

УДК: 004.85

DOI: <https://doi.org/10.47813/2782-2818-2022-2-3-0312-0330>

EDN: [FFSBLZ](https://www.edn.net/FFSBLZ)



Further exploration of deep aggregation for shadow detection

Islam Md Jahidul^{1,2*}, Omar Faruq^{2,1a}

¹*School of Software Engineering, Northeastern University, No. 195, Chuangxin Road, Hunnan, Liaoning, 110819, Shenyang, P. R. China*

²*School of Communication and Information Engineering, Chongqing University of Posts and Telecommunications, 2 Chongwen Rd, Chongqing, 400065, Nan An Qu, P. R. China*

Abstract. Shadow detection is a fundamental challenge in the field of computer vision. It requires the network to understand the global semantics and local details of the image. All existing methods depend on the aggregation of the features of a multi-stage pre-trained convolution neural network, but in comparison to high-level capabilities, low-level capabilities provide less detection performance. Using low-level features not only increases the complexity of the network but also reduces its time efficiency. In this article, we propose a new shadow detector that only uses high-level features and explores the complementary information between adjacent feature layers. Experiments show that the technique in this paper can accurately detect shadows and perform well compared with the most advanced methods. The detailed experiments performed on three public shadow detection datasets, SUB, UCF, and ISTD, we demonstrate that the suggested method is efficient for detecting any sort of shadow image, which provides the maximum percentage of accuracy and stability.

Keywords: partial decoder module, adjacent feature, shadow detection.

For citation: Jahidul, Islam Md. & Faruq, O. (2022). Further Exploration of Deep Aggregation for Shadow Detection. Modern Innovations, Systems and Technologies, 2(3), 0312–0330. <https://doi.org/10.47813/2782-2818-2022-2-3-0312-0330>

INTRODUCTION

A shadow is an unseen entity that blocks light from a point of light. Behind a light source, it fills the entire 3-dimensional space. Hand-crafted and deep convolution neural

^aThese authors contributed equally to this work.

networks (CNNs) have recently been used to achieve considerably better results than the prior state-of-the-art for additional durable characteristics than shadow detection [1]. The image with prior and crafted features does not make high-level semanticity successful. Initial shadow detection approaches are mainly models focused predominantly on invariant hypotheses regarding color chromatics or illumination and use artisanship features such as illumination points [2] and others. The assumption model only works well with high-quality as well as well-restrained images. For that reason, shadow regions are identified in consumer photos when consumer photos perform so poorly on complex photos, for that reason, shadow regions are identified in consumer photos. Later, data-driven strategies design certain manual features on annotated data and transmit them into various classifications [3]. Although these strategies achieve improvements in accuracy, they are usually degraded in complex cases in which handcrafted features do not discriminate sufficiently for shadow regions to be found. Various hand-crafted features were used to generate user image suggestions. The initial reviews focused on edge and pixel specifics. For example, Guo et al. [4] measured illuminating features for segmented areas and then assembled a graphic classifier using details from each area and relationships in pairs. And classify them based on indications of texture, gradient, and intensity. Instead of looking at individual pixel-level signals, the researchers looked at region-level signals. Vicente et al. [5] Highly qualified shadow and shadow field classificatory and Markov Random Field (MRF) were used to improve. performance in a parallel context. Many of the methods mentioned above are hand-crafted, making them ineffective in complex scenes. However, picture priors and hand-built features for the removal of high-level semantics are not successful.

More recently, approaches based on deep convolutional neural networks have shown promising results in a variety of visual regions, yielding an accurate map with a high computational cost. For shadow images, CNN can pick up on global spatial settings [6]. Learn more about spatial context and how to increase the efficiency of shadow detection, the analysis and integration of different contexts on multiple scales into the global context on objects and conditions of illumination in the image, and local contexts on data in the shadow type. This leads us to explore shadowy contexts across multiple layers of CNNs, which require shallow sections to reveal local contexts and deep layers. Due to the clear potential for the generalization of deep CNN models, not only image-level grading tasks, but pixel-level grading tasks have

been developed and used [7]. In terms of function, the encoder-decoder models are based on fully convolution networks (FCNs) like semantic images, and edge detection SOD (Salient Object Detection) have greatly improved their performance on pixels [8]. The trend towards mainstream SOD methods, particularly in the past few years, indicates that most work is done inside the decoder scheme. The encoder is a multilevel, deeply-trained image classification model (i.e., ResNeXt-101) [9] Semantic information is used in low-resolution high-level characteristics, while low-level encoding represents spatial precise awareness. Using the combined detection of shadows, we evaluated shadow edge data and built a multi-task CNN for shadow detection, shadow edges of a single image, and other unlabeled datasets. These functions are built into the decoder to create an accurate performance chart. Various decoders [10] have been developed by researchers to combine low and high-level characteristics.

In a deep aggregation approach connected with high-level functioning, low-level attributes play a lesser role in success. When aggregating characteristics from high to low levels, performance appears to be readily saturated. Low-level resolution features that integrate these features with high-level device output improvements. When CNN becomes deep, the functionalities turn from low to high-level. Deep mapping models can only recover details on the spatial map by combining the deeper layer's features [11]. The effect of this mechanism, however, is dependent on map accuracy. Because deeper layer fusion characteristics create a fairly right shadow chart, this map may be used for the direct refinement of characteristics.

