Implicit Neural Representation for Cooperative Low-light Image Enhancement – Supplemental Document –

Abstract

This is the supplementary material for the paper: "Implicit Neural Representation for Cooperative Low-light Image Enhancement". Firstly, we provide more results of our NRN module in ***Section A*** *for a comprehensive illustration. Besides, in* ***Section B****, we train the framework with different prompts and compare their performance with other ablation settings, which verifies the effect of text-driven supervision. In* ***Section C****, we further conduct ablation experiments on TAD to determine the role of each path. In* ***Section D****, to demonstrate the semantic advantage of our method, we classify the enhanced results of different methods with a pre-trained vision-language model and report their accuracy. Our results are considered best for textual description of high-light image. Finally, more qualitative analyses on three well-known benchmarks are displayed in* ***Section E****, including LSRW dataset, LOL dataset and LIME dataset. It is obvious that the proposed NeRCo achieves the best performance, further verifying our superiority.*

A. Normalized Results

Deep learning based models learn to map a sample from the input domain to the target domain. In real-world application, however, degradation conditions are various. For some inputs far from the learned input domain, it is hard for a trained model to perform stable superior performance. Hence, we developed Neural Representation Normalization (NRN) module to normalize different conditions, which has been illustrated in Sec. 3.2. In order to provide more convincing proof, here we add here more experimental results.

As shown in Fig. 8, we adopted three sets of images from the SICE [3] dataset, each set contains three images with different brightness and the same content. We box them with blue, red and green lines respectively. All of them are processed by our NRN module, the corresponding results are boxed with the same color. One can see that the brightness of original inputs varies a lot, while the output brightness of NRN is similar. For more intuitive, on the right, we provide a visualization of their pixel distribution on the Y channel. It is obvious that NRN constricts the the range of brightness changes and normalizes degradation levels.

B. Experiments with Alternative Textual Prompts

To investigate the contribution of our proposed text-driven supervision, we compare the performance of models trained on different prompts. Specifically, we consider three pairs of alternative prompts to guide model training: i) *dark* and *bright*. ii) *dim* and *light*. iii) *night* and *day*. These alteration experiments are conducted on the "#3" ablation setting mentioned in Sec. 4.3, which removes the neural representation function from the NeRCo.

As shown in Tab. 3, we report the results of different settings on LSRW dataset [13]. We design different prompts to study the impact of different texts on model performance. One can see that the "#3" settings with diverse prompts realize decent scores on all four metrics. Although their values are different, they are within a stable range, *i.e.*, better than other ablation settings and worse than NeRCo. On the one hand, we can see that text-driven supervision does have a gain in performance. On the other hand, it also indirectly proves the contribution of our proposed Neural Representation Normalization (NRN) module. Since *low-light image* and *high-light image* are two texts widely adopted to describe images in this task, we used this pair in other experiments for more intuitive validation.

C. TAD Ablation

TAD contains three paths: color discrimination, edge supervision, and text-driven discrimination. To further define the role of each component, we conduct ablation study on TAD, which removes different paths from the ablation setting "#3" in the submitted paper. Note that as at least one supervision is required, we adopt color discrimination as the base discriminator. Results are given in Tab. 5, one can see that both edge path and text supervision improve the effect.

