Dynamic Low-Light Image Enhancement for Object Detection via End-to-End Training

Haifeng Guo[1], Tong Lu[1,*], Yirui Wu[2]

[1] Department of Computer Science and Technology, Nanjing University

[2] College of Computer and Information, Hohai University

Abstract

Object detection based on convolutional neural networks is a hot research topic in computer vision. The illumination component in the image has a great impact on object detection, and it will cause a sharp decline in detection performance under low-light conditions. Using low-light image enhancement technique as a pre-processing mechanism can improve image quality and obtain better detection results. However, due to the complexity of low-light environments, the existing enhancement methods may have negative effects on some samples. Therefore, it is difficult to improve the overall detection performance in low-light conditions. In this paper, our goal is to use image enhancement to improve object detection performance rather than perceptual quality for humans. We propose a novel framework that combines low-light enhancement and object detection for end-to-end training. The framework can dynamically select different enhancement subnetworks for each sample to improve the performance of the detector. Our proposed method consists of two stage: the enhancement stage and the detection stage. The enhancement stage dynamically enhances the low-light images under the supervision of several enhancement methods and output corresponding weights. During the detection stage, the weights offers information on object classification to generate high-quality region proposals and in turn result in accurate detection. Our experiments present promising results, which show that the proposed method can significantly improve the detection performance in low-light environment.

Overall Framework

As is illustrated in Fig. 1, the proposed method mainly consists of two stages: the enhancement stage and the detection stage. We unify these two stages in one framework for end-to-end joint optimization.

