DSDNet Neural Network for Shadow Detection from Urban Satellite Images

Oleg Naidovich
Belarusian State University
Minsk, Belarus
o.naidovich@gmail.com

Alexander Nedzved
Belarusian State University
Minsk, Belarus
nedzveda@gmail.com

Shiping Ye
Zhejiang Shuren University
Hangzhou, China
zjsruysp@163.com
ORCID 0000-0002-9771-7168

***Abstract.* Shadow detection is one of the fundamental and challenging tasks in the field of computer vision and image processing. The increase of computing power has enabled many deep learning approaches to solve this problem. In this article we consider a DSDNet neural network in order to detect shadows on the base of texture analysis of the shadow area and bright area of the urban area.**

***Keywords*: shadow detection, DSDNet, deep neural networks, segmentation, Satellite image**

I. Introduction

Shadow is an illumination phenomenon, which is caused by the occlusion of light by some object, resulting in color and intensity changes in the local surfaces. Knowing where the shadow is allows us to infer, for example, scene geometry, lighting direction, and camera parameters. However, the presence of a shadow can degrade the performance of many fundamental computer vision tasks such as semantic segmentation, object detection and visual tracking. Consequently, the shadow detection problem has been studied for many years and presents a severe problem among computer vision tasks.

A shadow appears when an object partially or completely obscures a direct light source. The structure of the shadow is strongly dependent on the features of the object such as geometry and height. That is why it is more efficient to conduct research on structured objects like buildings. Generally, shadows are classified into attached shadows and cast shadows.

- Cast shadows arise when a light source is obstructed by a part of the same or another object. Such shadows appear on knowledge with a flat roof (see Fig. 1).
- Attached shadows are the shadows that form on the objects themselves. They arise when the angle between a surface normal and a light source direction is obtuse. They usually appear on buildings with hipped roofs (see Fig. 2).

Knowing where the shadow is, allows us to infer, for example, scene geometry, lighting direction, and camera parameters. Moreover, deleting shadows from the image can be used to detect objects such as buildings, trees, roads, etc. In addition, after removing shadows the objects will be displayed more evidently, hence they can be correctly recognized.

The presence of a shadow can degrade the performance of many fundamental computer vision tasks such as semantic segmentation, object detection, and visual tracking. The main problem caused by shadows is the loss of information in the image. It can lead to distortion of various parameters derived from pixels, so areas cannot be interpreted correctly.

The shadow detection problem has been studied for many years and it is a major concern among computer vision tasks. There are a lot of methods for detecting shadows in images. The purpose of this paper is to figure out the workflow of the neural network DSDNet and compare the results of its work to the other methods.

II. Other approaches

There are two groups of methods in detection shadows on images: traditional based methods and deep learning based methods.

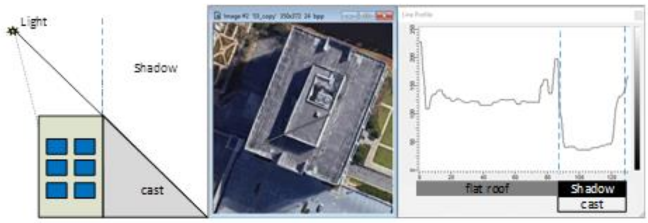

Fig. 1. Representation of the building with flat roofs

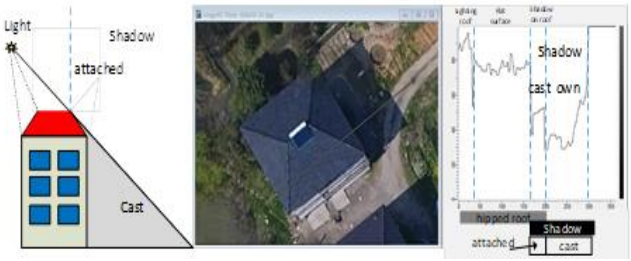

Fig. 2. Representation of the building with hipped roofs

Pattern Recognition and Information Processing (PRIP'2021) : Proceedings of the 15th International Conference, 21–24 Sept. 2021, Minsk, Belarus. – Minsk : UIIP NASB, 2021. – 246 p. – ISBN 978-985-7198-07-8.

