Boosting Object Detection with Zero-Shot Day-Night Domain Adaptation

Supplementary Material

(M, N)	mAP(%)
(M=2, N=2)	**23.5**
(M=1, N=2)	21.9
(M=4, N=2)	22.3
(M=2, N=1)	22.4
(M=2, N=4)	20.3
(M=2, N=7)	18.6

Table 5. Parameter variation (M, N) in reflectance decoder.

Method	mAP(%)
Ours	23.5
w/ RF	24.2

Table 6. Alternative of Retinex decomposition net.

Method	mAP(%)
Baseline	15.2
$+\mathcal{L}_{ref}$	18.9
$+\mathcal{L}_{decom}$	**20.5**

Table 7. Analysis on reflectance decoding losses.

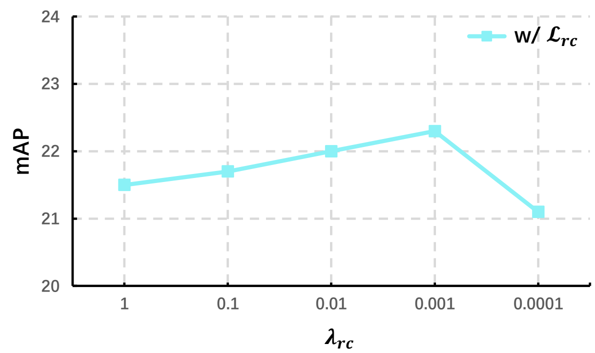

Figure 6. Parameter variation of loss weight λ_{rc} for redecomposition cohering loss.

This supplementary material provides 1) more implementation details; 2) additional ablation study; 3) more examples for visualization and comparison.

6. Implementation Details

We implement our method with PyTorch 1.11.0 and all experiments are conducted on four V100 GPUs, except for the image classification task we use a single V100 GPU.

7. Ablation Study

The experiments below are performed on WIDER FACE $\rightarrow$ DARK FACE using the DSFD detector [29] as the same to Sec. 4.1 in the paper.

More results for reflectance decoder. In the paper, we report that the best performance occurs when the reflectance decoder is appended after the M-th conv layer of backbone and it consists of N conv layers. Both M and N are set to 2 by default. Here, we conduct parameter variation over different M and N in Table 5.

Variation of M. We first fix N and conduct experiments on different M. The result shows that either increasing or decreasing M would slightly degrade the performance. $M = 2$ performs the best.

Variation of N. Next, we fix M to 2 and vary the value of N to find whether a bigger N benefits the reflectance decoder. When N increases, the performance actually drops. The default one $N = 2$ performs the best.

Alternative of the Retinex decomposition net. The Retinex decomposition net (see Sec. 3.4 and Fig. 2 in the paper) can be replaced with other methods with similar functions. Here, we replace it with a recent development called RetinexFormer [4] in the paper. RetinexFormer explicitly models noises in reflectance and illumination maps through a transformer architecture, thus offering more robust pseudo ground truth than the original RetinexNet [69] used in the paper. We achieve better result by using RetinexFormer (w/ RF) in Table 6. This illustrates the generalizability of our method with a stronger decomposition net.

Analysis of the reflectance decoding losses. Referring to Sec. 3.4-DAI-Net in the paper, for the optimization of reflectance decoder, we also add the reflectance learning loss $\mathcal{L}_{ref}$ and image decomposition loss $\mathcal{L}_{decom}$. In Table 7, we give an analysis by adding these losses sequentially to our Baseline (see Sec. 4.4 in the paper). Every employed loss contributes clearly to the overall performance.

More results for redecomposition cohering loss. We vary the loss weight λ_{rc} for redecomposition cohering loss in Fig. 6. Ranging from 0.0001 to 1 for λ_{rc}, we observe that using the redecomposition cohering loss consistently shows better performance compared to not using it (20.5 in Table 4 in the paper). $\lambda_{rc} = 0.001$ achieves the largest improvement.

8. Visualization

Qualitative examples. We present additional qualitative examples of our method on WIDER FACE $\rightarrow$ DARK FACE and COCO $\rightarrow$ ExDark in Fig. 7 and Fig. 8, respectively.

Reflectance visualization. Given an image I from the DARK FACE dataset, we show the pseudo ground truth $\{\hat{R}, \hat{L}\}$ and the decomposed reflectance R from the reflectance decoder of our DAI-Net in Fig. 9. More visualizations of R are given in Fig. 10 (Top). We also provide the two-round decomposition results on LOL v2. We show them in Fig. 10 (Bottom). The generated reflectances are consistent between well-lit and low-light image pairs, and also between the two rounds.

Figure 7. Qualitative examples of DSFD [29] and our method. Images are taken from the DARK FACE dataset and are enhanced only for visualization.

Figure 8. Qualitative examples of our method on the ExDark dataset. Predicted categories and their confidence scores are given along with bounding boxes in each image. Zoom in for details.

