

Supplementary material for: Learning Exposure Correction Via Consistency Modeling

Ntumba Elie Nsambi
elientumba@mail.nwpu.edu.cn

Zhongyun Hu
zy_h@mail.nwpu.edu.cn

Qing Wang
qwang@nwpu.edu.cn

School of Computer Science
Northwestern Polytechnical University
Xi'an 710072, China.

1 Implementation details

Figure 1 and Table 1 show the architecture and the hyper-parameters of our network respectively. Our network is implemented as a fully convolutional encoder-decoder. Specifically, two convolutional layers are used to extract general image features. The extracted features are subsequently passed on to a series of Residual Dense Blocks (RDB) [15, 16]. The dense connections in RDBs enable each layer within the block to receive features from all previous layers. As a result, RDBs preserve the number of channels and facilitate collective features to be reused. We use two RDB layers at each level of the encoder, for a total of three levels. After each RDB, we double the number of channels and downsample the feature map by first applying a 1x1 convolutional layer, followed by a max-pooling layer.

In our network We use a total of 6 transformer [17] layers in the Global Attention Block. Each layer has an internal representation of 512 and uses 8 attention heads. The decoder is mainly composed of convolutional layers and upsampling layers. Details lost during encoding are recovered in the upsampling process via skip connections between corresponding layers. Our network parameters count is approximately 5M, which is significantly smaller than that of of Afifi *et al.* [18] 7M.

2 Quantitative Results

In Table 2 we supplement the results presented in the main paper with detailed additional results. We provide complete quantitative comparisons between our results and those of other methods on the Full test set of Afifi *et al.* [18]. The other methods include both learning and non-learning methods: Histogram Equalization (HE) [19], Contrast-Limited Adaptive Histogram Equalization (CLAHE) [20], Weighted Variational Model (WVM) [21], low light enhancement (LIME) [22], HDR-CNN [23], mobile image enhancement (DPED) [24], Deep Photo Enhancer (DPE) [25], High Quality Exposure Correction (HQEC) [26], RetinexNet [27],

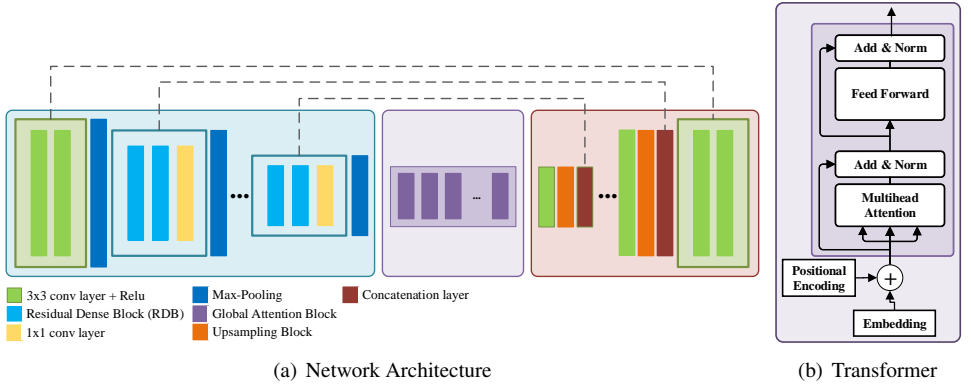


Figure 1: (a) Our network architecture composed of an encoder, Global Attention Block and Decoder. (b) Detailed structure of a Transformer layer as used in our network.

Deep Under-exposed Photo Enhancer (Deep-UPE) [14], Zero-reference Deep-curce Estimation (Zero-DCE) [15], Multi-Scale Exposure Correction (MSEC) [16]. In the main paper, we only report the average results on the 5 expert sets. Here, we report detailed results evaluated on Expert-A, Expert-B, Expert-C, Expert-D, and Expert-E. As in the main paper, we use the Peak Signal to noise ratio (PSNR) and the Structural Similarity Index Metric (SSIM) as our main evaluation metrics. Note that our method outperforms previous work at under-exposure correction, over-exposure correction and joint under-, and over-exposure correction with higher PSNR and SSIM.

