Machine Learning Final Project: *Shadow detection*

Sepehr Nourmohammadi, 22001136 Navid Ghamari, 22001138

1. Introduction

Shadow Detection got too much interest in the field of Computer Vision in the recent years; The reason behind this is the fact that shadows are undesired visual effects on the images and cause lower accuracies in the other computer vision applications e.g., Image Segmentation and Object Recognition, Object Tracking, and Surveillance. To define a shadow: Whenever an object blocks a light source, shadows appear. An object may cast shadow on its own surface (self-shadow), or on another surface (cast shadow).

2. Proposed Framework

One of the sophisticated detection tasks in Computer vision is Camouflage Detection regarding which numerous well-performing models were proposed. Since in this task, camouflaged objects should be detected into the wild images and not in the images captured by humans (with the purpose of researching), camouflage detection is assumed as one of the difficult detection tasks. As a result, we propose a framework which is inspired by multiple well-performing camouflage detection models [1].

This framework is divided into two main modules, searching module and identifying module. The first section has capability of searching a shadow in an image and the latter one is almost focusing on detecting shadow precisely. The former section (search module) is based on the verification that neuroscience experts set out; In a biological visual system, there are different receptive fields which pave way to robust the area that is nearby the retinal fovea, and more over this part is highly sensitive to a tiny fraction of changes of spatial shifts. Based on this preconception knowledge, receptive field provides an outstanding schism between the features which has been represented during searching evaluation. The overall view of the structure is given in the Figure 1 below:

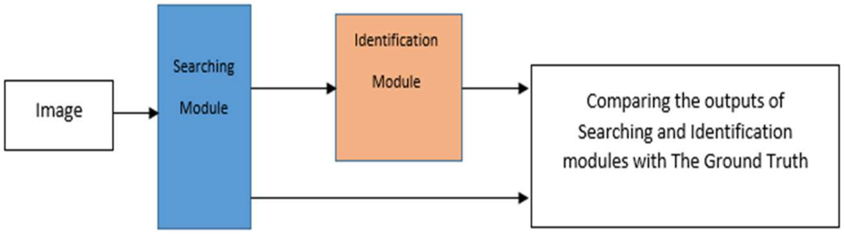
Image
Searching Module
Identification Module
Comparing the outputs of Searching and Identification modules with The Ground Truth

Fig 1. The overall structure of the shadow detection

2.1 Searching module

In the searching section, there are 5 layers of the convolutional networks to extract features (As common approach in feature extraction procedure). The shallow layers almost conserve the components or details of spatial for determining the boundaries. In other words, low level features produce edge detection requirements. On the contrary, high level features include semantic segmentation information. Because of this interdisciplinary property of neural networks, extracted features were separated into low level, middle level, and high level. First and second convolution layers are dedicated to low level, third is the middle level and fourth and fifth convolution layers are dedicated to high level. Finally, they are concatenated after up sampling and down sampling to synchronize the feature sizes.

This approach preserves more data of variety layers, and moreover the receptive field is going to be intensified by using RF. To give an example, by down sampling the features of the low layers, resolution will decrease further and the new outputs fed into the receptive field algorithm in order to produce the output features.

Three levels of components were combined, then set of enhanced features is utilized to robust the learning. Figure 2 illustrates the structure of the Searching Module.