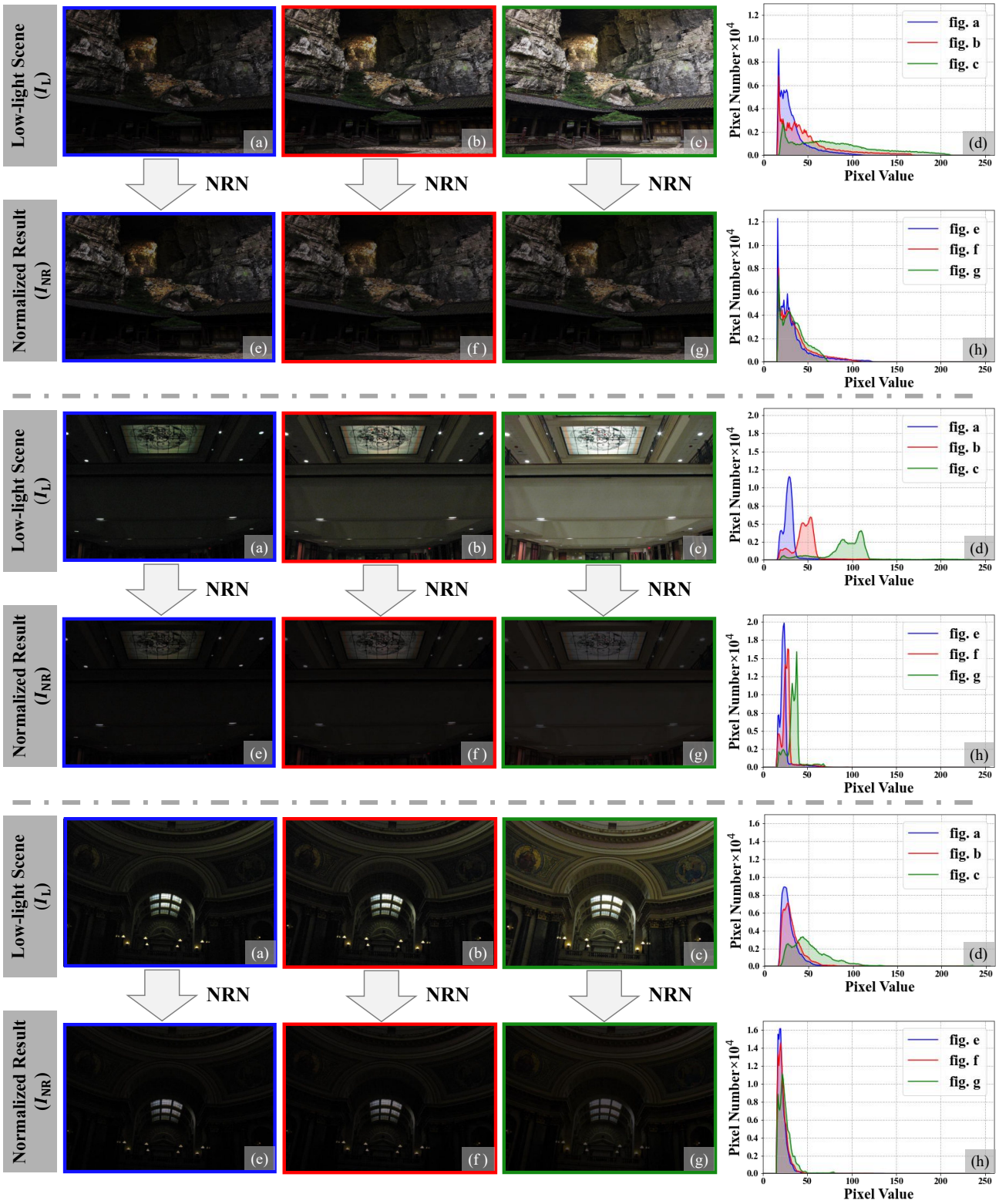

Figure 8: Comparisons between the captured low-light scenes ($\boldsymbol{I}_{\mathrm{L}}$) and the results of NRN ($\boldsymbol{I}_{\mathrm{NR}}$). The low-light samples are from the SICE [3] dataset. It contains numerous image sets, each with common content and varying degradation conditions. The pixel value distribution of images on the Y channel is given on the right. One can see that NRN normalizes the brightness to be similar.

Models	#1	#2	#3 (*low-light image / high-light image*)	#3 (*dark / bright*)	#3 (*dim / light*)	#3 (*night / day*)	NeRCo
PSNR ↑	16.77	17.65	18.32	**18.58**	18.23	18.56	**19.00**
SSIM ↑	0.4565	0.5023	0.5201	0.5118	0.5208	**0.5277**	**0.5360**
NIQE ↓	12.34	10.60	10.83	9.45	9.59	**9.33**	**9.23**
LOE ↓	272.4	247.9	230.9	202.9	**191.0**	206.6	**189.5**

Table 3: Ablation study with different prompt options. The best and the second best results are highlighted in **red** and **blue** respectively. One can see that settings trained with prompts outperform other versions, and prompts can be replaced with synonyms. It proves that text-driven supervision has a gain in model performance.