3 Qualitative Results

3.1 Qualitative Comparison against other methods

We provide additional qualitative results produced by our method and compare them against those produced by other methods. We compare against previous methods on under-, and over-exposure correction.

In Figure 2, our method is compared against HDR-CNN [17], Deep Hdr Reconstruction [18] and MSEC [16]. HDR images produced by DHDR [19] and HDR-CNN [17] are tone-mapped for display using [20]. HDR-CNN [17] fails at correcting the over-exposure images. Instead, the resulting images contain color distortions and artifacts. Deep Hdr Reconstruction [18] produces wrong colors in row 4 and fails at correcting the over-exposure present in row 1. MSEC [16] produces incorrect colors in row 1 and inconsistent corrections can be observed in rows 2, 3, and 4. Our results on the other hand have consistent corrections and are closer to the well-exposed ground truth.

Additional visual comparisons on under-exposure correction are shown in Figure 3. We compare our method against DPED [21] which is a method for image enhancement and MSEC [16]. Results produced by our method are on par with these methods, with MSEC [16] producing results closer to ours.

Layer	Parameters						
	k	s	d	p	chns	input	
conv3x3 ₁	3	1	1	1	3/32	I_{in}	
conv3x3 ₂	3	1	1	1	32/32	conv ₁	
Pool ₁	2	1	-	0	32/32	conv ₂	
	k	s	d	chns	n-l	g-r	input
RBD _{1,1}	3	1	1	32/32	8	16	pool ₁
RBD _{1,2}	3	1	1	32/32	8	16	RBD _{1,1}
	k	s	d	p	chns	input	
conv1x1 ₁	1	1	1	0	32/64	RBD _{1,2}	
Pool ₂	2	1	-	0	64/64	conv1x1 ₁	
	k	s	d	chns	n-l	g-r	input
RBD _{2,1}	3	1	1	64/64	8	16	pool ₂
RBD _{2,2}	3	1	1	64/64	8	16	RBD _{2,1}
	k	s	d	p	chns	input	
conv1x1 ₂	1	1	-	0	64/128	RBD _{2,2}	
Pool ₃	2	1	-	0	128/128	conv1x1 ₂	
	k	s	d	chns	n-l	g-r	input
RBD _{3,1}	3	1	1	128/128	8	16	pool ₃
RBD _{3,2}	3	1	1	128/128	8	16	RBD _{3,1}
	k	s	d	p	chns	input	
conv1x1 ₃	1	1	-	0	128/256	RBD _{3,2}	
Pool ₄	2	1	-	0	256/256	conv1x1 ₃	
	e-s	n-l	n-h	i-r	input		
GAB	256	6	8	512	Pool ₄		
	k	s	d	p	chns	input	
conv3x3 ₃	3	1	1	1	256/128	GAB	
Up ₁	-	-	-	-	-	conv3x3 ₃	
conv1x1 ₄	1	1	-	0	512/256	Up ₁ + RDB _{3,2} + RDB _{3,1}	
conv3x3 ₄	3	1	1	1	256/128	conv1x1 ₄	
Up ₂	-	-	-	-	-	conv3x3 ₄	
conv1x1 ₅	1	1	-	0	256/128	Up ₂ + RDB _{2,2} + RDB _{2,1}	
conv3x3 ₅	3	1	1	1	128/64	conv1x1 ₅	
Up ₃	-	-	-	-	-	conv3x3 ₅	
conv1x1 ₆	1	1	-	0	128/64	Up ₃ + RDB _{1,2} + RDB _{1,1}	
conv3x3 ₆	3	1	1	1	64/32	conv1x1 ₆	
Up ₄	-	-	-	-	-	conv3x3 ₆	
conv3x3 ₇	3	1	1	1	64/32	Up ₄ + conv3x3 ₂	
conv3x3 ₈	3	1	1	1	32/3	conv3x3 ₇	

Table 1: Our network architecture, where **k** is the kernel size, **s** the stride, **d** the kernel dilation, **p** the image padding. **chns** and **input** are the number of input/output channels and the input to the layer. For RDBs and the GAB, **n-l** is the number of layers, **g-r** is the growth rate, **e-s** is the embedding size, **n-h** is the number of head and **i-r** is the internal representation size.

