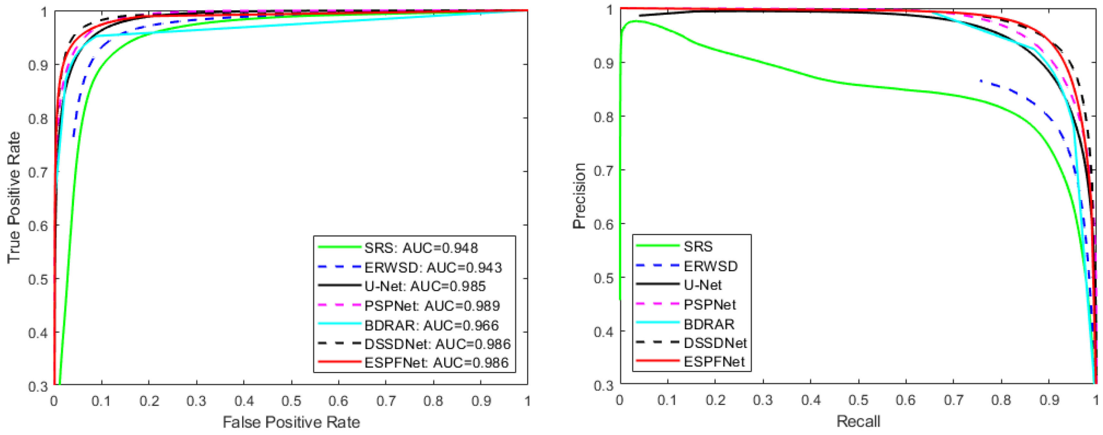
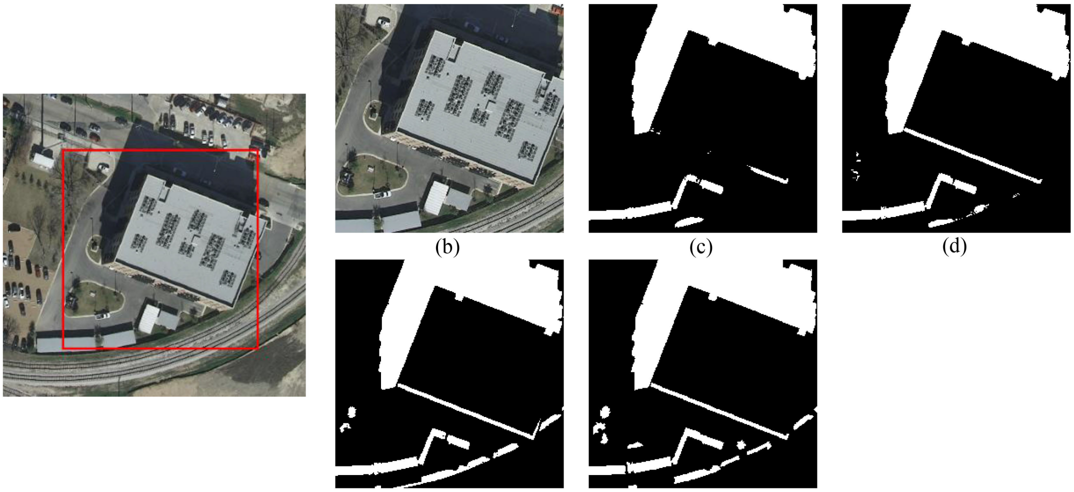


Fig. 9. Box plots of the F-score distributions for the compared methods on the salient test images.



(a) (b)

Fig. 10. ROC and PR curves of the compared methods on the salient test images. (a) ROC curves. (b) PR curves.



(a) (b) (c) (d) (e) (f)

Fig. 11. Visual comparison of a detailed region (red box) for ESPFNet-WoEDM, ESPFNet-WoEFF, and ESPFNet. (a) Shadow image. (b) Detailed region cropped from (a). (c) ESPFNet-WoEDM. (d) ESPFNet-WoEFF. (e) ESPFNet. (f) Salient shadow ground truth.

ROC and PR curves of the compared methods on the salient test images are presented in Fig. 10. The ROC curves of the proposed ESPFNet method (red solid curve) and DSSDNet (black dashed curve) are close to each other, and are better than the other methods in Fig. 10(a). The AUC scores of ESPFNet and DSSDNet are both 0.986, and are higher than the scores of the other methods, except for PSPNet. It should be noted that the AUC score of PSPNet is 0.989, which is the highest score, but the ROC curve of PSPNet is not the closest to the upper left corner. Besides, the PR curves are shown in Fig. 10(b), which also indicates the ESPFNet and DSSDNet achieve superior performances on the salient shadow test images.