Datasets	Input	LECARM [39]	SDD [14]	RetinexNet [46]	KinD [55]	URetinexNet [47]	ZeroDCE [11]	SSIENet [53]	RUAS [25]	EnGAN [16]	SCI [28]	NeRCo	Reference
LOL [46]	0.1590	0.3685	0.3397	0.4831	0.5130	0.5554	0.3463	0.4639	0.4381	0.4450	0.3402	**0.6366**	**0.5910**
LSRW [13]	0.3164	0.6028	0.5907	0.6653	0.6052	0.6539	0.6584	0.6746	0.6418	0.5969	0.6176	**0.7581**	**0.6955**
LIME [12]	0.3281	0.5168	0.4669	0.5657	0.5589	**0.6265**	0.4719	0.6147	0.5427	0.4912	0.5781	**0.7499**	-

Table 4: The average semantic scores of different settings on three benchmarks. The best and the second best results are highlighted in **red** and **blue** respectively. One can see that the pre-trained vision-language model classifies our results more accurately than those of other methods, which demonstrates the better semantic consistency of our method.

Color Path	Edge Path	Text Supervision	PSNR ↑	SSIM ↑	NIQE ↓	LOE ↓
✔	✗	✗	17.37	0.4942	11.72	252.1
✔	✔	✗	**17.65**	**0.5023**	**10.60**	**247.9**
✔	✔	✔	**18.32**	**0.5201**	**10.83**	**230.9**

Table 5: Ablation study on TAD, based on the ablation setting "#3". We conduct experiments on LSRW dataset. The best and the second best results are highlighted in **red** and **blue** respectively.

D. Semantic Evaluation

In order to demonstrate the superiority of our NeRCo at the semantic level, we employ the pre-trained CLIP model [35] to calculate semantic score of different methods. Concretely, we first design a prompt *high-light image*. The image vector and the text vector are then generated by CLIP model. We calculate their cosine discrepancy, and use a softmax function to obtain the semantic score, which values from 0 to 1. A higher score represents the better semantic consistency between the enhanced image and the text *high-light image*.

We report the average prediction accuracy of the results from different methods in Tab. 4. One can see that the pre-trained CLIP considers the input low-light image to be the least likely high-light image, while the high-light references are classified accurately, except for LIME [12] which only contains degraded scenarios. Although some methods output semantically impressive results, our NeRCo achieves the best scores, even better than the ground truth. It proves that the quality of the reference images from the dataset is semantically good, as they achieve the second-best scores. Due to the text-driven discrimination during training, our method produces more perceptual-friendly results than references.

E. Qualitative Analysis

We have provided adequate quantitative results (Tab. 1) in our paper. However, due to the limit of space, only parts of visual comparisons are given (Fig. 6). Here, we supplement more qualitative analysis compared with other SOTA methods, including LECARM [39], SDD [14], RetinexNet [46], KinD [55], URetinex-Net [47], ZeroDCE [11], SSIENet [53], RUAS [25], EnGAN [16], and SCI [28].

Fig. 9 displays the enhanced results on LSRW dataset. One can see that conventional model-based methods cannot recover sufficient brightness, while some other comparison methods suffer from color cast. RetinexNet, KinD, ZeroDCE, and RUAS, *etc.* develop the post-processing denoising operations to remove the inherent noise in dark regions, but they tend to discard details. In general, our NeRCo is capable of color adjustment and detail preservation, demonstrating its superiority over other algorithms. Furthermore, we provide visual comparisons between our proposed NeRCo and other SOTA mthods on other well-known benchmarks. Fig. 10 shows the comparisons on LOL dataset and Fig. 11 displays the qualitative results on LIME dataset. Obviously, across all these comparisons, our method recovers the most authentic tones and provides visual-friendly results, which proves its effectiveness.