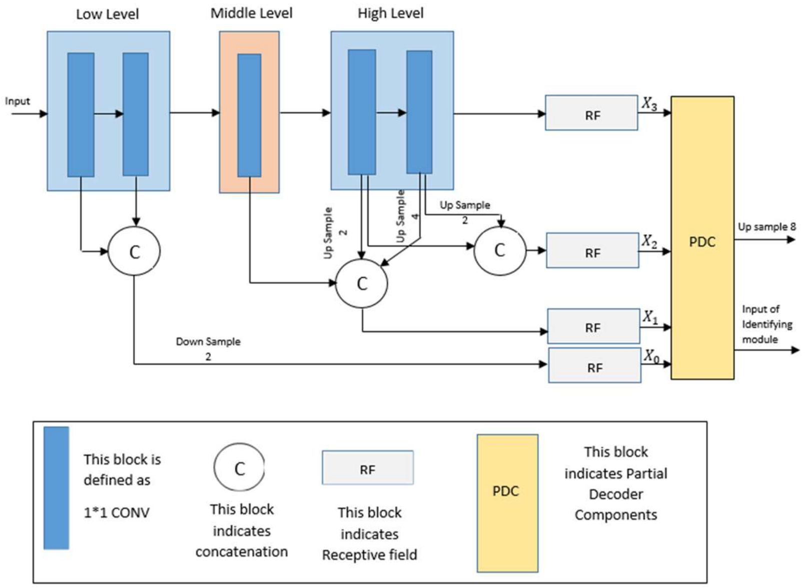

Fig 2. Structure of the Searching Module and definitions of the components of Searching Module

The outputs of partial decoder component are up sampled (to 8 times of the size of original outputs) to be in the same size of the ground truth

The structure of the RF component is given in Figure 4:

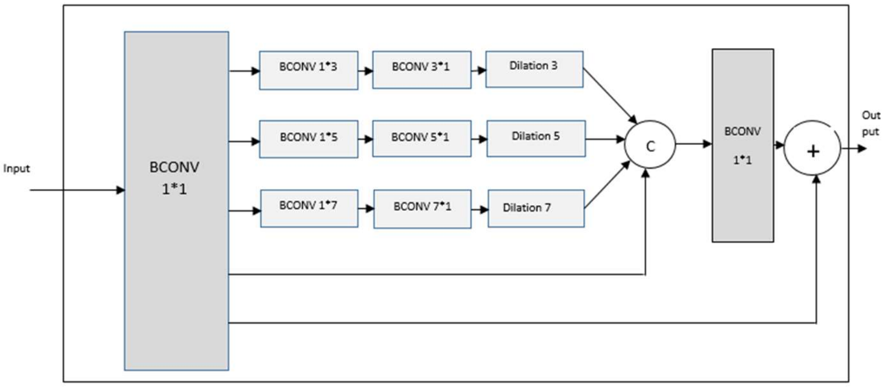

Fig 3. Inside the Receptive field (RF)

As you see, the RF component involves 5 branches, each has 1×1 convolutional layers which reduces the channel size to 32. In the first, second and third branches, $(2k - 1) \times (2k - 1)$ convolutional layers with a dilation rate $(2k - 1)$ for $k = \{2, 3, 4\}$ are used respectively. The results of first four branches are concatenated and their channel size reduces to 32 respectively. At the end, the last branch is added to the former output and fed to a RLU activation function.

2.1.1 Partial Decoder Component

The duty of Partial Decoder Component (PDC) is making pernicious differences between the shallowest layers and deeper layers (by ignoring shallow levels). It integrates high level features after feeding through the convolutional layers.

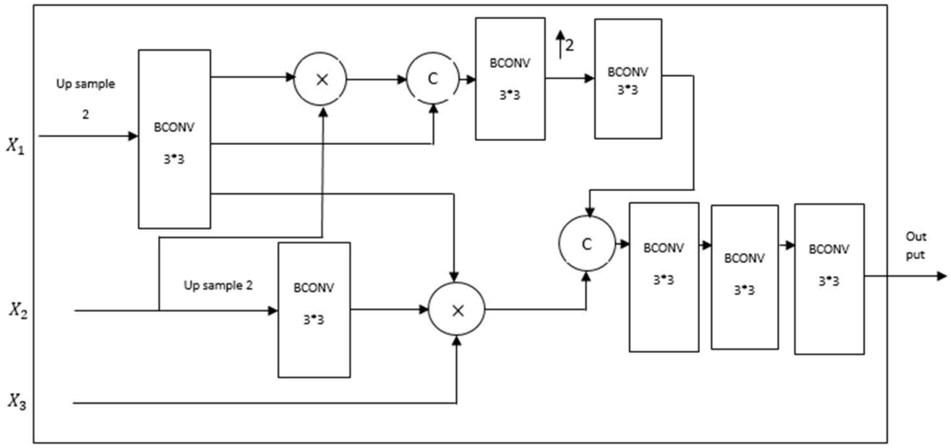

Fig 4. Structure of partial decoder components

In the figure above, X_1 is the output of middle and high level concatenation which is fed through a Receptive Field . In follows, X_2 is the of concatenation of 4th and 5th layers in the high level and then assigned to RF in order to heighten receptive field, and finally, X_3 is the output of the 5th layer which is passed through the RF structure.