E. Ablation Analysis

As the EDM is the main innovation of the proposed ESPFNet method, two baseline networks were evaluated, i.e., ESPFNet without the EDM (ESPFNet-WoEDM) and ESPFNet without the edge feature fusion (ESPFNet-WoEFF), to demonstrate the effectiveness of the EDM. It should be noted that ESPFNet-WoEDM means that the EDM has been completely removed, while ESPFNet-WoEFF means that the EDM still exists, but the edge features are not fused with the multibranch features.

Visual comparisons for ESPFNet-WoEDM, ESPFNet-WoEFF, and ESPFNet are presented in Figs. 11 and 12. Parts of shadow regions are enlarged, and labeled with the red boxes, as shown in Figs. 11(a) and 12(a). It can be seen that ESPFNet-WoEDM has left out some salient shadow regions in Fig. 11(c), and it falsely recognizes the dark road as shadow in Fig. 12(c). ESPFNet-WoEFF can partially correct the problems of the missing salient shadow regions and the false detection of dark road in Figs. 11(d) and 12(d), respectively. Comparing the ESPFNet-WoEDM and ESPFNet-WoEFF, ESPFNet can not only detects the salient shadows completely, but it also

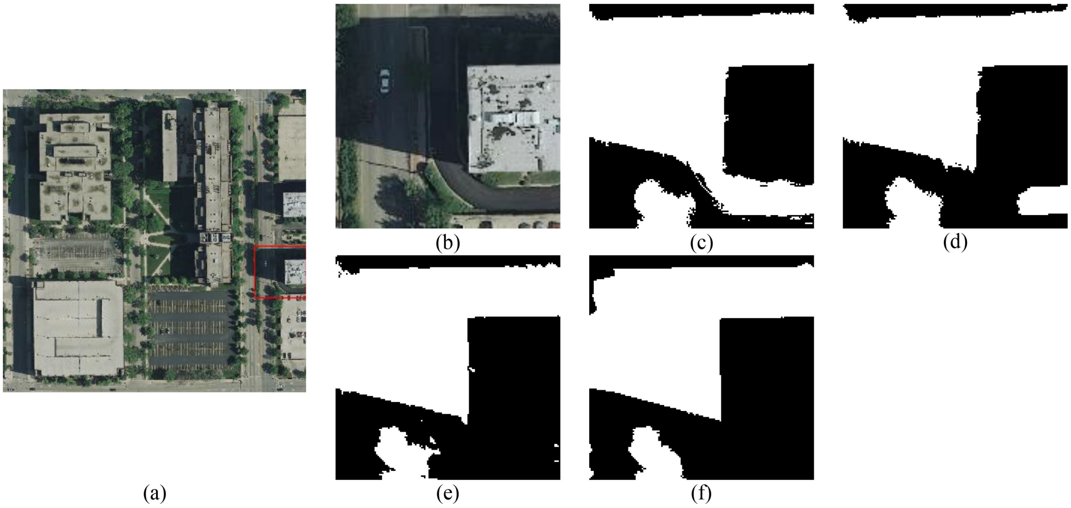


Fig. 12. Visual comparison of a detailed region (red box) for ESPFNet-WoEDM, ESPFNet-WoEFF, and ESPFNet. (a) Shadow image. (b) Detailed region cropped from (a). (c) ESPFNet-WoEDM. (d) ESPFNet-WoEFF. (e) ESPFNet. (f) Salient shadow ground truth.

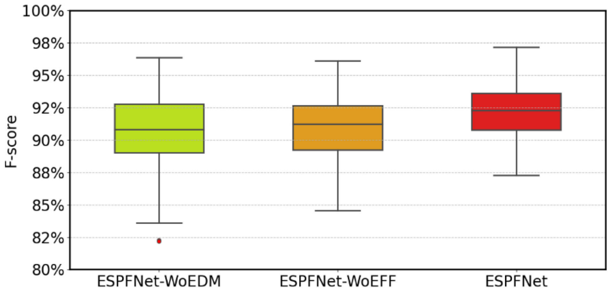


Fig. 13. Box plots of the F-score distributions for the salient test images.

effectively excludes the disturbance of the dark road in Figs. 11(e) and 12(e), which demonstrates the effectiveness of the EDM and the edge feature fusion with multibranch features.