A. Traditional methods

Traditionally, unique image shadow detection methods are based on exploitation of physical models of color and illumination. Other traditional methods use hand-crafted features which are based on marked shadow images. First of all, such methods describe image regions by feature descriptors and then categorize the regions into shadow and non-shadow regions. Such feature descriptors, for example, color, edges, and texture are frequently used in decision trees or SVM algorithms.

However, in satellite images, there are often non-shadow regions that emerge like shadows and are detected as shadows. And there are also shadow regions that emerge like non-shadow patterns and are detected as non-shadows as well. That is why these traditional predicated methods, which are based on color chromaticity and illumination, or use hand-craft features like illumination cues and color cannot deal with complex cases.

B. Deep learning methods

Due to the growth of computing power, it became possible to apply deep learning techniques to computer vision tasks. Recent state-of-the-art neural networks can be learned to detect shadows, which achieve significant performance improvements over the traditional ones.

For example, convolutional neural network is demonstrated to be a very powerful tool to learn features for detecting shadows, especially when large data is available. Also the generative adversarial networks (GANs) and recurrent neural networks (RNNs) adopted to detect shadows. They are based on capturing contextual information and exploring spatial context of the image. In general, the task of training any neural network is reduced to the task of minimizing the loss function by adjusting the parameters using the gradient descent method. Below it is presented a Distraction-aware Shadow Detection Network (DSDNet), which can be called a deep CNN.

III. DSDNet APPROACH

For shadow detection it was used convolutional neural network DSDNet which is proposed in [1] and shown in Fig. 3. In order to construct a multiscale network, it was used RexNet-101 as a backbone network. Backbone features run through each scale (i.e. conv conv1, conv2_x, conv3_x, conv4_x, conv5_x) and get into DS module. At each scale, an encoder converts the backbone features to image features. Each of DS modules will take as input an image features and produce DS features, which catch the distraction semantics. In the end, the DS features are concatenated from top to down and finally sent to a fusion layer. This size provides on the output one feature map acquisition. Finally this map is followed by a sigmoid activation function to output a binary shadow map as the final output.

A. Distraction-aware Shadow (DS) module

The DS module is used in order to learn semantic features of the distraction regions and concatenate the distraction features with the input image features to produce distraction-aware features, which are used for shadow segmentation. As the input data, it used image features (size: H*W*32), which were produced by a backbone network and Encoder layer (Fig. 3). The output data is DS features with the same size. DS module is made of a FP sub-module and a FN sub-module, and operations to combine different features.