The methods above all rely on integrating multi-scale pre-trained neural network functionality, although we find that the advanced characteristics are more important for the effectiveness of shadow detection. To improve the performance and efficiency of shadow identification, we propose a modern, lightweight end-to-end network for shadows that only leverages profound and specialized features. Then, additional information is collected from adjacent layer functions to enhance the efficiency of shadow detection. More specifically, the following are the primary contributions to this paper:

- First, a new, lightweight end-to-end network for shadow detection has been developed, which accepts shadow RGB images as input and produces shadow RGB images as output. maps that detect shadows. It includes a cascaded partial decoder that only uses deep advanced features to enhance time efficiency. It obtains complementary information from adjacent layer features to boost shadow detection performance.

- The results of the experiments on three public data sets suggest that the technique used in this paper is more reliable and efficient.

RELATED WORK

Researchers have developed comprehensive shadow detection algorithms over the last two decades. Early studies investigated a physical model using the variant of illumination as well as invariant assumptions, texture, and strange derivative characteristics to differentiate shadows from a single monochromatic image. For more specific shadow measurements in monochromatic photos by using shadow variance and invariant indicators in a learning-based way that is data-driven [12]. Other models of color information assumptions will only operate well on high quality and accuracy when performing poorly on complex user images. In this case, to meet the demands of multitask learning, we will create a new large-scale dataset that includes picture shadow triplets such as shadow, shadow mask, and shadow-free images. ST-CGAN outperforms various state-of-the-art techniques in terms of both detection and extraction. The scGAN [6] generator must be trained to incorporate the shadow mask into the input scene image to generate an output image that is conditioned on an input image. For global structure and context, the generator of scGAN features a full view of the whole image and doesn't line an area region classifier. To learn more about the spatial environment and to increase the efficiency of shadow detection [13]. Guo et al. measured illuminating features for segmented sections, then build a graphic classifier using data from each region and relationships between pairs of regions that are likely to be made of the same material, and decide whether or not they have identical lighting circumstances. The existence of an encoding area and the limited range of edge capabilities indicates whether the same or different luminescence in the two regions Vicente et al. [14] region categorizer will be identical to each region's shadow probability, which is supported by contextual indications between the neighboring regions.

The relational indications are inserted into the context as feature classifications of the shadow and shadow region using the MRF model to improve performance by using the parallel meaning of the region. Here we speak mostly about deep learning frameworks for detecting shadows: Inspirational CNN approaches were created for shadow detection to obtain deep shadow inference functions from tagged information sets, and they were motivated by the remarkable development of deep learning in many computer vision problems. Deep learning's

effectiveness in computer vision problems has recently increased the popularity of shadow detection approaches [1]. Researchers first considered CNN's mainly as a strong feature extractor and drastically improved its performance with strong, deep characteristics. Khan et al. [15] developed a method to classify picture pixels as shadows or non-shadows through the creation of a 7-layer CNN that extracts deep features from pixels in the image and feeds them into the conditional random field (CRF) model to smooth shadow detection effects. Authors in [16] Shen et al. Shadow edges were retrieved using structured CNNs, and so shadow restoration was handled as an optimization debate. The occurrence of complete convolution networks (FCN) later proposed end-to-end CNN models [17]. As an example, Vicente et al. [18] had previously studied a shadow-level image and used a shadow mask-based patch to coach CNNs. Then, using a conditional generator for the input picture, a shadow detector called scGAN forecasts a shadow mapping. A quick, deep shadow detection grid uses the preceding map's shadows to anticipate the shadow patch masks, then includes the results of non-linear and non-patched dispatches for the overall shadow map prediction. Nowadays, researchers have shown that combining multi-level features increases performance on dense time series forecasting [19]. High-level aspects of CNN include contextual and low-level knowledge. Spatial information is useful in improving object boundaries. This method is used in the course of several works, specifically fragments of the relevant subjects. Integrate multi-level function maps in numerous resolutions at identical times with semantic information and spatial data [20]. Then predict and fuse the shadow map at every resolution to make the ultimate shadow map. [21] retrieves multi-level context-aware features and uses a two-fold gated framework to transport messages between them.

Introduced their shadow-aware distraction module (DS) in each shadow sheet and fuse distraction functions. As the shadow bar grows, the creation of their detection networks requires a tremendous amount of information, and annotations at the pixel level almost have a downside to existing methodologies. Deep models of state-of-the-art shadow detection have primarily emphasized the degree to which global contexts are extracted. But here we introduce the utility of shadow detection's deeply advanced functionality. We propose a replacement, lightweight, end-to-end shade network that uses only proven and advanced functionality to boost shadow detection performance and effectiveness. Then additional information will come from the adjacent functions for shadow detection.

PROPOSED METHOD

We suggest in this paper, an additional exploration of deep aggregation for a unique partial decoder framework.