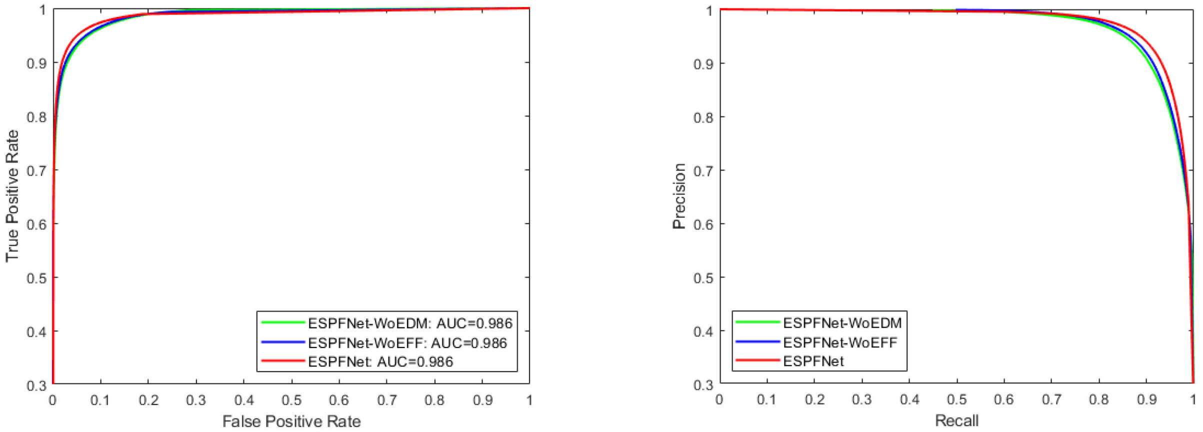
The F-scores of ESPFNet-WoEDM, ESPFNet-WoEFF, and ESPFNet obtained with the salient test images are given in Table II. It shows that the F-score of ESPFNet-WoEFF is 90.68% higher than the F-score of ESPFNet-WoEDM (90.37%), demonstrating the effectiveness of the EDM. The F-score of ESPFNet is 92.04%, higher than the F-score of ESPFNet-WoEFF (90.68%), which means that the edge feature fusion with multibranch features is helpful for salient shadow detection. The box plots of the F-score distributions for the salient test images are shown in Fig. 13. The box size of ESPFNet-WoEFF is smaller than that of ESPFNet-WoEDM, and the box location of ESPFNet-WoEFF is also higher than that of ESPFNet-WoEDM. The box size and box location of ESPFNet are both the smallest and the highest.

Further, ROC and PR curves are provided in Fig. 14 to compare ESPFNet with ESPFNet-WoEDM and ESPFNet-WoEFF. The ROC curves indicate that ESPFNet (red curve) is closest to the upper left corner, while ESPFNet-WoEFF (blue curve) is slightly better than ESPFNet-WoEDM (green curve) in Fig. 14(a). Moreover, the PR curves show that ESPFNet is again closest to the upper right corner, while ESPFNet-WoEFF shows a better performance than ESPFNet-WoEDM in Fig. 14(b). Overall, the proposed ESPFNet can improve the salient shadow detection accuracy significantly, compared to ESPFNet-WoEDM and ESPFNet-WoEFF.

V. Discussion

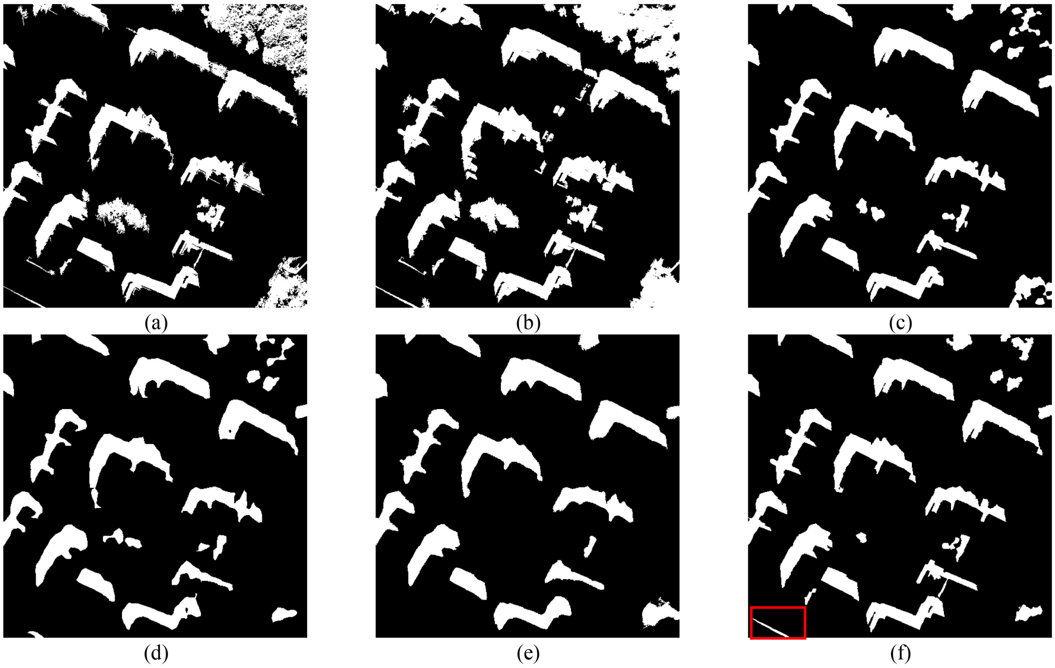
Since the large area is the most important feature for salient shadows, it should be investigated whether salient shadow detection results can be achieved through some simple postprocessing of the compared methods. Therefore, a morphological operation i.e., small shadow region removal, was conducted to verify this point. Through a trial-and-error test, the area threshold value was set to 130, which can generate accurate salient shadow detection results for DSSDNet, because the original results of DSSDNet are the most accurate among the compared methods, but with small shadow regions.

