Supplemental Materials of Retinex-inspired Unrolling with Cooperative Prior Architecture Search for Low-light Image Enhancement

Risheng Liu[1,3], Long Ma[2,3], Jiaao Zhang[2,3], Xin Fan[1,3], Zhongxuan Luo[2,3]
[1]International School of Information Science & Engineering, Dalian University of Technology
[2]School of Software Technology, Dalian University of Technology
[3]Key Laboratory for Ubiquitous Network and Service Software of Liaoning Province
{rsliu, xin.fan, zxluo}@dlut.edu.cn, {longma, jiaaozhang}@mail.dlut.edu.cn

Abstract

In this supplementary material, we demonstrate more details on the implementation procedure. Besides, we provide more visual comparisons of algorithmic analysis to further evaluate the effectiveness of our proposed RUAS. Next, we present more visual results on different challenging datasets among state-of-the-art approaches and our RUAS. Finally, we execute the task of face detection in the dark to evaluate the practical values of low-light image enhancement methods.

1. Implementation Details

We considered the same search space (with 3 fundamental cell structures) for IEM and NRM, but defined their cells with different channel widths (i.e., 3 for IEM and 6 for NRM). During the search period, we searched the architectures for IEM and NRM by using our cooperation strategy. As for the parameter setting of the search process, the maximum epoch was set as 20, the batch size was set as 1, the initial learning rate was 3×10^{-4}, the momentum was randomly sampled from (0.5, 0.999) and the weight decay was 10^{-3}. After the search process, we obtained the network parameters by training the IEM and the NRM cooperatively. That is, following every iteration of the NRM, the IEM iterated 50 times to update its parameters. In the process, the learning rates for IEM and NRM were set to 0.015 and 0.0001, respectively. The momentums were set to 0.9 equally. We trained the IEM and the NRM for 800 iterations with a batch size of 1. Additionally, we used the synthesized noisy images, generated by adding Gaussian noises to clean images for pre-training the NRM.

2. Visual Comparison of Naively Determined Architectures and Our RUAS

We have demonstrated quantitative evaluation of naively determined architecture and our RUAS in the manuscript (see Fig. 8). We considered seven types of naively determined architectures including supernet and all the operations of a cell are the 1×1 convolution (1-C), the 3×3 convolution (3-C), the 1×1 residual convolution (1-RC), the 3×3 residual convolution (3-RC), the 3×3 dilation convolution with the dilation rate of 2 (3-2-DC), the 3×3 residual dilation convolution with the dilation rate of 2 (3-2-RDC), respectively. Fig. 1 presented visual results by using these architectures. Obviously, most of the naively determined architectures tended to produce strikingly over-exposure results (e.g., supernet, 1-C, 1-RC, 3-RC, 3-2-RDC). 3-C and 3-2-DC all failed to perform an ample brightness improvement. These appearances actually indicated the dramatic drawback of heuristically-designed architectures, i.e., they were supposed to conduct extensive experimental verifications to determine the best architecture, otherwise, the performance could not be guaranteed. In contrast, our searched architecture could realize a better visual expression.

3. Visual Comparison of Different Search Strategies

In our manuscript, we have provided the comparison in terms of searched architectures and quantitative results (see Fig. 9) of different search strategies. Here we further showed some visual comparisons in Fig. 2. In which, the separate search strategy is to search IEM and NRM one by one, that is, when searching for the NRM, the search procedure of the IEM has

Figure 1. Visual comparison of naively determined architectures and our proposed RUAS on the MIT-Adobe 5K dataset [1].