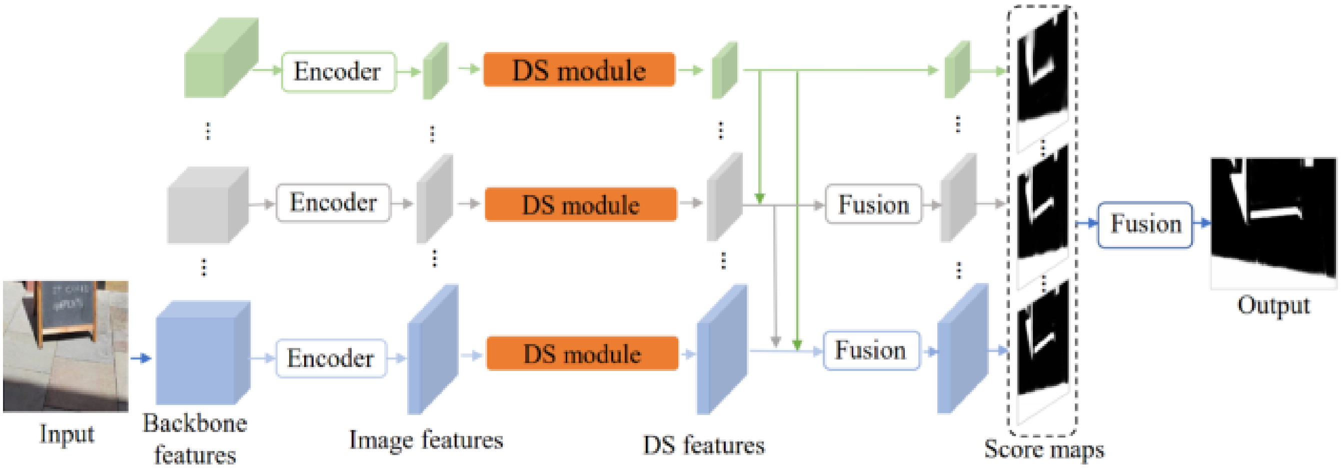

Fig. 3. Architectures of Distraction-aware Shadow Detection Network (DSDNet)

Pattern Recognition and Information Processing (PRIP'2021) : Proceedings of the 15th International Conference, 21–24 Sept. 2021, Minsk, Belarus. – Minsk : UIIP NASB, 2021. – 246 p. – ISBN 978-985-7198-07-8.

In short, image features are passed through FN sub-module in order to produce FN-masked image features, which are concatenated with Image features to produce FN-augmented features. Then, FP sub-module takes as input Image features and receives FN-augmented features and gets FP-aware image features, which are subtracted from FN-augmented features to get desired distraction-aware features.

B. False negative sub-module

FN sub-module is used in order to learn FN features and FN-masked features, which are used to augment the input image features. This module is designed to enable the network to focus on possible FN regions, which would help the network better discriminate the FN regions, whose visual appearances are highly variable from general shadows. Firstly, it employs a feature extractor on the image features to extract the FN features. It used the FN features for FN prediction, by estimating a soft binary map indicating the possible FN locations on the input image. It was necessary as we want to force the FN features to capture the semantics to recognize potential FN regions. Secondly, the FN features are concatenated with the image features and fed into an attention block to produce a mask. Thirdly, a masked image presentation is obtained by multiplying image features with duplicated soft mask element-wise. To enhance the feature activations on FN regions, received features are added to image features to produce FN-augmented image features.

C. False positive sub-module

The **FP sub-module** is used in order to learn the FP features which are used to enhance the FN-augmented features (Fig. 4). Similar to the previous scheme, it was used a feature extractor layer to get false positive features. Then we concatenate the received feature map from the FN sub-module and feed it into the Conv block in order to generate FP-aware image features. Finally, we subtract obtained FP features from FN features to weed out the negative effect of FP features on detection. This operation would make the network less sensitive to potential false positive distraction.

D. Convolutional Layers

The convolutional layers used in our network, except those listed below, are all followed by a Batch Normalization layer and ReLu activation function.

- *Encoder* consists of 2 convolutional layers, which consist of 32 filters with 3x3 kernels.
- Encoder consists of 2 convolutional layers, which consist of 32 filters with 3x3 kernels. *Fusion layer* consists of one convolutional layer with one filter with kernet size = 1 x 1.
- *The feature extractor* in FN Sub-module and FP Sub-module consists of 2 convolutional layers with 32 kernels of size 3 × 3.
- *The attention block* has one convolutional layer with 64 kernels of size 3×3, followed by a sigmoid activation function.
- *The Conv block* in FP Sub-module constituted of the 3 convolutional layers have 64 filters each, with kernel size = 1 × 1, 3 × 3 and 1 × 1, then it is followed by another 3 convolutional layers with 64 filters in the first layer and 32 filters in the others with the same kermel size.

The architecture of the DSDNet neural network is implemented in Python programming language using PyTorch library.