The results of two experiments in postprocessing for the compared methods are presented in Figs. 15 and 16 (from Figs. 5 and 7). It is clear that the results of the compared methods in Figs. 15 and 16 are improved significantly when compared to the results in Figs. 5 and 7. The salient shadow regions remain while the small shadow regions are excluded. However, it should also be noted that, for the results of SRS and ERWSD shown in Figs. 15(a) and (b) and 16(a) and (b), the falsely detected vegetation regions are not removed through this postprocessing, because some of the vegetation regions are too large. For the results of U-Net shown in Figs. 15(c) and 16(c), most of the non-salient shadow regions are removed, but some small regions are still connected with the building shadows, which are difficult to exclude with the area threshold. For the results of PSPNet and BDRAR in Figs. 15(d) and (e) and 16(d) and (e), the detected shadow regions are salient and the postprocessing is less effective on these results, but the problem is that the detected shadow boundaries are inaccurate, as mentioned in Section IV-C. For the results of DSSDNet in Figs. 15(f) and 16(f), the results are improved a lot, and DSSDNet can detect the salient shadow



(a) (b)

Fig. 14. ROC and PR curves of ESPFNet-WoEDM ESPFNet-WoEFF, and ESPFNet on the salient test images. (a) ROC curves. (b) PR curves.



(a) (b) (c) (d) (e) (f)

Fig. 15. Postprocessing for the compared methods from Fig. 5 with the area threshold value equal to 130. (a) SRS. (b) ERWSD. (c) U-Net. (d) PSPNet. (e) BDRAR. (f) DSSDNet.

TABLE III
Average Accuracies Obtained With the Salient Shadow Test Images, With the Area Threshold Value Equal to 130 (Except for ESPFNet)

Index	SRS	ERWSD	U-Net	PSPNet	BDRAR	DSSDNet	ESPFNet
F-score	73.69%	77.50%	89.01%	88.42%	88.23%	**92.15%**	92.04%

regions accurately while retaining regular shadow boundaries, but the slender shadow regions, as labeled with the red box, are difficult to remove, because these shadows also have a relatively large area.

The average accuracies obtained by the different methods with the salient shadow test images are given in Table III. Most of the compared methods show an improved accuracy, except for BDRAR with a 0.05% decrease, because the original results of BDRAR detect most of the salient shadow regions. The F-score of DSSDNet is 92.15% after small shadow region removal, which is higher than ESPFNet, with 92.04%.

Although the postprocessing can improve the salient shadow detection accuracy of the compared methods, an optimal area

