Supplementary Material for Real-Time Exposure Correction via Collaborative Transformations and Adaptive Sampling

Ziwen Li[1], Feng Zhang[1], Meng Cao[2], Jinpu Zhang[1], Yuanjie Shao[3], Yuehuan Wang[1]*, Nong Sang[1]
[1]National Key Laboratory of Multispectral Information Intelligent Processing Technology,
School of Artificial Intelligence and Automation, Huazhong University of Science and Technology
[2]Mohamed bin Zayed University of Artificial Intelligence
[3]School of Electronic Information and Communication, Huazhong University of Science and Technology
{D201980722,fengzhangaia,shaoyuanjie,yuehwang,nsang}@hust.edu.cn, {mengcaopku,zjphust}@gmail.com

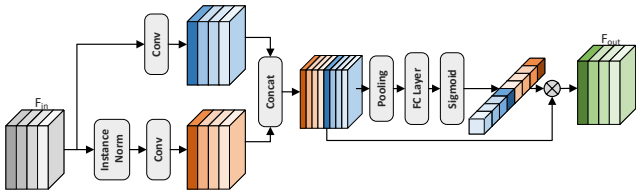


Figure 1. Architectural design of an improved ENC module.

Layers	Input Shape	Output Shape
Conv, LeakyReLU, IN	$3 \times 256 \times 256$	$8 \times 128 \times 128$
Conv, LeakyReLU, IN	$8 \times 128 \times 128$	$16 \times 64 \times 64$
Conv, LeakyReLU, IN	$16 \times 64 \times 64$	$32 \times 32 \times 32$
Conv, LeakyReLU, IN	$32 \times 32 \times 32$	$64 \times 16 \times 16$
Conv, LeakyReLU	$64 \times 16 \times 16$	$64 \times 8 \times 8$
AvgPooling	$64 \times 8 \times 8$	$64 \times 2 \times 2$
Flatten	$64 \times 2 \times 2$	256

Table 1. Detailed network architecture of the backbone for global transformation and adaptive sampling.

Layers	Input Shape	Output Shape
FC	256	M
FC	M	$3N^3$
Reshape	$3N^3$	$3 \times N \times N \times N$

Table 2. Architecture for predicting the 3D LUT. M and N denote the number and dimension of the basis 3D LUTs, respectively. In our experiments in the main text, $M = 3$ and $N = 17$.

Layers	Input Shape	Output Shape
FC, Split	256	$K_x - 1, K_y - 1$
Softmax	$K_x - 1, K_y - 1$	$K_x - 1, K_y - 1$
Accumulation	$K_x - 1, K_y - 1$	K_x, K_y
Cartesian product	K_x, K_y	$K_x \times K_y$

Table 3. Architecture for predicting the sampling grid. K_x and K_y denote the sample size along the x- and y- axes directions, respectively. In our experiments in the main text, $K_x = K_y = 256$.

In the supplementary material, we provide detailed network architectures, more experimental comparisons and future works.

1. Detailed Network Architecture

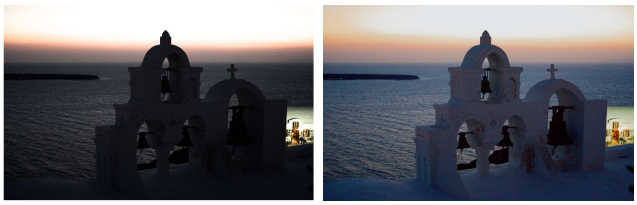
In order to ensure the generality of the proposed framework, we use simple and effective network architectures without bells and whistles.

In the pixel-wise transformation, we introduce an improved version of exposure normalization and compensation (ENC) module to reduce the disparity between different exposure features. The detailed structure is shown in Figure 1. Our ENC module first uses instance normalization to normalize the exposure to the exposure-invariant space. Then, we integrate the normalized features with the original features by channel attention in the channel dimension only. Compared to the original ENC module [3], we remove spatial interactions, which result in a large number of computations but limited performance gain.

For global transformation and adaptive sampling, we use the same backbone to analyze the content of the input image, which is the key to obtaining image-adaptiveness. The detailed network architecture is listed in Table 1. Specifically, the network employs only 5 convolutional layers to obtain an holistic understanding of the image content. Subsequently, we employ an adaptive average pooling layer and a flatten operation to obtain a compressed contextual representation.