Overview of the proposed framework's architecture

Our proposed method consists of two parts: the Adjacent Layers Shadow Feature Extraction Module (ALSFEM) and the Shadow Feature Redress Module (SFRM). Figure 1(a) shows the overall architecture: The feature extraction model discards the first three low-level features and then divides them into the attention branch and detection branch, which contain the fourth and fifth layers with the same structure, and are connected to the partial decoder module at the end. Figure 1(b) The current process of the shadow detection system leverages adjacent layer shadow characteristics to fully utilize both the two levels of the global and local contexts. Our network accepts one picture as an input and produces shadow detection as an output. First, it employs CNN to extract features of varying resolutions. The input picture is supplied into the backbone at various sizes to encourage backbone capabilities. The layer of final convolution of output for each field is used for backbone features in the network. Each field takes an image, which is then passed to the E3, E4, and E5 features by the Partial Decoder Module (PDM) [16], which only integrates features from a deeper layer. After initializing the shadow map, we'll go through the proposed holistic attention map(HAM) and then an improved HAM to refine the feature E3. The attention map efficiently eliminates the distractor in the feature and multiplies each feature to come up with an attention map. The shallow layer feature module identifies the deep aggregation layer feature module and accumulates shadow contextual information for the entire image, rather than just one shadow-specific detail inside the local area. Then, utilizing two neighboring feature modules as input for a concentrating module and an SFEM for context feature modification, we construct SFEM to gradually improve the features of each CNN layer. We incorporate SFEM with SFRM, which is then up-sampled using bilinear interpolation, thick relations with dense connectivity were concatenated from top to bottom and routed to a convolutional 1×1 layer to fusion. Finally, we forecast the scoring module based on the features and combine those two scores in a sustained attention layer as well as a sigmoid activation function to obtain the final shadow detection result using the soft binary shadow map.

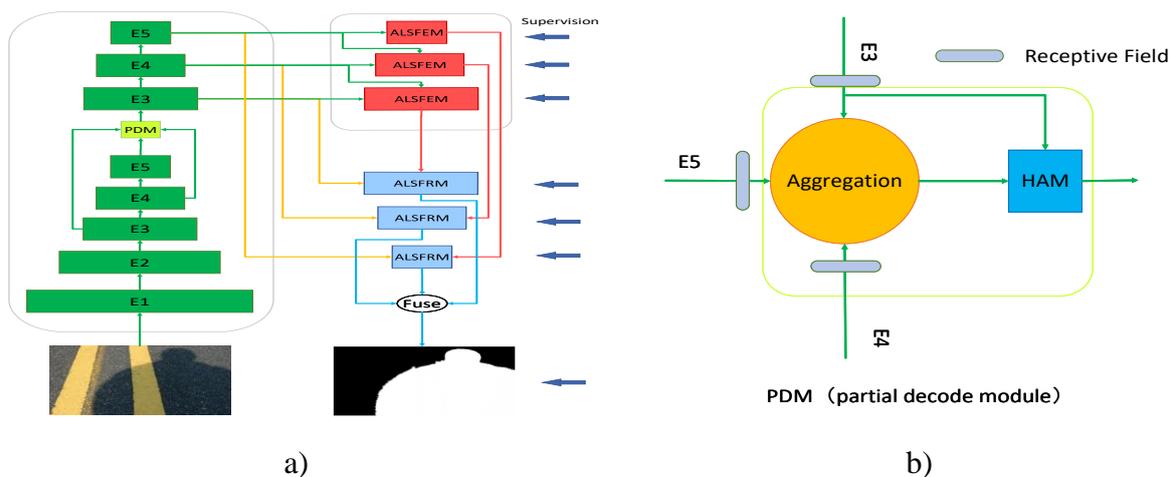


Figure 1. (a) Illustrates the network model architecture proposed in this paper, RGB image as the new input, and output shadow detection results. (b) is the structure of the PDM module, which aggregates deep features.

Partial Decoder Module (PDM)

Figure 1(b) shows that the Partial decoder module uses an improved Receptive Field Block (RFB) module, adds a branch to extend the receptive field, and uses 1×1 convolution to cut back on the quantity of calculation. Create an efficient context unit motivated by RFB. We have added three expanded receptive fields to the first RFB, and our context module has three branches $\{b_m, m = 1, 2, 3\}$. For acceleration, we utilize a 1×1 convolutional layer to reduce the channel number to 32 in each branch. For $\{b_m, m > 1\}$ we add two layers: $(2m - 1) \times (2m - 1)$ Convolutional layer $(2m - 1)$ and 3×3 layers of conversion $(2m - 1)$ dilation. We configure the outcomes of such divisions and use an extra 1×1 fully connected layer to minimize the channel to 32. The original RFB was then given a brief connection. In fact, given the split backbone network characteristics $\{f_i^c, i \in [1, \dots, L], c \in [a, d]\}$ we derive discriminatory features from the context module. Then, to close the gap between multi-level features, employ multiplication operations.