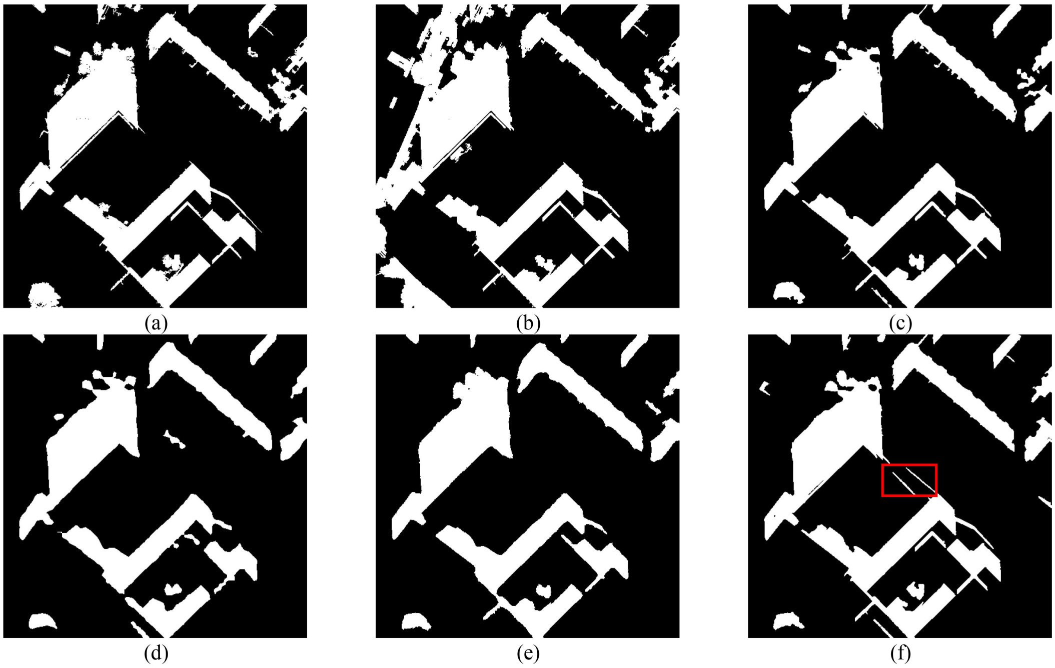


Fig. 16. Postprocessing for the compared methods from Fig. 7 with the area threshold value equal to 130. (a) SRS. (b) ERWSD. (c) U-Net. (d) PSPNet. (e) BDRAR. (f) DSSDNet.

threshold is obligatory for the postprocessing, which requires trial-and-error interaction. In contrast, salient shadows can be detected directly by ESPFNet. Overall, the proposed ESPFNet is a more effective and efficient way to obtain the salient shadows in aerial remote sensing images.

VI. Conclusion

In this article, the concept of salient shadows in remote sensing images has been presented and the novel ESPFNet under a multitask learning framework has been proposed to solve this problem in aerial remote sensing images. The key idea is the use of a multitask framework to achieve salient shadow detection by simultaneously detecting the shadow regions and shadow edges. ESPFNet is made up of three components, i.e., the PSP structure, the EDM, and the EMI.

The qualitative and quantitative analyses demonstrated that the proposed ESPFNet achieved competitive salient shadow detection performance, compared with traditional shadow detection methods and deep learning based methods. In addition, the ablation analysis has also been conducted, and verified the effectiveness of the EDM and the edge feature fusion with multibranch features. Furthermore, postprocessing (small shadow region removal) was performed to further highlight the advantage of the proposed ESPFNet.

Acknowledgment

The authors appreciate the editors and anonymous reviewers for their valuable suggestions.

References

[1] K. R. M. Adeline, M. Chen, X. Briottet, S. K. Pang, and N. Paparoditis, "Shadow detection in very high spatial resolution aerial images: A comparative study," *ISPRS J. Photogramm. Remote Sens.,* vol. 80, pp. 21–38, 2013.

[2] H. Li, L. Zhang, and H. Shen, "An adaptive nonlocal regularized shadow removal method for aerial remote sensing images," *IEEE Trans. Geosci. Remote Sens.,* vol. 52, no. 1, pp. 106–120, Jan. 2014.

[3] V. J. D. Tsai, "A comparative study on shadow compensation of color aerial images in invariant color models," *IEEE Trans. Geosci. Remote Sens.,* vol. 44, no. 6, pp. 1661–1671, Jun. 2006.

[4] S. Luo, H. Shen, H. Li, and Y. Chen, "Shadow removal based on separated illumination correction for urban aerial remote sensing images," *Signal Process,* vol. 165, pp. 197–208, 2019.

[5] Z. Li, H. Shen, H. Li, G. Xia, P. Gamba, and L. Zhang, "Multifeature combined cloud and cloud shadow detection in GaoFen-1 wide field of view imagery," *Remote Sens. Environ.,* vol. 191, pp. 342–358, 2017.

[6] S. Luo, H. Li, and H. Shen, "Deeply supervised convolutional neural network for shadow detection based on a novel aerial shadow imagery dataset," *ISPRS J. Photogramm. Remote Sens.,* vol. 167, pp. 443–457, 2020.