For the prediction of 3D LUTs, we use 2 fully connected

*Corresponding author



Input Ours

Figure 2. A failure case of our method for challenging scenes.

(FC) layers to accomplish this process, as shown in Table 2. The first FC layer is used to predict the fusion weights. The weights of the second FC layer represent the basis 3D LUTs. After linear combination of the second FC layer, we can get the final image-adaptive 3D LUT.

For the prediction of sampled grids, we use only one FC layer to accomplish this process, as detailed in Table 3. We use FC layers for the extracted contexts to obtain the sampling intervals in the x- and y- axis directions. Then accumulation, normalization and Cartesian product are applied to obtain the final sampling coordinate points.

2. More Experimental Results

2.1. Comparisons on NIQE metric

In this subsection, we further employ the no-reference image quality assessment metric NIQE to evaluate different methods on the LCDP dataset. As shown in Table 4, our method achieves the lowest NIQE scores, which means a higher perceptual quality and naturalness compared to existing state-of-the-art methods.

2.2. Exposure Consistency Comparisons

In this subsection, we perform an exposure consistency comparison with the state-of-the-art methods, as shown in Figure 3. The first and second rows represent the enhancement results for the underexposed and overexposed images, respectively. The third row represents the error map of the two corresponding results. It can be clearly seen from Figure 3 that our method is able to generate results with more favorable illumination. In addition, for input images with different exposures, our method is able to recover more similar exposure levels and achieve better exposure consistency.

2.3. More Results on UHD images

In this subsection, we show some visualization results on UHD images from the original SICE [2] dataset. Most of the methods fail to process UHD images due to out-of-memory. We provide visual comparison results with the image-adaptive 3D LUT method [5] as presented in Figure 4 . Compared with the global transformation of 3D LUTs, our method is able to produce more pleasing visual results with better local contrast in some local regions with the help of the collaborative transformation framework.

2.4. More Visual Comparisons.

In this subsection, we provide more visual comparison results with the state-of-the-art methods. Specifically, Figure 5 and Figure 6 present the results on the SICE [2] dataset. Figure 7 and Figure 8 show the visual results on the MSEC [1] dataset. Figure 9 and Figure 10 present the results on the LCDP [4] dataset. The input images cover a variety of lighting conditions. Our proposed method is able to reconstruct satisfactory visual results with reasonable illumination and harmonious colors. These competitive visual results demonstrate the superiority of our method.

2.5. Runtime on CPU.

In this section, we report the running time of our method on the CPU in Table 5. It can be clearly seen that even on CPUs, our method strikes a better balance between efficiency and performance compared to other state-of-the-art methods.

3. Limitations and Future Work

We will continue to strive for higher image quality in the future. Figure 2 illustrates a failure case of our method for challenging scenes. For extreme exposure regions where structural information is lost, transformation-based methods may be difficult to recover. In the future, we will explore generative models to handle such challenging situations. In addition, to ensure the generality of the proposed framework, we only borrowed simple CNN networks as the backbone. More effective architectural designs, such as Transformer, deserve further investigation. We also plan to extend our approach to video processing and consider both temporal and spatial operations.

References

[1] Mahmoud Afifi, Konstantinos G Derpanis, Bjorn Ommer, and Michael S Brown. Learning multi-scale photo exposure correction. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 9157–9167, 2021. 2

[2] Jianrui Cai, Shuhang Gu, and Lei Zhang. Learning a deep single image contrast enhancer from multi-exposure images. *IEEE Transactions on Image Processing*, 27(4):2049–2062, 2018. 2

[3] Jie Huang, Yajing Liu, Xueyang Fu, Man Zhou, Yang Wang, Feng Zhao, and Zhiwei Xiong. Exposure normalization and compensation for multiple-exposure correction. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 6043–6052, 2022. 1

[4] Haoyuan Wang, Ke Xu, and Rynson WH Lau. Local color distributions prior for image enhancement. In *Proceedings of the European Conference on Computer Vision*, pages 343–359. Springer, 2022. 2

Methods	HE	CLAHE	LIME	WVM	RetinexNet	URetinexNet	DRBN	SID	MSEC	ZeroDCE
NIQE	3.7327	3.5745	3.4205	3.3705	3.8659	3.4242	3.7554	4.2326	3.4499	3.2817
Methods	Zero-DCE++	RUAS	SCI	PairLIE	ENC-SID	ENC-DRBN	CLIP-LIT	FECNet	LCDPNet	Ours
NIQE	3.3543	3.8307	3.4913	3.4725	3.2771	3.2343	3.3858	3.6465	3.2723	**3.2123**

Table 4. Quantitative results of different methods on the LCDP dataset in terms of the NIQE metric.