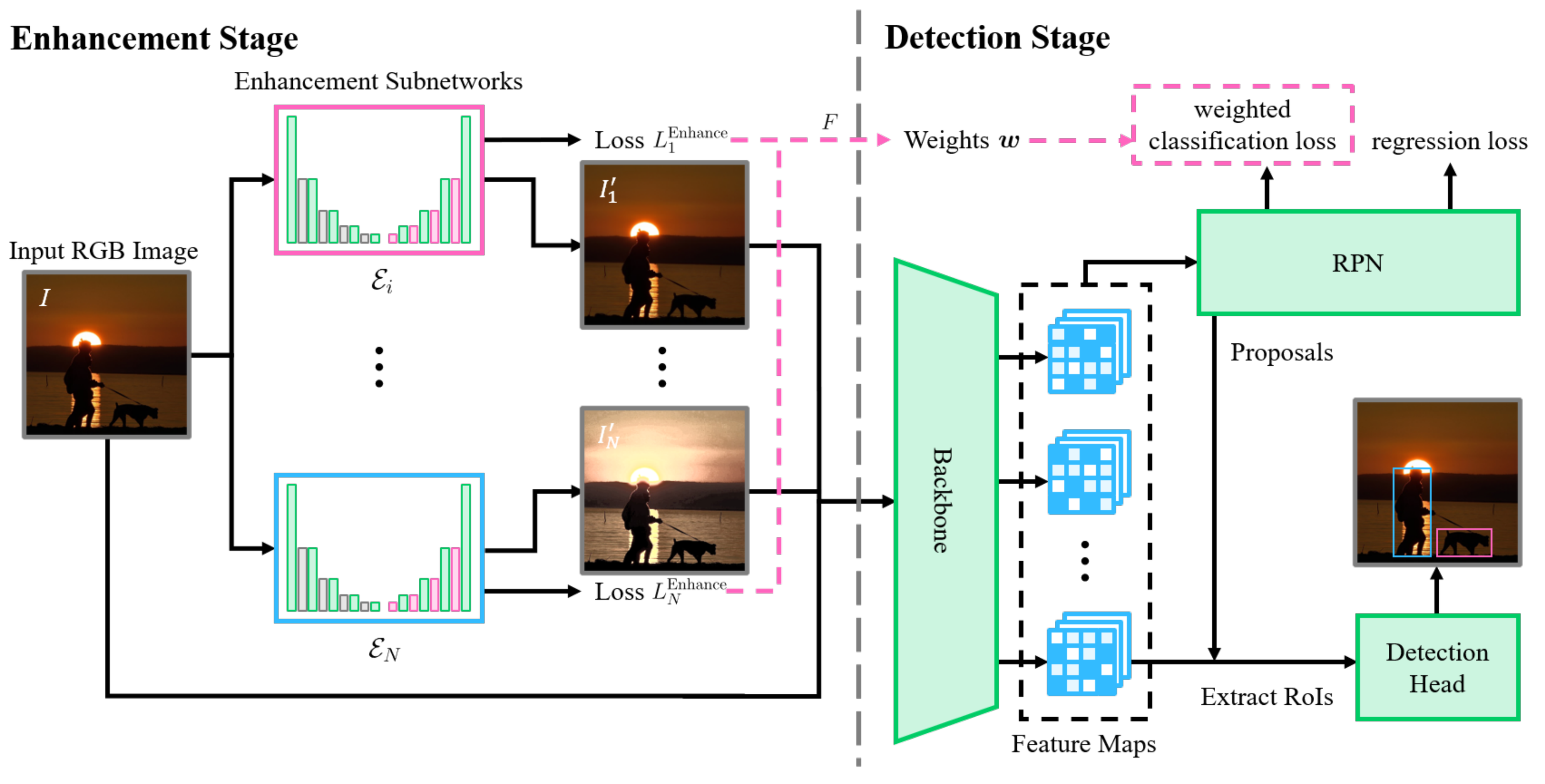

Fig. 1. Our proposed end-to-end framework. We show the end-to-end framework of multiple enhancement subnetworks and the detector. I refers to the input of the enhancement stage. $I_1', \ldots, I_N'$ denote the output of each enhancement subnetwork, which is supervised by different enhancement methods.

In the enhancement stage, we use a dynamic filter networks to generate sample-specific convolution kernels. These convolution kernels are used to dynamically enhance low-light images. We employ common enhancement methods to constrain the behavior of each enhancement subnetwork so that the model can adaptively choose the most effective enhancement methods. In the detection stage, we use a variant of Faster RCNN to perform object detection based on the enhanced images generated by the enhancement stage. We assign weights to the classification losses of RPN to improve the classification performance at this stage. The weight is calculated from the losses in the enhancement stage, which represents the importance of each enhancement subnetwork for each sample.

Enhancement Network

We propose a dynamic enhancement network composed of several sub-networks which are independent of each other.