[7] X. Kang, Y. Huang, S. Li, H. Lin, and J. A. Benediktsson, "Extended random walker for shadow detection in very high resolution remote sensing images," *IEEE Trans. Geosci. Remote Sens.,* vol. 56, no. 2, pp. 867–876, Feb. 2018.

[8] G. F. Silva, G. B. Carneiro, R. Doth, L. A. Amaral, and D. F. G. d. Azevedo, "Near real-time shadow detection and removal in aerial motion imagery application," *ISPRS J. Photogramm. Remote Sens.,* vol. 140, pp. 104–121, 2018.

[9] G. Sun *et al.*, "Combinational shadow index for building shadow extraction in urban areas from Sentinel-2A MSI imagery," *Int. J. Appl. Earth Observ. Geoinf.,* vol. 78, pp. 53–65, 2019.

[10] G. Liasis and S. Stavrou, "Satellite images analysis for shadow detection and building height estimation," *ISPRS J. Photogramm. Remote Sens.,* vol. 119, pp. 437–450, 2016.

[11] Y. Li, P. Gong, and T. Sasagawa, "Integrated shadow removal based on photogrammetry and image analysis," *Int. J. Remote Sens.,* vol. 26, no. 18, pp. 3911–3929, 2005.

[12] G. Tolt, M. Shimoni, and J. Ahlberg, "A shadow detection method for remote sensing images using VHR hyperspectral and LIDAR data," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, 2011, pp. 4423–4426.

[13] Q. Wang, L. Yan, Q. Yuan, and Z. Ma, "An automatic shadow detection method for VHR remote sensing orthoimagery," *Remote Sens.,* vol. 9, no. 5, pp. 469, 2017.

[14] P. M. Dare, "Shadow analysis in high-resolution satellite imagery of urban areas," *Photogramm. Eng. Remote Sens.,* vol. 71, no. 2, pp. 169–177, 2005.

[15] V. Arévalo, J. González, and G. Ambrosio, "Shadow detection in colour high-resolution satellite images," *Int. J. Remote Sens.,* vol. 29, no. 7, pp. 1945–1963, 2008.

[16] K. L. Chung, Y. R. Lin, and Y. H. Huang, "Efficient shadow detection of color aerial images based on successive thresholding scheme," *IEEE Trans. Geosci. Remote Sens.,* vol. 47, no. 2, pp. 671–682, Feb. 2009.

[17] H. Song, B. Huang, and K. Zhang, "Shadow detection and reconstruction in high-resolution satellite images via morphological filtering and example-based learning," *IEEE Trans. Geosci. Remote Sens.,* vol. 52, no. 5, pp. 2545–2554, May 2014.

[18] H. Zhang, K. Sun, and W. Li, "Object-Oriented shadow detection and removal from urban high-resolution remote sensing images," *IEEE Trans. Geosci. Remote Sens.,* vol. 52, no. 11, pp. 6972–6982, Nov. 2014.

[19] Y. Jin, W. Xu, D. Shao, X. He, and X. Zhang, "Object-Oriented automatic and accurate shadow detection for very high spatial resolution satellite images," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, 2019, pp. 1458–1461.

[20] L. Lorenzi, F. Melgani, and G. Mercier, "A complete processing chain for shadow detection and reconstruction in VHR images," *IEEE Trans. Geosci. Remote Sens.,* vol. 50, no. 9, pp. 3440–3452, Sep. 2012.

[21] Y. Zhang *et al.*, "Recurrent shadow attention model (RSAM) for shadow removal in high-resolution urban land-cover mapping," *Remote Sens. Environ.,* vol. 247, 2020, Art. no. 111945.

[22] T. Liu *et al.*, "Learning to detect a salient object," *IEEE Trans. Pattern Anal. Mach. Intell.,* vol. 33, no. 2, pp. 353–367, Feb. 2011.

[23] A. Borji, M. Cheng, H. Jiang, and J. Li, "Salient object detection: A benchmark," *IEEE Trans. Image Process.,* vol. 24, no. 12, pp. 5706–5722, Dec. 2015.

[24] R. Zhao, W. Ouyang, H. Li, and X. Wang, "Saliency detection by multi-context deep learning," in *Proc. IEEE Conf. Comput. Vis. Pattern*, 2015, pp. 1265–1274.

[25] Q. Hou, M. Cheng, X. Hu, A. Borji, Z. Tu, and P. H. S. Torr, "Deeply supervised salient object detection with short connections," *IEEE Trans. Pattern Anal. Mach. Intell.,* vol. 41, no. 4, pp. 815–828, Apr. 2019.