Methods	DRBN	SID	MSEC	ZeroDCE	Zero-DCE++	SCI	PairLIE	ENC-SID	ENC-DRBN	CLIP-LIT	FECNet	LCDPNet	Ours
Time(s)	3.7683	2.1927	1.5569	0.9783	0.0496	0.0371	5.7661	3.1936	5.6600	4.7009	4.1605	2.0874	0.7203

Table 5. The average runtime for 10 images when processing for 1024×1024 images on an Intel i7-6800K CPU.

[5] Hui Zeng, Jianrui Cai, Lida Li, Zisheng Cao, and Lei Zhang. Learning image-adaptive 3D lookup tables for high performance photo enhancement in real-time. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 44(4):2058–2073, 2022. 2

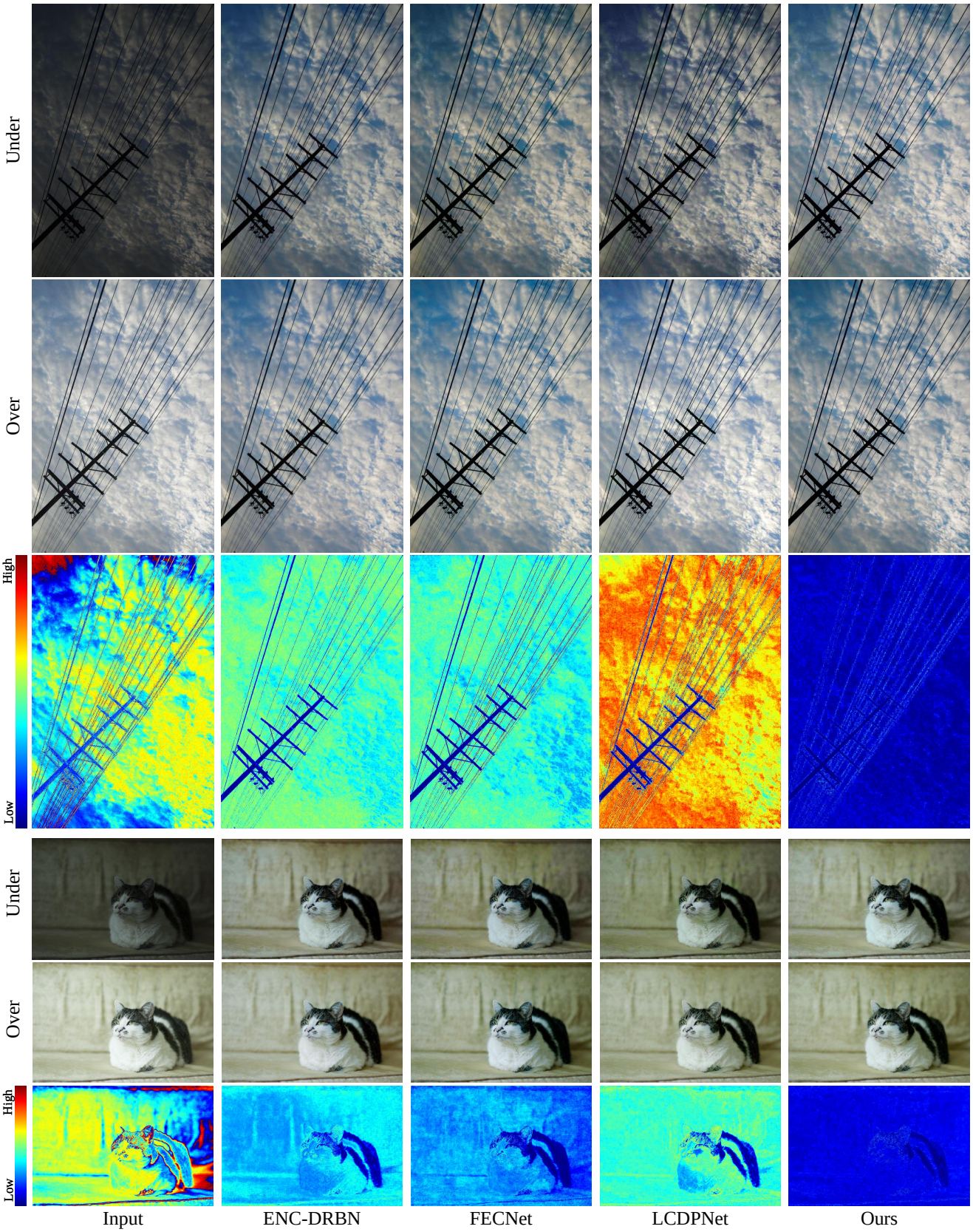


Figure 3. Exposure consistency comparison with state-of-the-art methods. The first and second rows represent the enhancement results for underexposed and overexposed images, respectively. The third row shows the error map of the two corresponding results.

Figure 4. Visual comparison with state-of-the-art methods on UHD images from the SICE dataset.

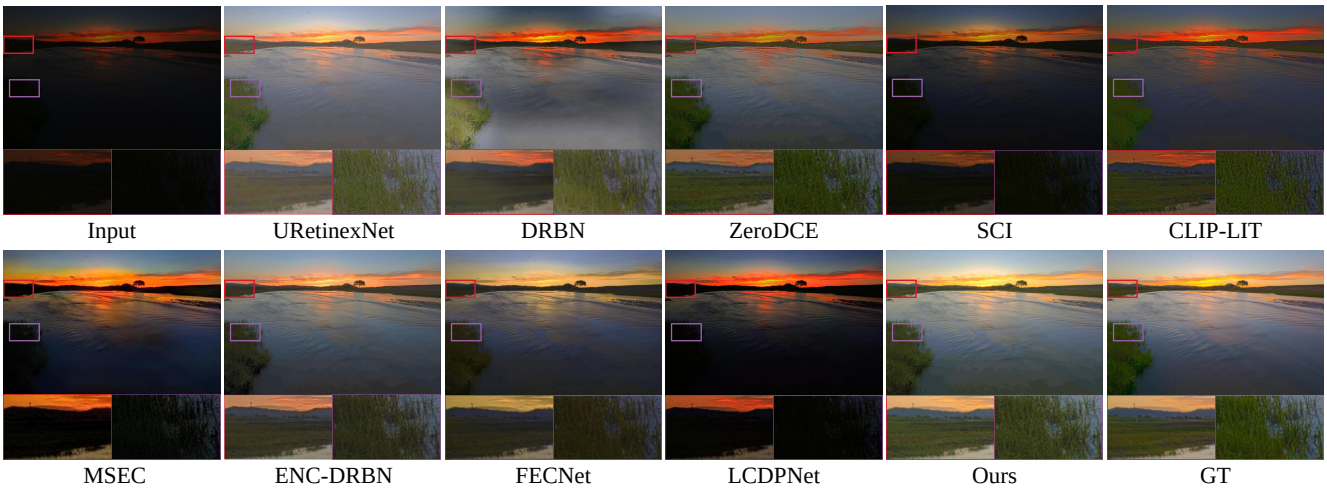


Figure 5. Visual comparison with state-of-the-art methods on underexposed images from the SICE dataset.