IV. Experiment

In training neural network it used three public datasets, where all images were reduced to a size of 320x320. In order to expand the training sample, all pictures were augmented by random horizontal flipping. The following sets of images were used as training data: SBU [2, 3] (4089 pictures for training, 638 for testing), UCF [4] (135 pictures for training, 110 for testing) and ISTD [5] (1870 pictures for training, 540 for the test). The neural network DSDNet for detecting shadows from satellite images allows us to reveal the semantics of the image due to the DS module. This module significantly increases the accuracy of shadow segmentation by double verification of geometric features, specifying their belonging to the shadow.

To evaluate the results BER metric was used (1):

$$BER = 1 - \frac{1}{2}\left(\frac{TP}{TP+FN} + \frac{TN}{TN+FP}\right), \qquad (1)$$

where TP, TN, FP, FN - denote the numbers of true positives, true negatives, false positive and false negative shadow pixels, respectively. BER is an effective metric to calculate the efficiency on the class imbalance results, and that is why it is extensively used for shadow evaluation. A lower score indicates a better performance.

DSDNet method was compared with many state-of-the-art shadow detection methods like DSCNet [6], scGAN [7], BDRAR [8] and ST-CGAN [9]. Table 1 presents the results of a quantitative comparison of the neural networks which were presented above. It shows that DSD network has the best scores on all test datasets. Fig. 5 shows visual results of the considered neural network workflow on satellite images.

Pattern Recognition and Information Processing (PRIP'2021) : Proceedings of the 15th International Conference, 21–24 Sept. 2021, Minsk, Belarus. – Minsk : UIIP NASB, 2021. – 246 p. – ISBN 978-985-7198-07-8.

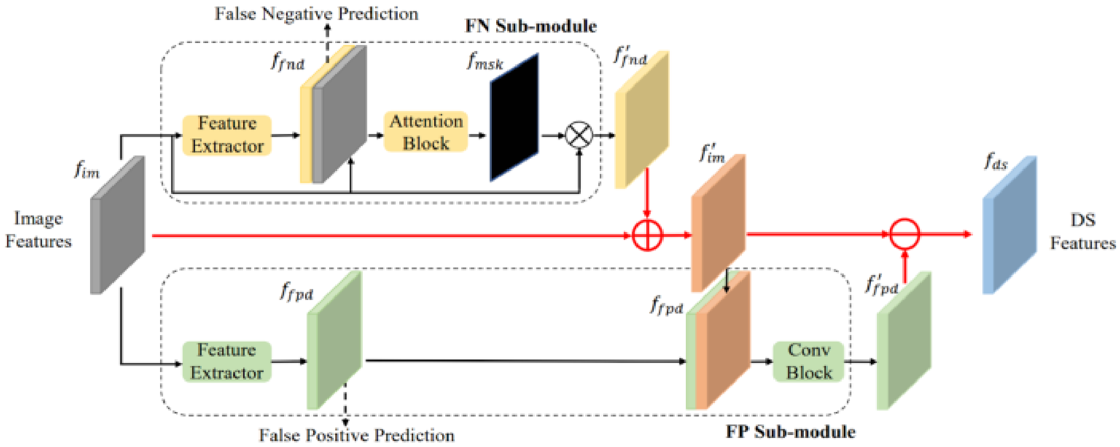

Fig.4. Architecture of the Distraction-aware Shadow module (DS module)

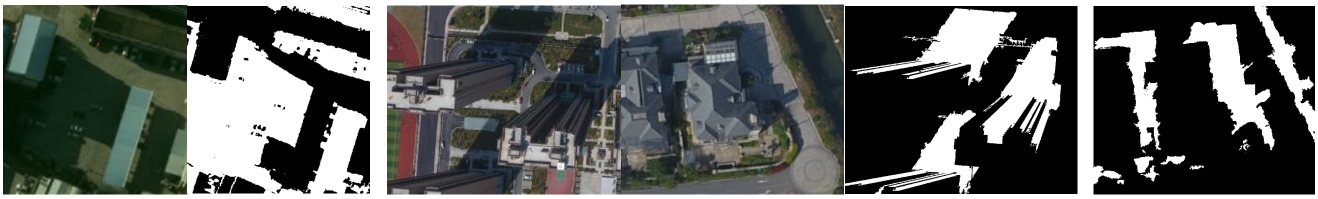

Fig.5. The result of DSDNet work. The white part is the shadow pixels. The black part is non shadow pixels