[26] X. Zhang, T. Wang, J. Qi, H. Lu, and G. Wang, "Progressive attention guided recurrent network for salient object detection," in *Proc. IEEE Conf. Comput. Vis. Pattern*, 2018, pp. 714–722.

[27] L. Wang, R. Chen, L. Zhu, H. Xie, and X. Li, "Deep Sub-region network for salient object detection," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 31, no. 2, pp. 728–741, Feb. 2021.

[28] M. Zhang *et al.* "LFNet: Light field fusion network for salient object detection," *IEEE Trans. Image Process.,* vol. 29, pp. 6276–6287, 2020.

[29] T. Liu, J. Sun, N. Zheng, X. Tang, and H. Shum, "Learning to detect a salient object," in *Proc. IEEE Conf. Comput. Vis. Pattern*, 2007, pp. 1–8.

[30] A. Borji, "What is a salient object? A dataset and a baseline model for salient object detection," *IEEE Trans. Image Process.,* vol. 24, no. 2, pp. 742–756, Feb. 2015.

[31] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards real-time object detection with region proposal networks," *IEEE Trans. Pattern Anal. Mach. Intell.,* vol. 39, no. 6, pp. 1137–1149, Jun. 2017.

[32] C. F. Flores, A. Gonzalez-Garcia, J. van de Weijer, and B. Raducanu, "Saliency for fine-grained object recognition in domains with scarce training data," *Pattern Recognit.,* vol. 94, pp. 62–73, 2019.

[33] R. Cong, J. Lei, H. Fu, F. Porikli, Q. Huang, and C. Hou, "Video saliency detection via sparsity-based reconstruction and propagation," *IEEE Trans. Image Process.,* vol. 28, no. 10, pp. 4819–4831, Oct. 2019.

[34] L. Shao and M. Brady, "Specific object retrieval based on salient regions," *Pattern Recognit.,* vol. 39, no. 10, pp. 1932–1948, 2006.

[35] X. Qin, Z. Zhang, C. Huang, C. Gao, M. Dehghan, and M. Jagersand, "BASNet: Boundary-aware salient object detection," in *Proc. IEEE Conf. Comput. Vis. Pattern*, 2019, pp. 7471–7481.

[36] M. Feng, H. Lu, and E. Ding, "Attentive feedback network for boundary-aware salient object detection," in *Proc. IEEE Conf. Comput. Vis. Pattern*, 2019, pp. 1623–1632.

[37] S. Lazebnik, C. Schmid, and J. Ponce, "Beyond bags of features: Spatial pyramid matching for recognizing natural scene categories," in *Proc. IEEE Conf. Comput. Vis. Pattern*, 2006, vol. 2, pp. 2169–2178.

[38] K. He, X. Zhang, S. Ren, and J. Sun, "Spatial pyramid pooling in deep convolutional networks for visual recognition," *IEEE Trans. Pattern Anal. Mach. Intell.,* vol. 37, no. 9, pp. 1904–1916, Sep. 2015.

[39] H. Zhao, J. Shi, X. Qi, X. Wang, and J. Jia, "Pyramid scene parsing network," in *Proc. IEEE Conf. Comput. Vis. Pattern*, 2017, pp. 6230–6239.

[40] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," in *Proc. Int. Conf. Mach. Learn.,* 2015, vol. 37, pp. 448–456.

[41] V. Nair and G. E. Hinton, "Rectified linear units improve restricted Boltzmann machines," in *Proc. Int. Conf. Mach. Learn.*, 2010, pp. 807–814.

[42] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern*, 2016, pp. 770–778.

[43] L. Chen, J. T. Barron, G. Papandreou, K. Murphy, and A. L. Yuille, "Semantic image segmentation with task-specific edge detection using CNNs and a discriminatively trained domain transform," in *Proc. IEEE Conf. Comput. Vis. Pattern*, 2016, pp. 4545–4554.

[44] D. Marmanis, K. Schindler, J. D. Wegner, S. Galliani, M. Datcu, and U. Stilla, "Classification with an edge: Improving semantic image segmentation with boundary detection," *ISPRS J. Photogramm. Remote Sens.,* vol. 135, pp. 158–172, 2018.

[45] H. Ding, X. Jiang, A. Q. Liu, N. M. Thalmann, and G. Wang, "Boundary-Aware feature propagation for scene segmentation," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2019, pp. 6818–6828.