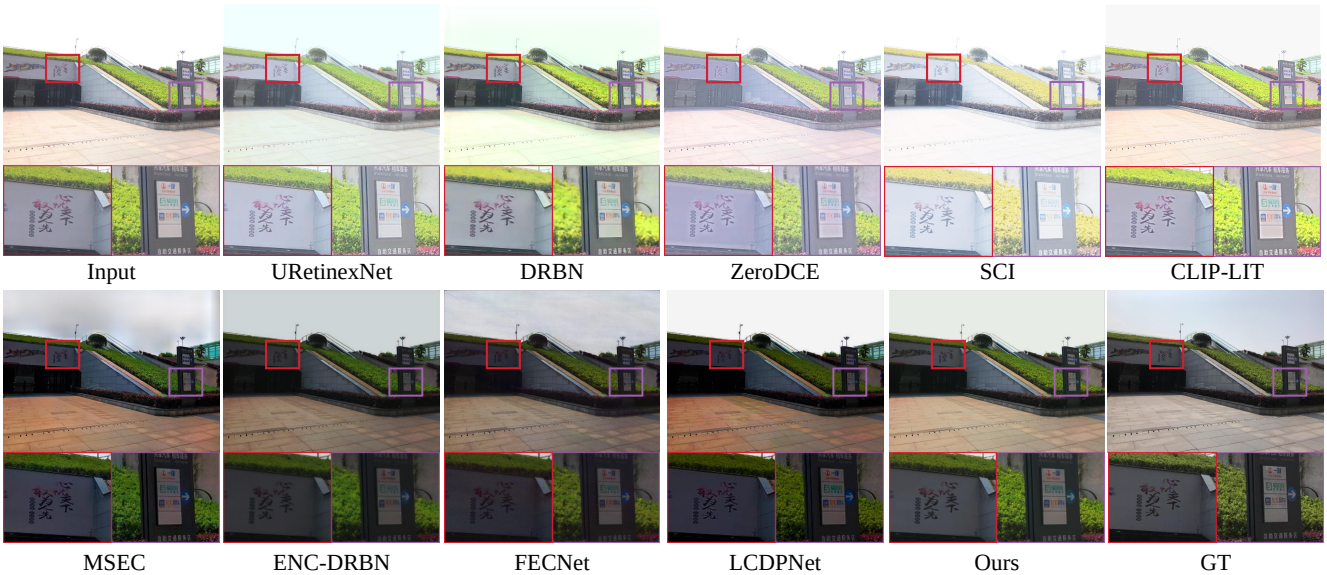


Figure 6. Visual comparison with state-of-the-art methods on overexposed images from the SICE dataset.