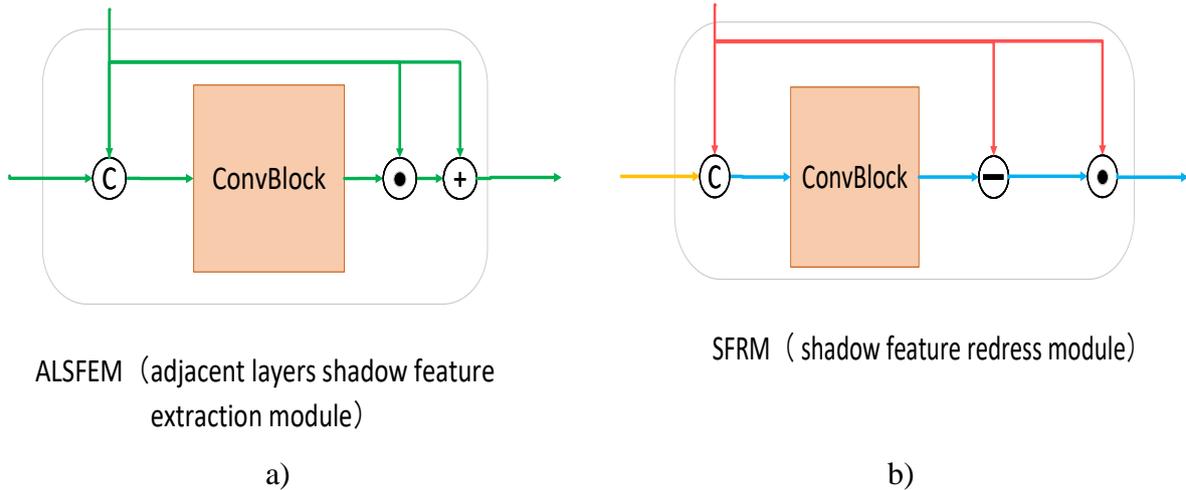


Figure 2. The structure diagram of ALSFEM and SFRM.

We specifically set $(f_L^{c2} = f_L^{c1})$ for the best feature ($i = L$). For features $\{f_i^{c1}, i < L\}$ with all deeper layer characteristics $\{f_i^{c2}\}$:

$$sf_i^{c2} = f_i^{c1} \odot \prod_{k=i+1}^L \text{Conv} \left(\text{Up}(f_k^{c1}) \right), i \in [l, \dots, L1] \quad (1)$$

where $\text{Up}(\cdot)$ is a factor 2^{k-j} upsampling feature while Conv can be a 3×3 convolutional layer.

Finally, we use an up-sampling approach for integrating multi-level functions. We built a provisional converter and used the convolutional layer 3×3 due to the optimized layer ($l=3, L=5$) to get a feature map of $\left[\frac{H}{3}, \frac{W}{3}\right]$ size and 64 channels. Get the extracted feature map and scale it to $[H, W]$ using 3×3 layers and 1×1 classification algorithms. In addition, when the aggregation of those works are want to integrate the functions of every branch the proposed framework may want to improve the prevailing deep aggregation model. Whether or not the mathematical formulation of the core network is increased and a decoder is introduced, the computational complexity of the backbone network is still greatly decreased due to the discarding of low-level operations inside the decoder. Furthermore, the proposed framework's cascading optimization technique enhances performance, and studies demonstrate that these two branches outperform the original model.

Adjacent layer shadow feature extraction module

First, we will explain how the ALSFEM works in the following parts. Figure 2(a) Refines the feature module in the adjacent layer. An adjacent layer is used as an input because the global information of the deep layer is richer than that of the shallow layer. So, the deep layer is used to refine the shallow layer. For instance, in Figure 1(b), E5 has richer global information than E4. We use E5 to refine E4 and supplement E4 as global information. The partial decoder module of the attention branch is multiplied by the corresponding elements of the third layer and then input to the detection branch, which is understood by individuals as a similar attention mechanism. For the sake of simplicity, the two branches of the model use the same partial decoder module structure. The difference is that the deeper receptive field is larger, so its global information should be more abundant. Taking them as input, we can obtain the redundant information caused by the dereferencing of the receptive field and use the redundant information to correct the features of the shallow layer. Such collar layer aggregation can better refine the shallow layer features. The two branches are jointly trained, the loss function adopts cross-entropy loss, and the total loss is the sum of the loss function output by the two branches of the partial decoder module. A unified architecture of the full decoder can be represented by: We built our architecture based on ResNeXt-101. It is highly renowned for producing almost state-of-the-art results in picture classification and strong generalization characteristics, as well as being the most often used backbone network in deep shadow detection models. We can abstract characteristics at three levels from the input image of size $H \times W$, which is indicated as, $\{E_i, i = 1, \dots, 5\}$ with output features map size: $\left[\frac{H}{2^{i-1}}, \frac{W}{2^{i-1}}\right]$. The decoder suggested above work as named complete decoders, incorporating all the functionality of the shadow map in Collar layer. Since the features of the shallow layers make a less effective contribution $D_T = g(E_1, E_2, E_3, E_4, E_5)$, where $g(\cdot)$ refers to an aggregation of a multi-level feature. We create a partial decoder that incorporates only the deeper layers. We designed a split communication infrastructure to increase features by utilizing the produced shadow map. As an optimization layer, we set Conv 3×3 and created two branches on the last two convolutional blocks.

For the additional branches, we created a Partial Digital Converter Module to incorporate three-level characteristics specified at $E_i^a = E_i, i = 3, 4, 5$. As a result, the partial decoder is denoted by $D_a = g_a(E_3^a, E_4^a, E_5^a)$ and creates a preliminary shadow map S_i .