2.2 Identification Module

As mentioned above, Identification has the intention to detect the shadow from the search modules output precisely. Throughout this procedure, partial decoder is used to integrate the features of the search module. The reason behind of using partial dense function is to reduce the computational complexity by this way: The high level features mostly integrated because low level features encounter high resolution. Most recent researches indicated that utilizing attenuation can significantly eliminate the irrelevant components, so, for this approach Search Attenuation (SA) module is already defined in which the middle level features are intensified to enhance the shadow map. SA is a normal Gaussian filter (σ = 32, λ = 4). The structure of Identifying Module (IM) is shown in Figure 5

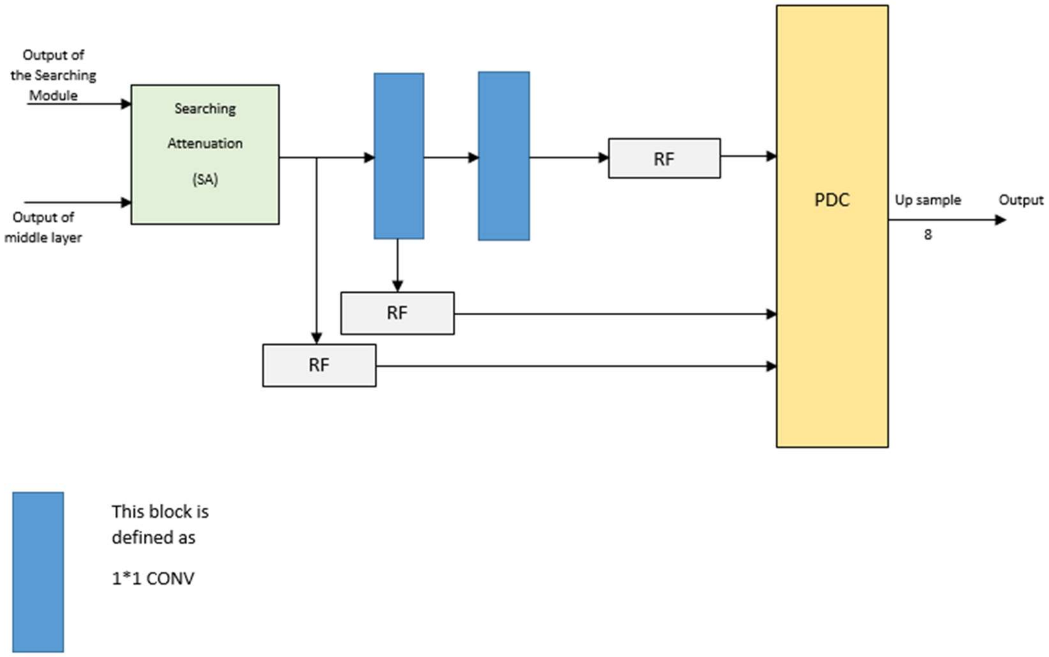

Fig 5. Structure of Identification Module

At this step, the outputs of Searching Module and middle level are assigned to searching attenuation to eradicate irrelevant features. Then, it passes through 1*1 convolutional layers. Three RF structure are applied to the output of searching attenuation, and other two convolutional blocks individually. At the end, the produced features of RF are added to a PDC structure in order to intensify the target which is basically involved in middle layer.

3. Experimental results

3.1 Dataset

Designed method is evaluated on two datasets: SBU and ISTD. SBU is already known as common complex and diversified dataset, covering different regions and shots such cities, snowy grounds, desert, etc. This dataset has 4089 images for training and 638 image for testing and all images have specific ground truth. ISTD is a dataset, combined with shadow image, shadow mask, and shadow free image. During this process we just use shadow cases and their shadow masks. In addition, the dataset has been augmented e.g., by flipping or twisting. The number of train and test samples are 1330 and 540 respectively.