[46] H. Han, Y. Chen, P. Hsiao, and L. Fu, "Using channel-wise attention for deep CNN based real-time semantic segmentation with class-aware edge information," *IEEE Trans. Intell. Transp. Syst.*, vol. 22, no. 2, pp. 1041–1051, Feb. 2021.

[47] S. Xie and Z. Tu, "Holistically-Nested edge detection," in *Proc. IEEE Int. Conf. Comput. Vis .*, 2015, pp. 1395–1403.

[48] C.-Y. Lee, S. Xie, P. Gallagher, Z. Zhang, and Z. Tu, "Deeply-supervised nets," in *Proc. Int. Conf. Artif. Intell. Statist.*, 2015, pp. 562–570.

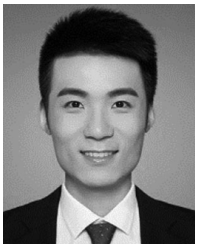
[49] N. Otsu, "A threshold selection method from gray-level histograms," *IEEE Trans. Syst. , Man, Cybern.,* vol. 9, no. 1, pp. 62–66, Jan. 1979.

[50] J. Canny, "A computational approach to edge detection," *IEEE Trans. Pattern Anal. Mach. Intell.,* vol. PAMI-8, no. 6, pp. 679–698, Nov. 1986.

[51] A. Paszke *et al.*, "Pytorch: An imperative style, high-performance deep learning library," in *Proc. Adv. Neural Inf. Process. Syst.*, 2019, pp. 8026–8037.

[52] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervention*, 2015, pp. 234–241.

[53] L. Zhu *et al.*, "Bidirectional feature pyramid network with recurrent attention residual modules for shadow detection," in *Proc. Eur. Conf. Comput. Vis.*, 2018, pp. 122–137.

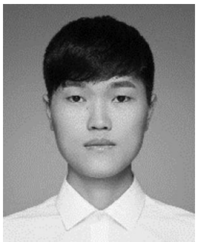


Shuang Luo received the B.S. degree in geographic information system from China University of Petroleum, Qingdao, China, in 2015. He is currently working toward the Ph.D. degree in cartography and geographic Information engineering at School of Resource and Environmental Sciences, Wuhan University, Wuhan, China.

His research interests include shadow detection and removal of high resolution remote sensing images.

Huifang Li (Member, IEEE) received the B.S. degree in geographical information science from China University of Mining and Technology, Xuzhou, China, in 2008, and the Ph.D. degree in photogrammetry and remote sensing from Wuhan University, Wuhan, China, in 2013.

She is currently an Associate Professor with the School of Resources and Environmental Science, Wuhan University. Her research interests include radiometric correction of remote sensing images, including cloud correction, shadow correction, and urban thermal environment analysis and alleviation.

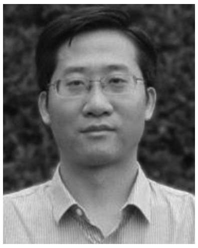


Ruzhao Zhu received the B.S. degree in physical geography and resource environment from Wuhan University, Wuhan, China, in 2019.

He is currently a R&D Engineer with KylinSoft, Changsha, China. He is dedicated to develop China's self-developed and self-reliant operating system.

Yuting Gong received the B.S. degree in geo-information science and technology from China University of Geosciences, Wuhan, China, in 2019. She is currently working toward the M.S. degree in surveying and mapping engineering at School of Resource and Environmental Sciences, Wuhan University, Wuhan, China.

Her research interests include reconstruction of land surface temperature of remote sensing data.



Huanfeng Shen (Senior Member, IEEE) received the B.S. degree in surveying and mapping engineering and the Ph.D. degree in photogrammetry and remote sensing from Wuhan University, Wuhan, China, in 2002 and 2007, respectively.

In 2007, he was with the School of Resource and Environmental Sciences (SRES), Wuhan University, where he is currently a Luojia Distinguished Professor and an Associate Dean of SRES. He was or is the PI of two projects supported by the National Key Research and Development Program of China, and six projects supported by the National Natural Science Foundation of China. He has authored more than 100 research papers in peer-reviewed international journals. His research interests include remote sensing image processing, multi-source data fusion, and intelligent environmental sensing.

Dr. Shen is a Council Member of China Association of Remote Sensing Application, Education Committee Member of Chinese Society for Geodesy Photogrammetry and Cartography, and Theory Committee Member of Chinese Society for Geospatial Information Society. He is currently a member of the Editorial Board of *Journal of Applied Remote Sensing* and *Geography and Geo-information Science*.