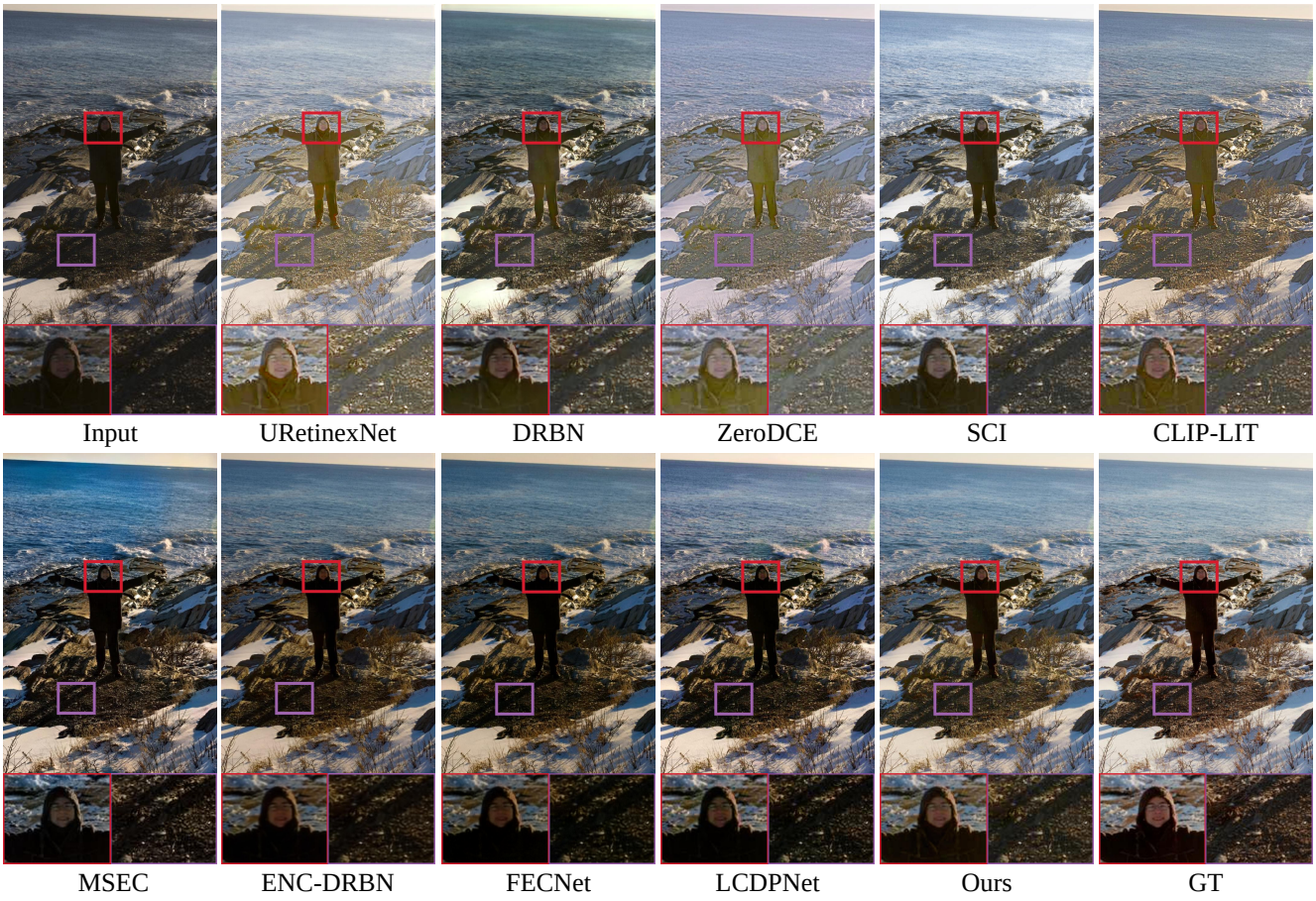


Figure 7. Visual comparison with state-of-the-art methods on underexposed images from the MSEC dataset.

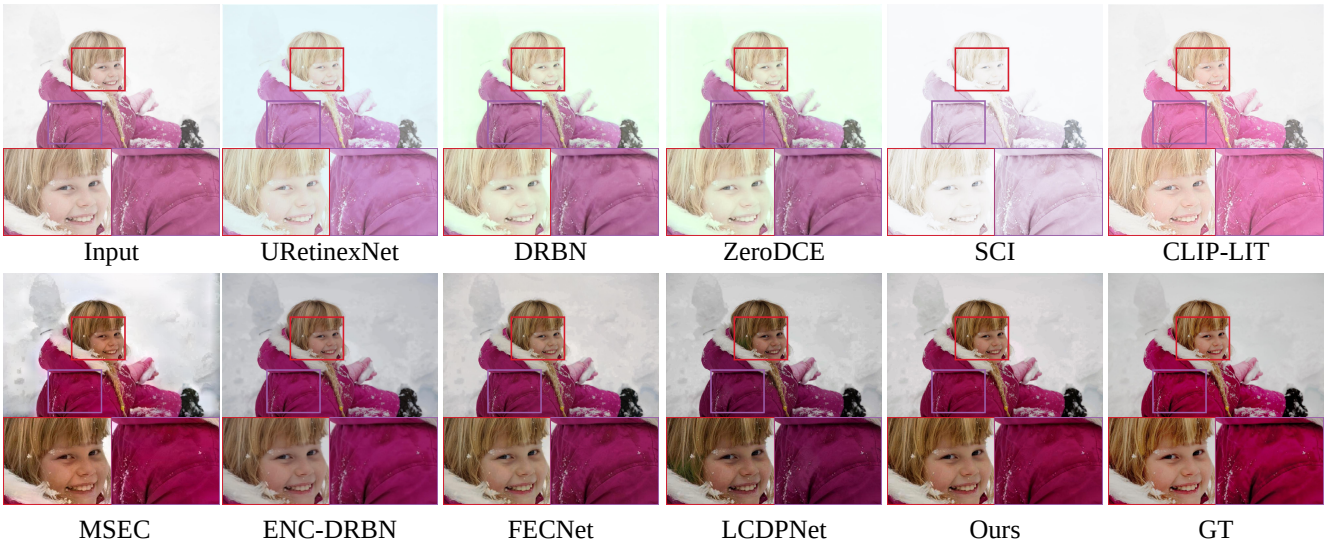


Figure 8. Visual comparison with state-of-the-art methods on overexposed images from the MSEC dataset.