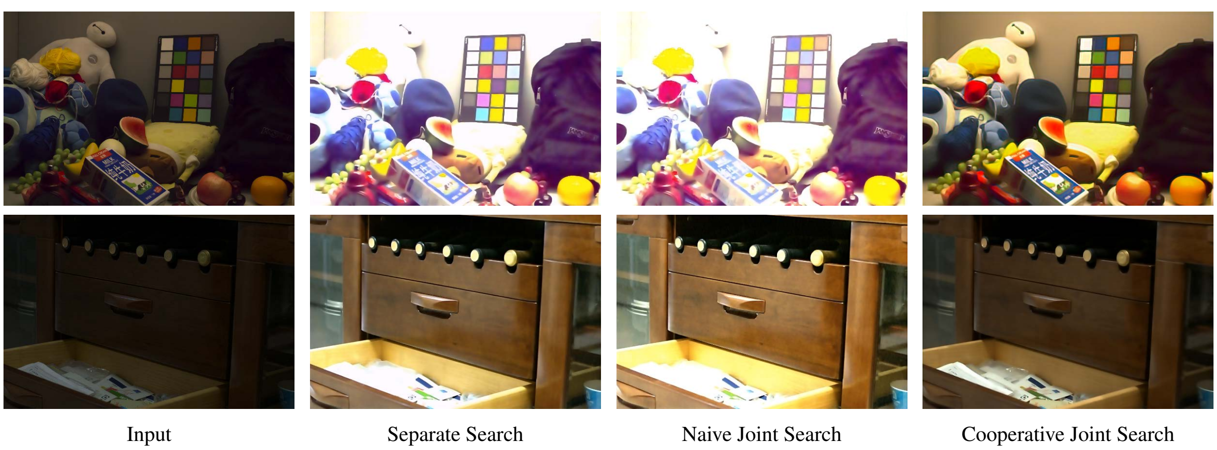

Figure 2. Visual comparison of different search strategies on the LOL dataset [2].

ended. The naive joint search is to view the IEM and NRM as parts of an entire network architecture, and just need to search for all the architecture once. The cooperative joint search is our proposed new strategy. We could see that separate and naive joint search strategies were difficult to provide a satisfying exposure improvement. A key reason was that they ignored the latent relationship between illumination estimation and noise removal. In contrast, our proposed cooperative search strategy could mutually assist in optimizing these two steps in a cooperative manner, achieving a prominent enhanced performance.

4. More Visual Comparisons

Here, we provided more visual comparisons by comparing our RUAS with recently-proposed state-of-the-art deep networks including RetinexNet [2], EnGAN [4], SSIENet [11], KinD [12], DeepUPE [6], ZeroDCE [3], FIDE [7], and DRB-N [9]. Fig. 3 and 4 demonstrated the enhanced results on the MIT-Adobe 5K dataset [1]. Fig. 5 showed the visual comparisons on the DarkFace dataset [10]. Fig. 6 presented the visual results on the ExtremelyDarkFace dataset (used as the sub-challenge in the CVPR 2020 UG2+Challenge[1]). From these evaluations, we can easily observe that our RUAS significantly outperformed those advanced deep networks.

5. Face Detection in the Dark

Finally, we also performed the task of face detection under low-light conditions, to evaluate the practical effects of the low-light image enhancement methods. We utilized the baseline face detection model, Dual Shot Face Detection (DSFD) [5] which was trained on the WIDER FACE dataset [8]. Table 1 reported the quantitative result (mAP and LAMR), we could easily see that our RUAS attained the highest numerical scores both in mAP and LAMR. Further, we demonstrated visual comparisons in Fig. 7. Obviously, our RUAS was remarkably superior to other methods, especially in the zoomed-in regions. This experiment fully verified our superiority and potential application prospect.

Table 1. The quantitative performance of face detection on the DarkFace dataset [10].

Methods	Input	RetinexNet	EnGAN	SSIENet	KinD	DeepUPE	ZeroDCE	FIDE	DRBN	Ours
mAP	44.52%	46.51%	46.34%	48.04%	44.34%	48.92%	49.07%	41.02%	41.35%	**50.13%**
LAMR	74%	74%	75%	74%	75%	73%	73%	76%	75%	**72%**

References

[1] Vladimir Bychkovsky, Sylvain Paris, Eric Chan, and Frédo Durand. Learning photographic global tonal adjustment with a database of input/output image pairs. In *CVPR*, pages 97–104, 2011. 2, 3

[2] Wei Chen, Wenjing Wang, Wenhan Yang, and Jiaying Liu. Deep retinex decomposition for low-light enhancement. In *BMVC*, 2018. 2, 3

[3] Chunle Guo, Chongyi Li, Jichang Guo, Chen Change Loy, Junhui Hou, Sam Kwong, and Runmin Cong. Zero-reference deep curve estimation for low-light image enhancement. In *CVPR*, pages 1780–1789, 2020. 3

[4] Yifan Jiang, Xinyu Gong, Ding Liu, Yu Cheng, Chen Fang, Xiaohui Shen, Jianchao Yang, Pan Zhou, and Zhangyang Wang. Enlightengan: Deep light enhancement without paired supervision. *IEEE TIP*, 30:2340–2349, 2021. 3

[5] Jian Li, Yabiao Wang, Changan Wang, Ying Tai, Jianjun Qian, Jian Yang, Chengjie Wang, Jilin Li, and Feiyue Huang. Dsfd: dual shot face detector. In *CVPR*, pages 5060–5069, 2019. 3

[6] Ruixing Wang, Qing Zhang, Chi-Wing Fu, Xiaoyong Shen, Wei-Shi Zheng, and Jiaya Jia. Underexposed photo enhancement using deep illumination estimation. In *CVPR*, pages 6849–6857, 2019. 3

[7] Ke Xu, Xin Yang, Baocai Yin, and Rynson WH Lau. Learning to restore low-light images via decomposition-and-enhancement. In *CVPR*, pages 2281–2290, 2020. 3