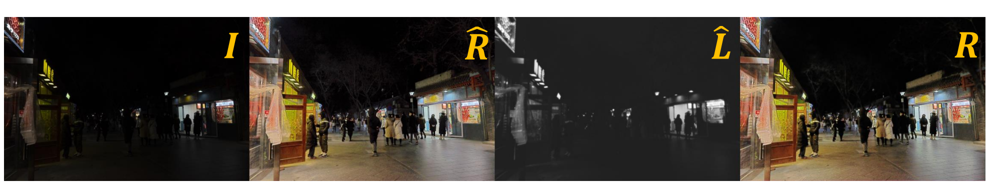

Figure 9. Visualization of original image, pseudo ground truth, and predicted reflectance (from left to right).

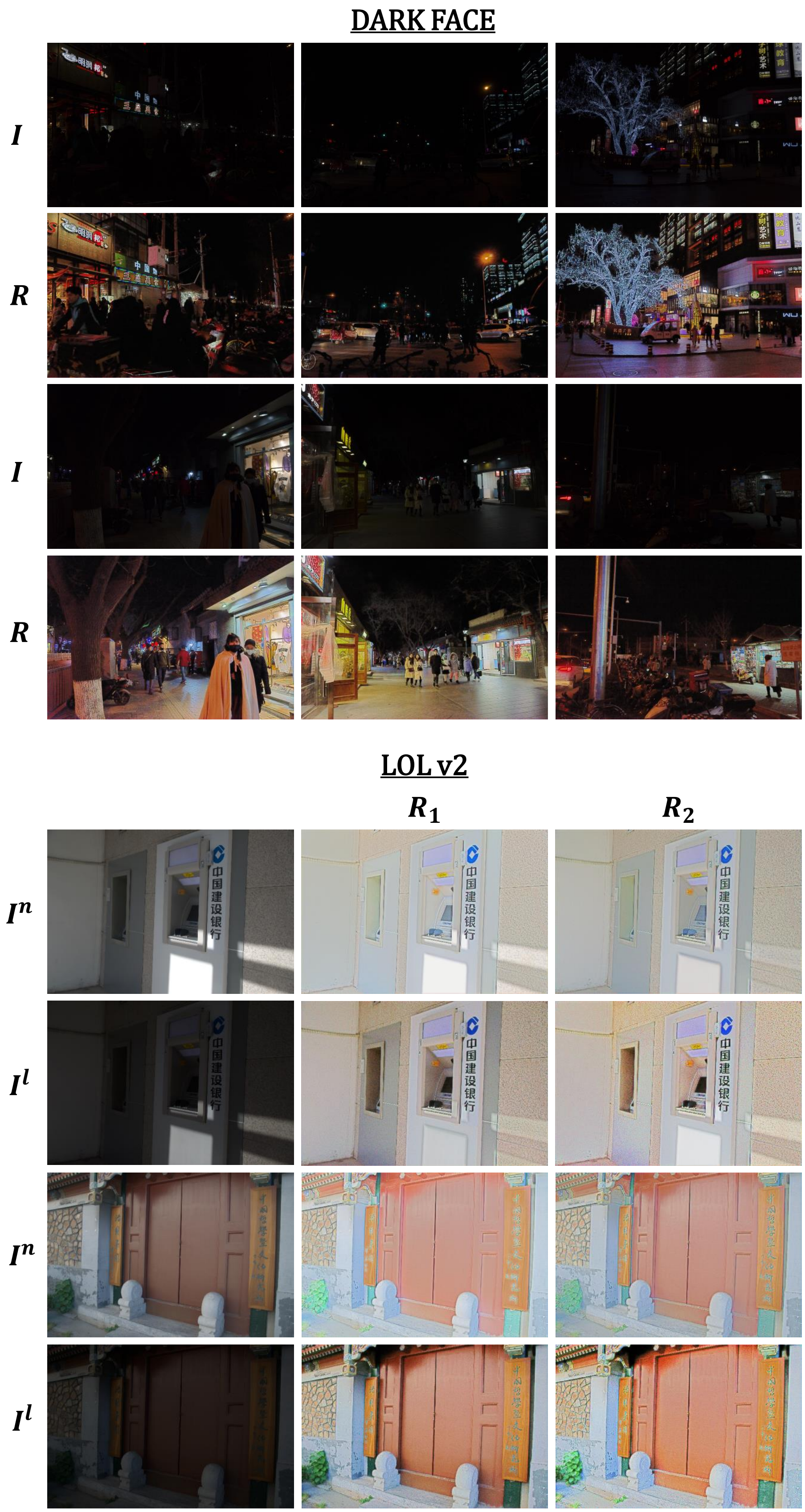

Figure 10. Reflectance visualization. We show the predicted reflectances R over extreme low-light images I from the DARK FACE dataset. We also give two-round decomposed reflectances $R_1^n, R_1^l, R_2^n, R_2^l$ of real paired images I^n, I^l from LOL v2.