Following completion of the planned Holistic Focus Module (HAM). Since the characteristics of the top three layers incorporate a reasonably reliable shadow diagram. Attention maps to features S_h of distraction maps, which E_3 essentially removes. Then we have a E_3^d refined detection function for the branch by multiplying the elemental characteristics and the concerned map: $E_3^d = E_3 \times S_h$. Therefore, the two-tier detection branch characteristics are defined as E_4^d, E_5^d . By constructing another partial decoder $D_d = g_d(E_3^d, E_4^d, E_5^d)$ The suggested model generates the ultimate shadow detection chart, S_d for the detection division. To make it clear, we set $g_a = g_d$. We train the two divisions along with the simple facts. There are no general qualifications for both divisions. Due to S_i, S_d and the associated mark i , complete total L_{total} . The formulation is as follows:

$$L_{total} = L_{ce}(S_i, l | \theta_i) + L_{ce}(S_d, l | \theta_i) \quad (2)$$

Shadow feature redress module (SFRM)

SFRM uses the output of SFEM Figure.2(b) and CNN's take features (E_3, E_4, E_5) as input. Because the global semantic information of features refined by SFEM is richer than E_3, E_4 , and E_5 they're aggregated to get redundant information (area for detecting errors). To correct redundant information, use E_3, E_4 , and E_5 . SFRM takes the adjacent features as input and outputs the corrected shadow features. First, it adds the collar layer feature and fed into a Conv-block and so gets the mask through the attention module. The mask image presentation is obtained by subtraction as an input layer to urge the context information and add this redundant information from the input layer to get the corrected shadow feature. SFRM is useful in discriminating between SFEM output and real shadows. This can be thanks to making the network less prone to getting accurate and fast shadows from any shadow images.

ANALYSIS AND EXPERIMENTAL RESULTS

In this section, we first present data sets and assessment parameters for shadow detection before comparing the proposed technique to existing shadow detectors and related works such as shadow deletion, detection, and classification techniques.

The dataset contains metrics for evaluation

Training conditions

We generate constants of a basic deep neural network by ResNeXt-101, It's quite good for ImageNet's photo categorization task, to speed up the training process and reduce the risk of overfitting, as some other function is given to discrete values. The entire network is optimized for 10000 iterations and an 8-batch size using the Stochastic gradient descent (SGD) optimizer with a momentum of 0.9 and weight decay of 5×10^{-4} . We start with a learning rate of 5×10^{-3} and lower it using a 0.9-strength polynomial approach. To train our network on a separate GTX 1080Ti, we scaled all labeled and unlabeled pictures to 320×320 and enrich the base classifier with random diagonal flips.

Datasets

To assess our technique, we utilize three available datasets: ISTD [22], SBU [23], and UCF [11]. The ISTD dataset includes 70 triples of shadow pictures, shadow-free pictures, and shadow maps of which 540 are tested. We have their shadow maps and photographs to estimate shadow detection for all of the evaluation metrics including pixel-by-pixel comments. SBU dataset including 4089 training and 638 assessing photos. The UCF dataset comprises 245 pictures, 110 of which were evaluated. ISTD only has pictures with shadows, but SBU and UCF have both self-shadowed and pictures with shadows in different situations. To demonstrate our model's generalization potential, we trained on the SBU and UCF datasets and tested both. We reorient the algorithm on the training sample for ISTD and then analyze the implementation of this testing data.

Metrics for evaluation

To quantify the performance of shadow detection, we use a balanced error rate (BER):

$$BER = 100 * \left(1 - \frac{1}{2} * \left(\frac{N_{tp}}{N_p} + \frac{N_{tn}}{N_n} \right) \right) \quad (3)$$

where N_p , N_n , N_{tp} , and N_{tn} , represent the recognition accuracy, and shadow, non-shadow, and true negative pixels in the shadow picture, separately. A better shadow accuracy rate is indicated by a lower BER rating.

Method of evaluation for shadow detection

We evaluate our technique with ideal shadow detection methods such as DSDNet [25], DC-DSPF, BDRAR [24], AD-Net [20], DSC [25], ST-CGAN, patched CNN [26], scGAN, and stack-CNN, that use deep learning, and Unary-Pairwise that use hand-crafted features. For a reasonable assessment, we used the author's published outcomes from the publication.

Visual contrast

As shown in the Figure, we use the shadow detection map created by our approach and the most recent method for offering to individuals (Figure).

The figure illustrates this. (3) that our approach can be successful, we give some visual data to qualitatively compare our methods to available methods, as shown in Figure 3. We can see from the top three rows that our technique is more capable of discriminating between actual shadows and non-shadows in terms of shadow appearance than alternative ways. In the last row, for example, our method can accurately detect the shadow areas, while other dominant methods (i.e. DSDNet, for example) appear to mist rack the shadows from barrel-shape cylinders inside the top of the corner. Dark work-like shadows were detected inappropriately in the second row, current ways (e.g. BDRAR). The last two rows display several challenging cases of shadow-detection, anywhere real shadows test like a background visually (potential false negatives). We'll be able to see that our strategy will clear the cases with continued progress, while alternate approaches neglect actual shadow areas. For instance, in the fourth row, all current ways except our methods, the black part at the bottom of the stairs sections as non-shadow (false negative) under the shadow regions, whereas DSDNet, BDRAR, and DSC will struggle to see the non-shadow region between two legs. Our method predicts them accurately in terms of distinction. Finally, as seen in Figure. (3), we would like to discuss the predictions of FP and FN made in our PD module. These findings will shed light on the integration of distraction linguistics, into certain challenging situations and will enable the identification of shadows. For example, our FP predictor estimates that the black part at the bottom of the stairs is a false positive in the prime row, which helps to ensure that the shadow area is correctly discriminated against by our model. In the second section, our FN predictor

works with a high visual resemblance to the adjacent building in the shadow zone. It will make our model simpler to fix possible ambiguities in the shadow area at intervals.