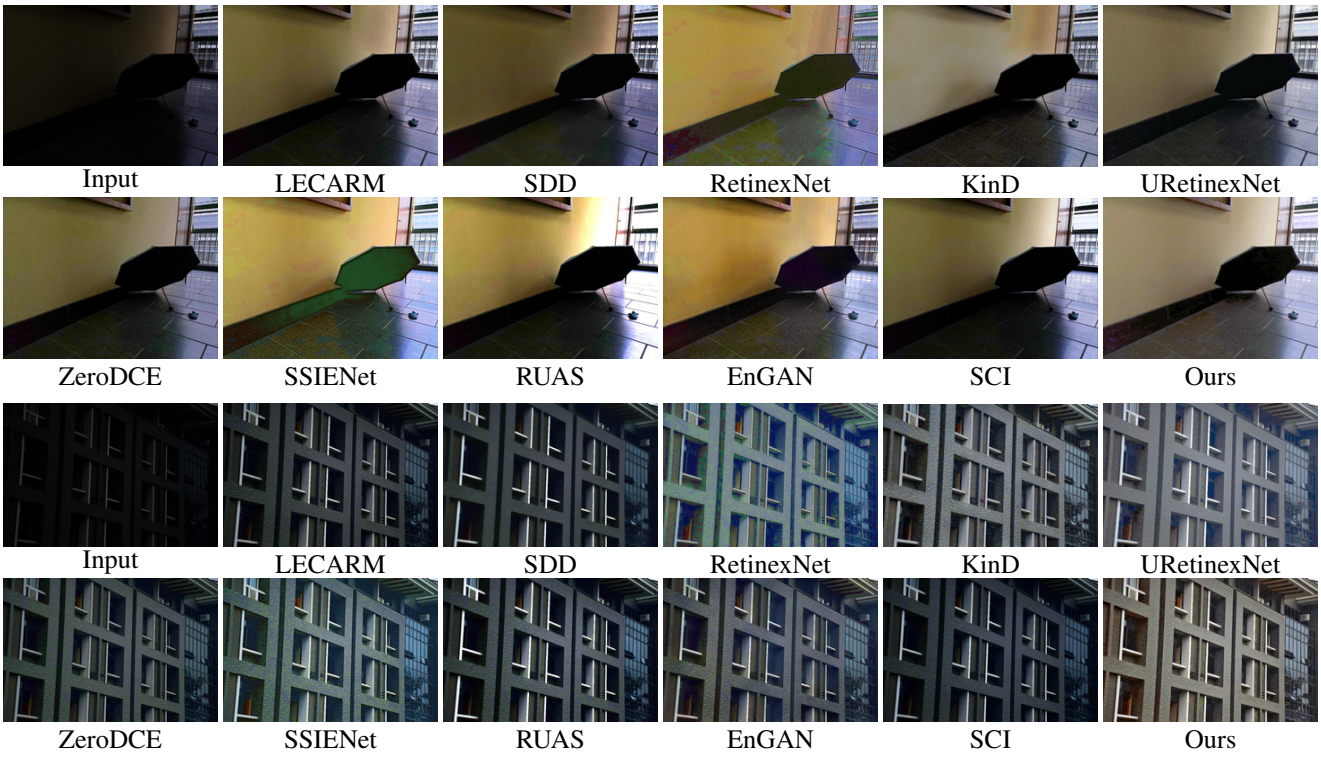

Figure 9: Subjective comparison on the LSRW dataset among state-of-the-art low-light image enhancement algorithms. Obviously, the proposed method has achieved the best performance, further verifying its effectiveness.

Figure 10: Subjective comparison on the LOL dataset among state-of-the-art low-light image enhancement algorithms. It is obvious that our method recovers the most authentic results, demonstrating its superiority.

Figure 11: Subjective comparison on the LIME dataset among state-of-the-art low-light image enhancement algorithms. Our model still performs best on this low-light image-only dataset, which proves its effectiveness.