TABLE I. QUANTUTATUVE SHADOW DETECTION RESULTS

Methods	BER		
	SBU	UCF	ISTD
DSDNet	**3.45**	**7.59**	**2.17**
DSCNet	5.59	10.54	3.42
scGAN	9.04	11.52	4.70
BDRAR	3.64	7.81	2.69
ST-CGAN	8.14	11.23	3.85

V. Conclusion

The basic difference of buildings patterns is geometric features on satellite images. It allows us to describe buildings as structured objects. The DS module is an additional clarification of their belonging, thus the accuracy of determining the shadow increases. The DS module augments input image features with explicitly learned distraction features by a specific fusion strategy to produce distraction-aware features for robust shadow detection. However, there are many cases where the network gives a wrong result. It can happen either on weak shadow images where the shadows have very similar brightness to the background or images with an extremely dark background, where the shadows are almost blended into the background. Due to a small number of labeled satellite images, the neural network was trained on public datasets, which contain not only satellite images, but also other cases. If the neural network is trained only on satellite images, the accuracy will be strongly increased. The creation of new datasets will be a further development of DSDNet to improve its efficiency in order to solve the issue of shadow segmentation from satellite images.

References

[1] Q. Zheng, X. Qiao, Y. Cao and R. W. H. Lau, "Distraction-Aware Shadow Detection," 2019 IEEE/CVF Conf. on Comp. Vis. and Pattern Recogn. (CVPR), 2019, pp. 5162-5171.

[2] T. F. Y. Vicente, L. Hou, C.-P. Yu, M. Hoai, D. Samaras, “Large-scale training of shadow detectors with noisily-annotated shadow examples,” in European Conference on Computer Vision, 2016, pp. 816–832.

[3] T. F. Y. Vicente, M. Hoai, and D. Samaras, “Noisy label recovery for shadow detection in unfamiliar domains,” in IEEE Conference on Computer Vision and Pattern Recognition, 2016, pp. 3783–3792.

[4] J. Zhu, K. G. Samuel, S. Z. Masood, and M. F. Tappen, “Learning to recognize shadows in monochromatic natural images,” in IEEE Conference on Computer Vision and Pattern Recognition, 2010, pp. 223–230.

[5] J. Wang, X. Li, and J. Yang, “Stacked conditional generative adversarial networks for jointly learning shadow detection and shadow removal,” in IEEE Conf. on Computer Vision and Pattern Recognition, 2018, pp. 1788–1797.d

[6] X. Hu, C.-W. Fu, L. Zhu, J. Qin, P.-A. Heng, "Direction-Aware Spatial Context Features for Shadow Detection and Removal," in IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 42, no. 11, pp. 2795-2808.

[7] V. Nguyen, T. F. Y. Vicente, M. Zhao, M. Hoai, D. Samaras, “Shadow detection with conditional generative adversarial networks,” in IEEE International Conference on Computer Vision, 2017, pp. 4510–4518

[8] L. Zhu, Z. Deng, X. Hu, C.-W. Fu, X. Xu, J. Qin, and P.-A. Heng, “Bidirectional Feature Pyramid Network with Recurrent Attention Residual Modules for Shadow Detection,” in IEEE International Conference on Computer Vision, 2018.

[9] J. Wang, X. Li, J. Yang, "Stacked Conditional Generative Adversarial Networks for Jointly Learning Shadow Detection and Shadow Removal," 2018 IEEE/CVF Conf. on Comp. Vis. and Pattern Recogn., 2018, pp. 1788-1797.

Pattern Recognition and Information Processing (PRIP'2021) : Proceedings of the 15th International Conference, 21–24 Sept. 2021, Minsk, Belarus. – Minsk :
UIIP NASB, 2021. – 246 p. – ISBN 978-985-7198-07-8.