[8] Shuo Yang, Ping Luo, Chen-Change Loy, and Xiaoou Tang. Wider face: A face detection benchmark. In *CVPR*, pages 5525–5533, 2016. 3

[9] Wenhan Yang, Shiqi Wang, Yuming Fang, Yue Wang, and Jiaying Liu. From fidelity to perceptual quality: A semi-supervised approach for low-light image enhancement. In *CVPR*, pages 3063–3072, 2020. 3

[10] Wenhan Yang, Ye Yuan, Wenqi Ren, Jiaying Liu, Walter J Scheirer, Zhangyang Wang, Taiheng Zhang, Qiaoyong Zhong, Di Xie, Shiliang Pu, et al. Advancing image understanding in poor visibility environments: A collective benchmark study. *IEEE TIP*, 29:5737–5752, 2020. 3

[11] Yu Zhang, Xiaoguang Di, Bin Zhang, and Chunhui Wang. Self-supervised image enhancement network: Training with low light images only. *arXiv preprint arXiv:2002.11300*, 2020. 3

[12] Yonghua Zhang, Jiawan Zhang, and Xiaojie Guo. Kindling the darkness: A practical low-light image enhancer. In *ACM MM*, 2019. 3

[1] http://cvpr2020.ug2challenge.org/dataset20_t1.html

Figure 3. Visual comparison on the MIT-Adobe 5K dataset.

Figure 4. Visual comparison on the MIT-Adobe 5K dataset.

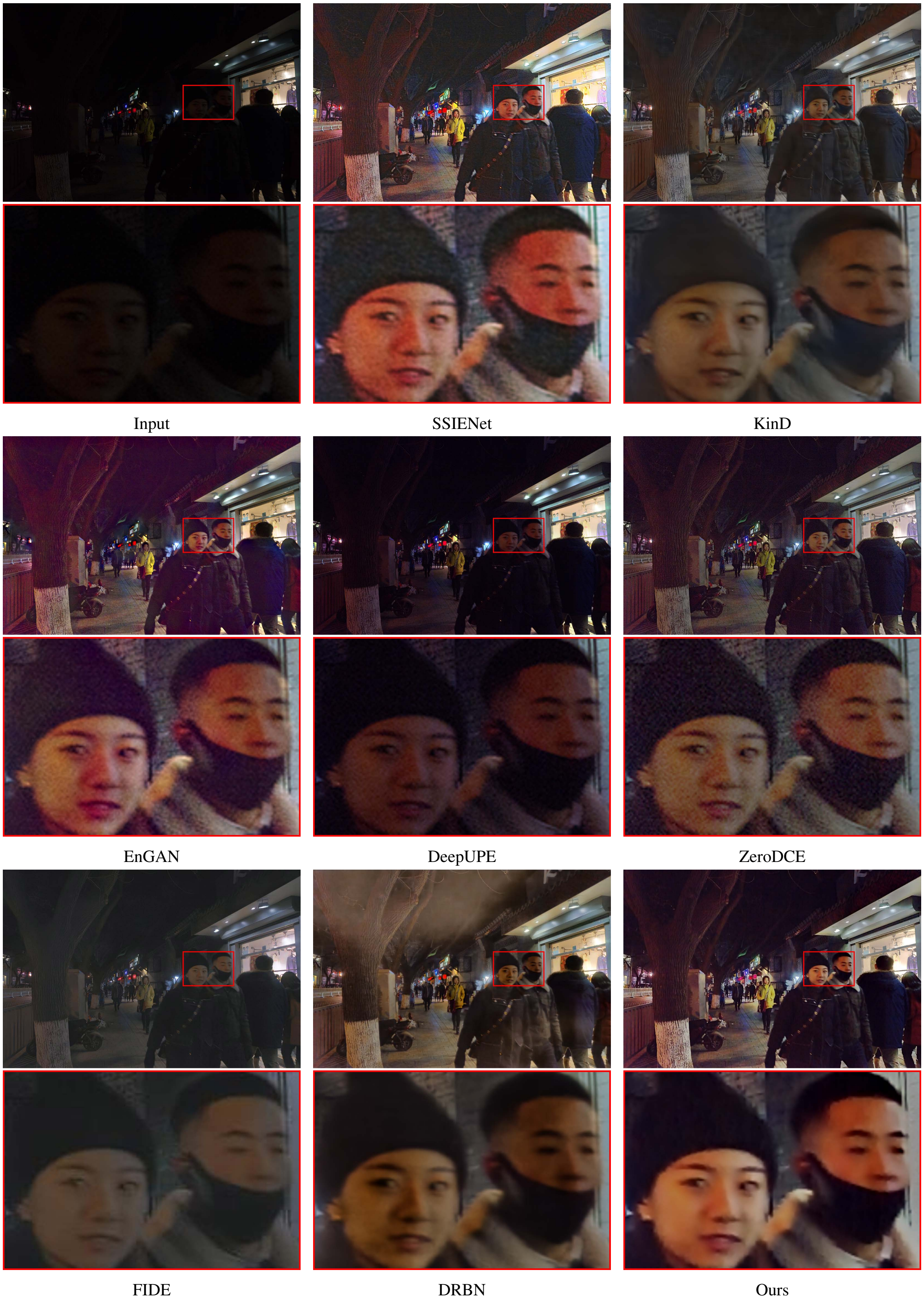

Figure 5. Visual comparison on the DarkFace dataset.