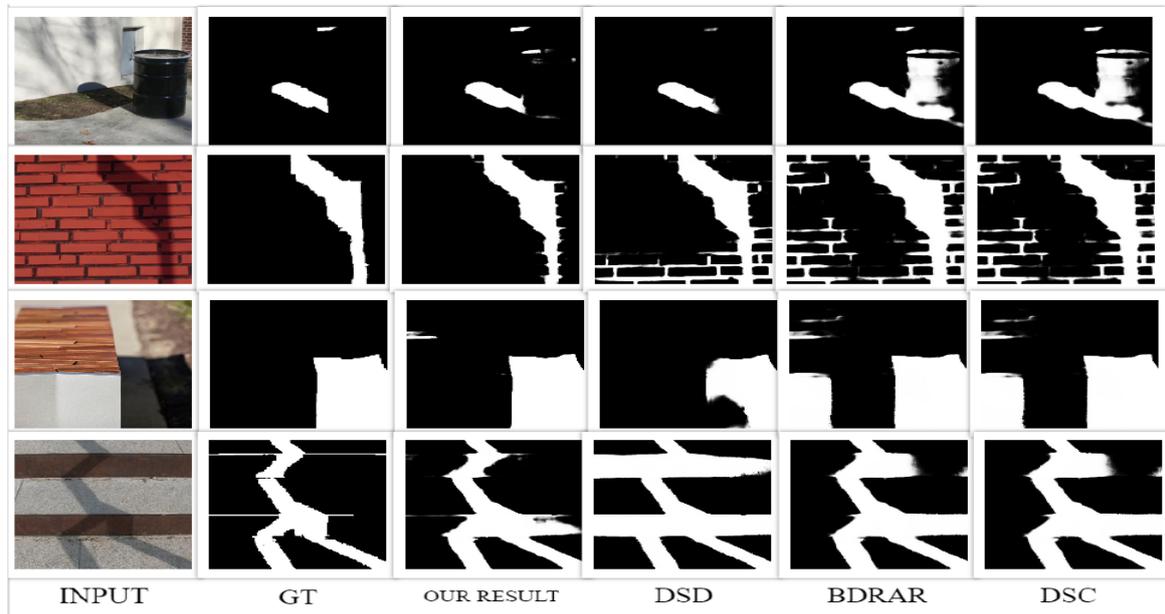


Figure 3. Compares our methods to available methods.

Table 1. Our strategy is realistic, especially in comparison to other methods.

Dataset	ISTD			UCF			SBU		
Method	BER	Shadow	Non-Shadow	BER	Shadow	Non-Shadow	BER	Shadow	Non-Shadow
PSPNet	4.26	4.51	4.02	11.75	0.00	0.00	8.57	0.00	0.00
AADEF-Net	0.00	0.00	0.00	0.00	0.00	0.00	3.77	4.24	3.33
GateNet	0.00	0.00	0.00	0.00	0.00	0.00	3.73	3.37	4.10
EGNET	1.85	1.75	1.95	9.20	11.28	7.12	4.49	5.23	2.50
SRM	7.92	13.97	1.86	12.51	21.41	3.60	6.51	10.52	2.50
Amulet	0.00	0.00	0.00	15.17	0.00	0.00	15.13	0.00	0.00
Deshado	0.00	0.00	0.00	8.92	0.00	0.00	6.92	0.00	0.00
wNet									
Patched-CNN	0.00	0.00	0.00	0.00	0.00	0.00	11.56	15.60	7.52
sgGAN	4.70	3.22	6.18	11.50	7.74	15.30	9.10	8.39	9.69
Stacked-CNN	8.60	7.69	9.23	13.0	9.00	17.10	11.0	8.84	12.76
UnaryPa	0.00	0.00	0.00	0.00	0.00	0.00	25.03	36.26	13.80
irwise									
ST-CGAN	3.85	2.14	5.55	11.23	4.94	17.52	8.14	3.75	12.53
DSC	3.42	3.85	3.00	10.54	18.08	3.00	5.59	9.76	1.42
AD-Net	0.00	0.00	0.00	9.25	8.37	10.14	5.37	4.45	6.30
BDRAR	2.59	0.50	4.87	7.81	9.69	5.94	3.64	3.40	3.89
DC-DSPF	0.00	0.00	0.00	7.90	6.50	9.30	4.90	4.70	5.10
DSDNet	2.17	1.36	2.98	7.59	9.74	5.00	3.45	3.33	3.58
Ours	1.96	1.40	2.52	7.52	9.80	5.24	3.26	3.14	3.38

Qualitative contrast

In three data sets, Table 1, our strategy is realistic, especially in comparison to others. The lower the shadow and non-shadow error rates, the more comprehensive and redundant the

shadow detection map is, and the BER is their approximate amount, which determines the performance of the shadow detection map in a neutral way.