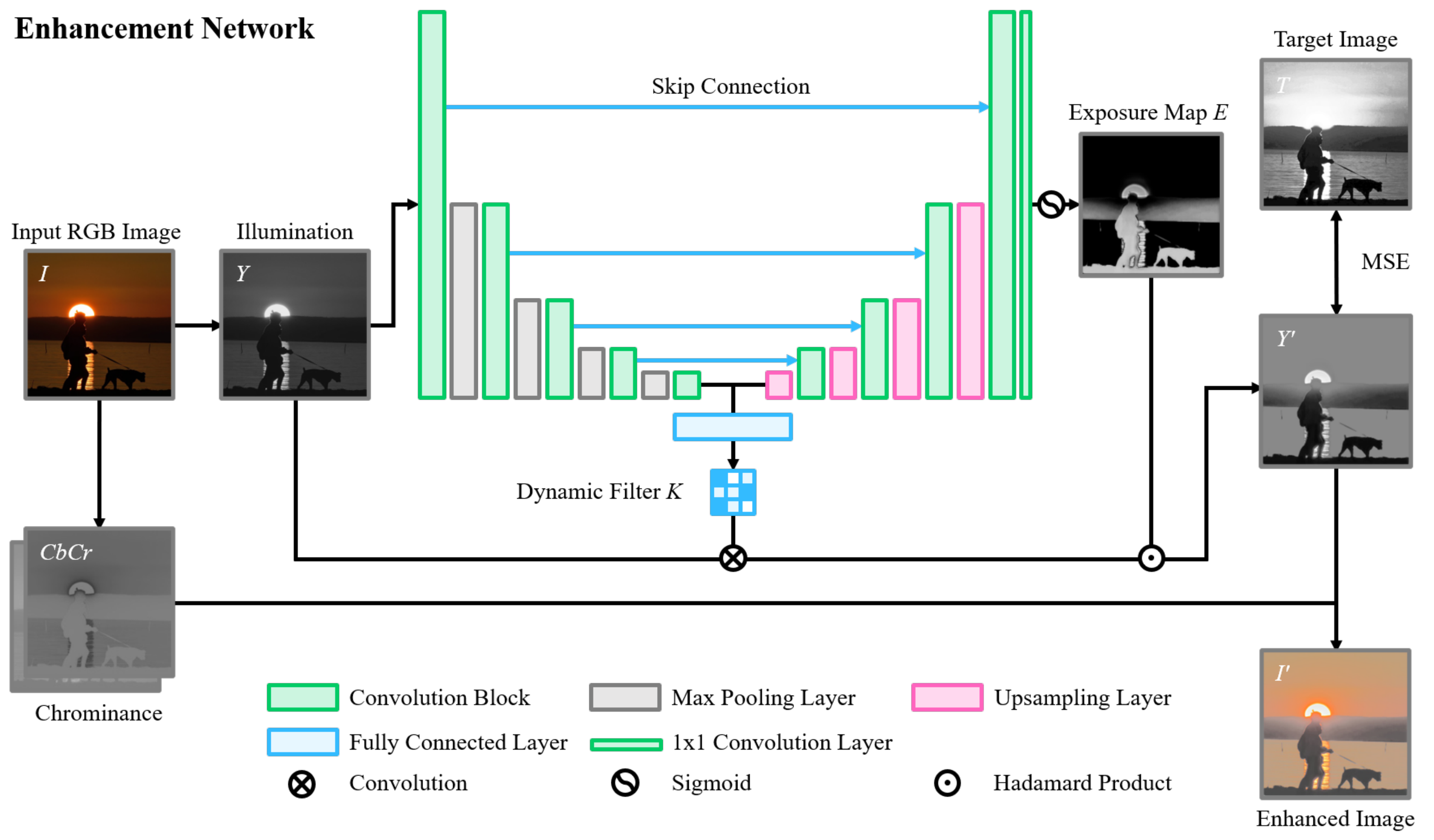

Fig. 2. The architecture of the enhancement networks. The heights of the rectangles corresponding to different layers indicate the scale of the corresponding feature maps. For clarity, we have omitted the reshape operations of the input and output of U-Net-like architecture. Convolution block consists of two consecutive convolutional layers activated by ReLU. During the downsampling process, the number of channels of the feature map is continuously increased, from 1 to 512, and the upsampling process is the opposite.

In Fig. 2, we illustrate the detailed architecture of the enhancement network. The enhancement network includes two components: the dynamic filter generator and the adaptive exposure module (AEM). The dynamic filter is designed to simulate a specific enhancement method, and AEM is to further activate areas in the image that are critical to improve detection performance. We proposed a convolutional architecture based on U-Net to combines the two parts in one subnetwork. In this way, the two modules can share the feature maps to reduce the computation cost.

Experiments

TABLE I
QUANTITATIVE EVALUATION OF THE PROPOSED METHOD.

	Enhanced Channel(s)*	AP	AP_{50}	AP_{75}	AP_S	AP_M	AP_L
RetinaNet	-	27.6	52.7	25.9	4.5	16.0	31.8
Faster R-CNN w FPN	-	30.4	61.0	27.3	4.3	**18.8**	35.2
Bilateral Filter (BF)	Y	27.8	57.6	23.2	2.2	17.3	32.0
Guided Filter (GF)	Y	25.8	53.9	21.7	1.9	14.6	30.2
Histogram Equalization (HE)	Y	28.4	57.4	25.6	2.6	17.6	32.7
Image Sharpening (IS)	Y	29.0	59.2	25.7	2.9	17.2	34.0
Loh *et al.*	Y	29.0	58.4	25.4	5.2	17.4	33.2
EnlightenGAN (EGAN)	Y	29.2	59.7	25.5	4.5	18.1	33.6
Loh *et al.*	RGB	27.5	55.8	23.3	4.3	16.6	31.6
EnlightenGAN (EGAN)	RGB	29.4	58.8	26.1	6.8	18.6	33.8
Proposed Method (based on Loh *et al.*'s & EGAN)	Y	31.6	61.7	28.8	**7.4**	18.5	36.3
Proposed Method (based on HE & IS)	Y	**32.1**	**62.1**	**29.9**	5.4	**18.8**	**36.4**

* Enhanced channel means which channel we apply enhancement methods. Y denotes the illumination component in YCbCr color space and RGB indicates all the channels in RGB color space.

All the quantitative results are shown in Table I. we show the results of two classic one-stage and two-stage object detection algorithms. Second, we independently compare the effects of several image enhancement methods on the detection results. Finally, we apply our method to untrainable and trainable methods mentioned before. For the untrainable methods (HE and IS), despite their poor overall performance, they still performed well on some samples. The advantage of untrainable methods is that they usually do not have a large computational cost and do not require complicated adjustment for hyperparameters. Experimental results show that our method has good compatibility with these methods and can further improve the detection performance on the basis of them.