Figure 6. Visual comparison on the ExtremelyDarkFace dataset.

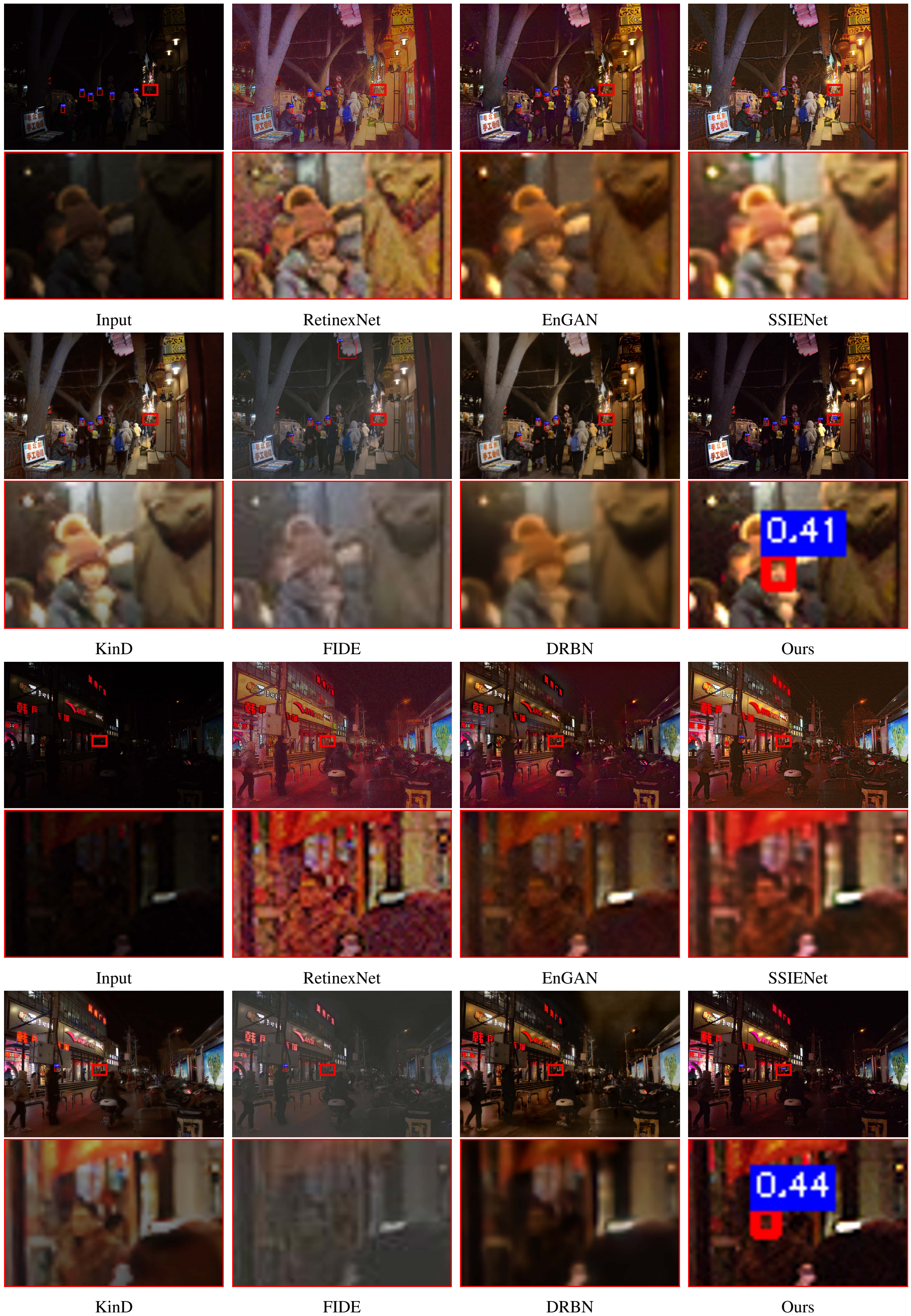

Figure 7. Visual comparison of face detection on the DarkFace dataset.