The red annotation represents the highest result, and the blue represents the second-highest result. In most circumstances, our technique has the lowest BER score. On ISTD datasets, where identity is more suited, EGNNet works well. Furthermore, in the UCF and SBU data sets, our technique has the lowest BER score, indicating that our method performs better in dealing with varied settings and shadows, not only self- shadows.

Furthermore, the solution based on custom feature identification has a higher BER score than deep learning-based techniques, demonstrating that supervised learning photos may be used to build stronger shadow detection features using deep learning approaches. These models may be re-trained, trained on shadow detection samples, and utilized to identify shadows.

To further compare, we re-train and evaluate a contemporary shadow classification model, a semantic segmentation model, and a shadow removal model on shadow detection datasets using the researchers' codes and change the variables for optimum results.

As demonstrated in Table 1, our technique beats current models in some instances, despite their greater spectral efficiency over some current shadow detectors.

CONCLUSION

This research describes a unique link for detecting single-image shadows. To thoroughly investigate the domestic and global relevant information decoded in various layers of a convolutional neural network (CNN), three strategies are tried to be introduced: partial decoder module (PDM), neighboring layer shadow function extractor unit ALSFEM, and shadow feature repair module SFRM (CNN). By studying the attentiveness strengths to pick an end-to-end approach, the PDM module delivers a unique feature improvement technique for the scenario in the additional layer. The shadow context is aggregated in one direction at multiple levels, thus improving the shadow limits and also removing the non-shadow areas. Experiment results demonstrate that our model can tackle tough and unclear scenarios in the field of shadow detection positively, offering new state-of-the-art SBU, UCF, and ISTD dataset capabilities. But even if our techniques are designed to handle the most complicated concepts, poor shadow images with such a dark background don't fail to indicate, as shown in Figure (3) in the last row.

Finally, we test our network on three datasets and compare it to a variety of state-of-the-art approaches, as well as display our network's accuracy results and BER statistics.

Funding: The author(s) received no financial support for the research, authorship, and/or publication of this article.

Conflicts of interest/Competing interests: currently available research influences us to study shadow detection. We think we can make something better for detecting shadows in still, continuous, and live images. For this research, we collected some existing data sets from the internet and we collected images for the application of our method from the internet also. After analyzing our method we detected shadows accurately. We will continue our research in the future for better shadow-removing systems in computer vision.

Availability of data and material: Internet and some physical sources.

Code availability (software application): We cannot share our custom code. But we wrote our code in Python and simulated it with PyCharm.

Authors' contributions: Successfully detected shadows by this research and compared them with other methods. We proved that our method can identify the best results rather than others.

REFERENCES

- [1] Guanbin Li, Yu Y. Visual saliency based on multiscale deep features. IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 2015; 5455-5463. doi: 10.1109/CVPR.2015.7299184
- [2] Finlayson Graham D, Drew Mark S, Lu Cheng. Entropy minimization for shadow removal. International Journal of Computer Vision. 2009; 85 (1): 35-57.
- [3] Lalonde J.F., Efros A.A., Narasimhan S.G. Detecting Ground Shadows in Outdoor Consumer Photographs. In: Daniilidis, K., Maragos, P., Paragios, N. (eds) Computer Vision – ECCV 2010. ECCV 2010. Lecture Notes in Computer Science. Springer, Berlin, Heidelberg. 2010; 6312. https://doi.org/10.1007/978-3-642-15552-9_24
- [4] Ruiqi Guo, Qieyun Dai, Derek Hoiem. Paired regions for shadow detection and removal. IEEE transactions on pattern analysis and machine intelligence/ 2012; 35(12): 2956-2967.
- [5] Vicente T.F.Y., Hoai M., Samaras D. Leave-One-Out Kernel Optimization for Shadow Detection and Removal. In IEEE Transactions on Pattern Analysis and Machine Intelligence.

2018; 40 (3): 682-695. doi: 10.1109/TPAMI.2017.2691703

[6] Ding B., Long C., Zhang L., Xiao C. ARGAN: Attentive Recurrent Generative Adversarial Network for Shadow Detection and Removal. IEEE/CVF International Conference on Computer Vision (ICCV). 2019; 10212-10221. doi: 10.1109/ICCV.2019.01031

[7] Hearst Marti A., Dumais Susan T, Osuna Edgar, Platt John, Scholkopf Bernhard. Support vector machines. IEEE Intelligent Systems and their applications. 1998; 13 (4): 18-28.

[8] Liu C., Jia K., Liu P. Fast Intra Coding Algorithm for Depth Map with End-to-End Edge Detection Network. IEEE International Conference on Visual Communications and Image Processing (VCIP). 2020; 379-382. doi: 10.1109/VCIP49819.2020.9301859

[9] Wang Yupei, Zhao Xin, Li Yin, Hu Xuecai, Huang Kaiqi. Densely cascaded shadow detection network via deeply supervised parallel fusion. In Proceedings of the 27th International Joint Conference on Artificial Intelligence (IJCAI'18). AAAI Press. 2018; 1007-1013.