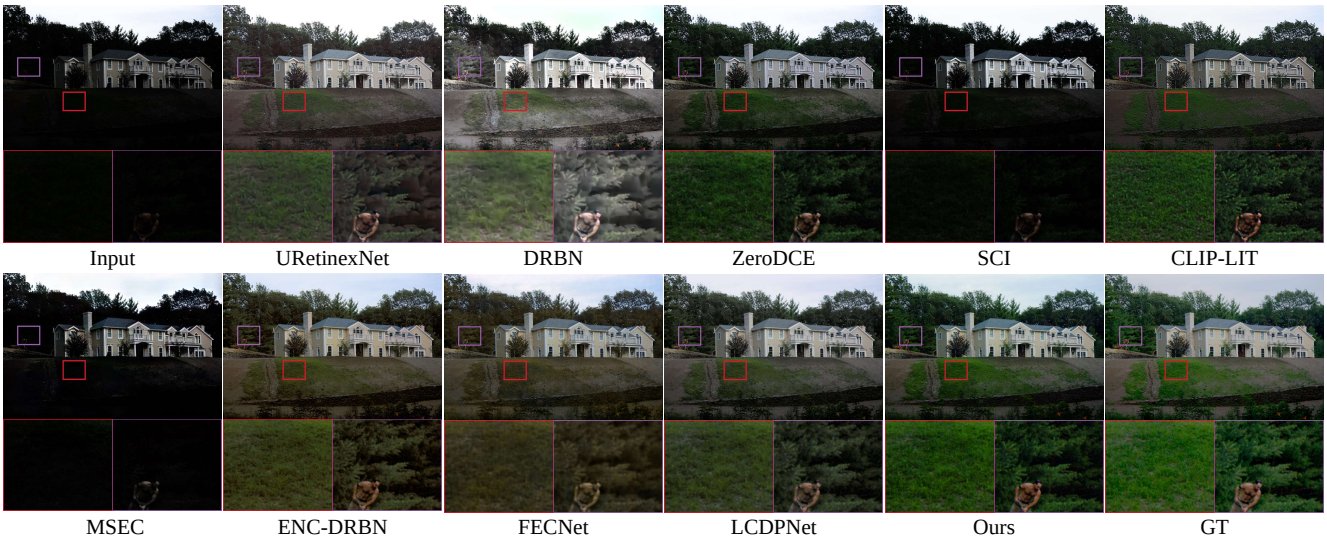


Figure 9. Visual comparison with state-of-the-art methods on underexposed images from the LCDP dataset.

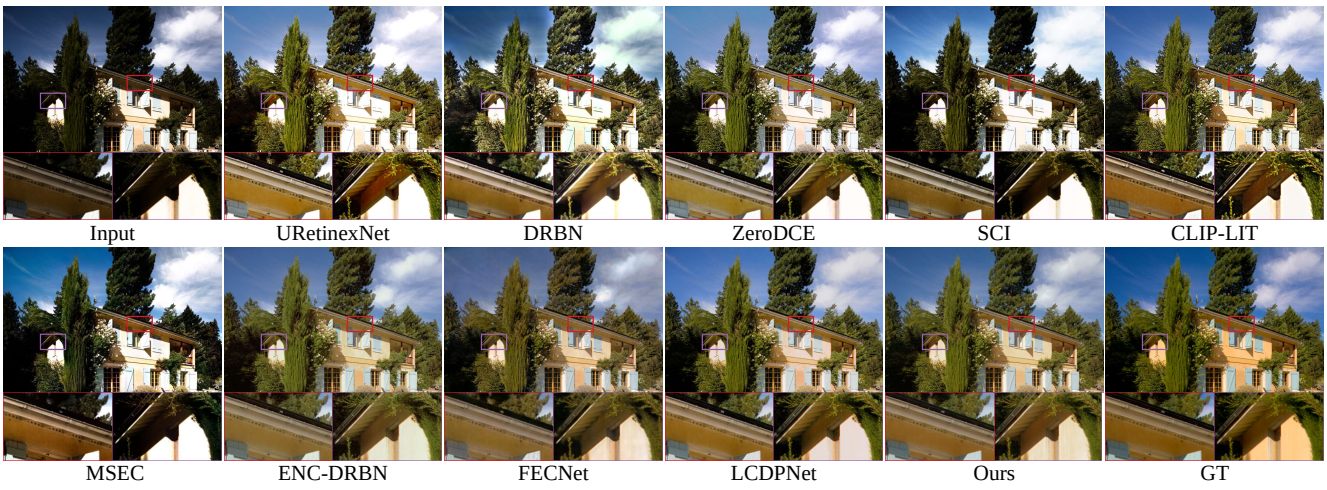


Figure 10. Visual comparison with state-of-the-art methods on under- and overexposed scenes from the LCDP dataset.