3.2 Evaluation Metrics

Mean absolute error (MAE) metric provides an evaluation approach to disclose the pixel-level accuracy between the predicted and the ground truth. However, for determining the exact pixel error location, MAE is not capable to indicate where the error incidents.

E-measure metric is known as human visual perceptron, which evaluates how far pixels matched and provides an image level statistic simultaneously. Based on the intrinsic features of this metric, it is compatible for both overall and localized accuracy of the shadow detection results.

While shadow might be in casual shapes, S-measure is provided to overcome this difficulty. S-measure can judge structure similarities. Another metric we want to suggest is weighted F-measure which can produce reliable and confidential results than the normal F-measure, since it considers both recall and precision in even stages.

3.3 Results

The results were obtained by 40 iterations via python and MATLAB 2017b and are shown at the table below:

Table 1 The overall results

Datasets	E-measure	F-measure	S-measure
SBU	0.914	0.85	0.862
ISTD	0.946	0.888	0.918

To evaluate the performance of proposed method, its results have been compared with some previous research. The comparison results are summarized in Table 2 and Table 3.

ISTD dataset link: https://drive.google.com/file/d/1I0qw-65KBA6np8vIZzO6oeiOvcDBttAY/view

SBU dataset link: **https://www3.cs.stonybrook.edu/~minhhoai/projects/shadow.html**

Table 2 Comparison based on SBU dataset

Contributors	Accuracy (%)
Stacked-CNN [2]	88
cGAN [3]	87
ScGAN [3]	90
DenseASPP [4]	91.4
Guoet al. [5]	88
Freitaset al. [6]	87.2
DSAN [7]	96.2
Proposed method (our)	**96.32**

Table 3 Comparison based on ISTD dataset

Contributors	Accuracy (%)
Stacked-CNN [2]	88
cGAN [3]	91
scGAN [3]	96
DenseASPP [4]	96.1
Guoet al. p [5]	91.7
Freitaset al. [6]	91.6
DSAN [7]	97
Proposed method (our)	**97.20**

The reason behind that SBU results are low is that this dataset contains casual and non-uniform shadow samples which makes the recognition challenging.

Dawei Li et.al. [7] reached to 96.2 accuracy for SBU and 97 for ISTD in 150 iterations which is increasing the training duration, but we reached to these result with the number of iteration 40.

References of related works

[1] D.-P. Fan, G.-P. Ji, G. Sun, M.-M. Cheng, J. Shen, and L. Shao, “Camouflaged object detection,” in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., Jun. 2020, pp. 1–11.

[2] T. F. Y. Vicente, L. Hou, C.-P. Yu, M. Hoai, and D. Samaras. Large-scale training of shadow detectors with noisily annotated shadow examples. In Proceedings of the European Conference on Computer Vision, 2016

[3] Nguyen, V., Yago Vicente, T.F., Zhao, M., Hoai, M., Samaras, D.: Shadow detection with conditional generative adversarial networks. In: ICCV. (2017)

[4] M. Yang, K. Yu, C. Zhang, Z. Li, K. Yang, DenseASPP for Semantic Segmentationin Street Scenes, in: 2018 IEEE/CVF Conference on Computer Vision andPattern Recognition, Salt Lake City, UT, 2018, pp. 3684-3692, doi: 10.1109/CVPR.2018.00388

[5] R. Guo, Q. Dai, and D. Hoiem. Single-image shadow detection and removal using paired regions. In Proceedings of the 2011 IEEE Conference on Computer Vision and Pattern Recognition, CVPR ’11, pages 2033–2040, Washington, DC, USA, 2011. IEEE Computer Society

[6] Freitas, V.L.S.; Reis, B.M.F.; Tommaselli, A.M.G. Automatic shadow detection in aerial and terrestrial images. Bull. Geod. Sci. 2017, 23, 578–590.

[7] Dawei Li, Sifan Wang, Xue-song Tang, Weijian Kong, Guoliang Shi, Yang Chen, Double-stream atrous network for shadow detection, Neuro computing,417,2020,167-175, ISSN 0925-2312.