[10] Chen S., Fu Y. Progressively Guided Alternate Refinement Network for RGB-D Salient Object Detection. In: Vedaldi, A., Bischof, H., Brox, T., Frahm, JM. (eds) Computer Vision – ECCV 2020. ECCV 2020. Lecture Notes in Computer Science. 2020; 12353. Springer, Cham. https://doi.org/10.1007/978-3-030-58598-3_31

[11] Simonyan Karen, Zisserman Andrew. Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv. 2014; 1409.1556.

[12] Jiandong Tian, Xiaojun Qi, Liangqiong Qu, Yandong Tang. New spectrum ratio properties and features for shadow detection. Pattern Recognition. 2016; 51: 85-96.

[13] Liangqiong Qu, Jiandong Tian, Shengfeng He, Yandong Tang, Rynson W. H. Lau. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 2017; 4067-4075.

[14] Caijuan Shi, Weiming Zhang, Changyu Duan, Houru Chen. A pooling-based feature pyramid network for salient object detection. Image and Vision Computing. 2021; 107: 104099.

[15] Salman Hameed Khan, Mohammed Bennamoun, Ferdous Sohel, Roberto Togneri. Automatic Feature Learning for Robust Shadow Detection. In Proceedings of the 2014 IEEE Conference on Computer Vision and Pattern Recognition (CVPR '14). IEEE Computer Society, USA. 2014; 1939-1946. <https://doi.org/10.1109/CVPR.2014.249>

[16] Li Shen, Teck Wee Chua, Karianto Leman. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 2015; 2067-2074.

- [17] Weiwei Sun, Ruisheng Wang. Fully convolutional networks for semantic segmentation of very high resolution remotely sensed images combined with dsm. *IEEE Geoscience and Remote Sensing Letters*. 2018; 15(3): 474-478.
- [18] Hieu Le, Tomas F. Yago Vicente, Vu Nguyen, Minh Hoai, Dimitris Samaras. *Proceedings of the European Conference on Computer Vision (ECCV)*. 2018; 662-678.
- [19] Chen L.C., Papandreou G., Kokkinos I., Murphy K., Yuille A.L. DeepLab: Semantic Image Segmentation with Deep Convolutional Nets, Atrous Convolution, and Fully Connected CRFs. In *IEEE Transactions on Pattern Analysis and Machine Intelligence*. 2018; 40 (4): 834-848. doi: 10.1109/TPAMI.2017.2699184
- [20] Zhang L., Dai J., Lu H., He Y., Wang G. A Bi-Directional Message Passing Model for Salient Object Detection. *IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2018; 1741-1750. doi: 10.1109/CVPR.2018.00187
- [21] Zhang P., Wang D., Lu H., Wang H., Ruan X. Amulet: Aggregating Multi-level Convolutional Features for Salient Object Detection. *IEEE International Conference on Computer Vision (ICCV)*. 2017; 202-211. doi: 10.1109/ICCV.2017.31
- [22] Wu Zhe, Su Li. Qingming Huang. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. 2019; 3907-3916.
- [23] Wang J., Li X., Yang J. Stacked Conditional Generative Adversarial Networks for Jointly Learning Shadow Detection and Shadow Removal. *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2018; 1788-1797. doi: 10.1109/CVPR.2018.00192
- [24] Mingliang Xu, Jiejie Zhu, Pei Lv, Bing Zhou, Marshall F Tappen, Rongrong Ji. Learning-based shadow recognition and removal from monochromatic natural images. *IEEE Transactions on Image Processing*. 2017; 26(12): 5811-5824.
- [25] Zheng Q., Qiao X., Cao Y., Lau R. W. H. Distraction-Aware Shadow Detection. *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. 2019; 5162-5171. doi: 10.1109/CVPR.2019.00531
- [26] Lei Zhu, Zijun Deng, Xiaowei Hu, Chi-Wing Fu, Xuemiao Xu, Jing Qin, Pheng-Ann Heng. *Proceedings of the European Conference on Computer Vision (ECCV)*. 2018; 121-136. <https://link.springer.com/conference/eccv>

INFORMATION ABOUT THE AUTHORS

Islam Md Jahidul, School of Software Engineering, Northeastern University, No. 195, Chuangxin Road, Hunnan, Liaoning, 110819, Shenyang, P. R. China; School of Communication and Information Engineering, Chongqing University of Posts and Telecommunications, 2 Chongwen Rd, Chongqing, 400065, Nan An Qu, P. R. China
e-mail: jahidulislamneu@gmail.com

Omar Faruq, School of Communication and Information Engineering, Chongqing University of Posts and Telecommunications, 2 Chongwen Rd, Chongqing, 400065, Nan An Qu, P. R. China; School of Software Engineering, Northeastern University, No. 195, Chuangxin Road, Hunnan, Liaoning, 110819, Shenyang, P. R. China
e-mail: smomarfaruq@diu.edu.bd

Статья поступила в редакцию 20.09.2022; одобрена после рецензирования 23.09.2022; принята к публикации 26.09.2022.

The article was submitted 20.09.2022; approved after reviewing 23.09.2022; accepted for publication 26.09.2022.