3.2 Correction Consistency

The method of Afifi *et al.* [4] tends to produce images that suffer from a lack of Correction Consistency. In other words, their method fails at correcting some pixels. In our network we model Correction Consistency by ensuring that distant pixels can interact with each other via self-attention, for the purpose of global image properties (e.g., color distribution, average brightness) adjustment. Consequently, pixels sharing similarities in brightness (illumination) or color, tend to be corrected in a similar manner. In Figure 4 we visualize the attention maps learned by the Global Attention block (GAB). For a given query pixel, the GAB attends to all image pixels. From Figure 4 we observe that the GAB attends more

to pixels that are similar to the query pixels in terms of brightness (illumination) or color. For instance in row 2, pixels getting the highest attention include the girls faces and the furniture behind. Although the Furniture does not have the same color as that of the faces, it nonetheless shares approximately the same brightness intensity. The same can be observed in other images.

3.3 Exposure Consistency

We provide additional qualitative results on the exposure consistency modeling of our method. For two given images sharing the same content, but different in exposure, the resulting corrected images should be as close as possible to each other, and to the well-exposed ground truth image. We compare our results against those of Afifi *et al.* [4]. As can be observed in Figure 5, our method tends to produce results with consistent exposure as opposed to that of Afifi *et al.* [4] where the produced images contain large differences as can be observed from insets images. Our better results are due to our explicit exposure consistency modeling, which encourages our network to learn exposure-invariant feature representation.

3.4 Generalization

We demonstrate our method’s ability to generalize beyond the images on which it was trained. We randomly collect under- and over-exposed images from the internet and process them using our method. In Figure 6 and Figure 7 we show results produced by our method when applied on internet images. Our method can consistently correct over-exposed and under-exposed images, recovering both colors and brightness intensity, resulting in more appealing images.

4 Limitations and Future Work

Limitations. Although the proposed network performs fairly well it is not without limitation. We identify the following cases where our network struggles or in the worst case fails at producing corrected images. (i) Extremely over-exposed images where parts or most of the image are without semantic information are challenging to our method. (ii) Underexposed images with very dark regions lacking color information pose a challenge to our method as well, as can be observed in Figure 8. In the upper row, we attempt to correct an extremely over-exposed image with many saturated regions. Results produced by our method show a level of failure in recovering missing content. The bottom row shows our attempt at correcting a very under-exposed image with very dark regions. We observe that our method tends to increase the brightness in regions with sufficient light but fails at recovering correct brightness and colors in very dark regions.

Future Work. Existing works on image exposure correction (our work included) treat the exposure correction as an image translation or regression problem, however, an over-exposed or under-exposed image can have multiple corresponding well-exposed images. As Future work, we are interested in learning a multi-modal distribution on well-exposed images. Specifically, we believe that learning a full distribution of well-exposed images and focusing only on either form of exposure error has the potential to yield improved and diverse results.

Methods	Expert-A		Expert-B		Expert-C		Expert-D		Expert-E	
	PSNR \uparrow	SSIM \uparrow	PSNR \uparrow	SSIM \uparrow	PSNR \uparrow	SSIM \uparrow	PSNR \uparrow	SSIM \uparrow	PSNR \uparrow	SSIM \uparrow
Overexposed										
HE [□] *	16.140	0.686	16.277	0.672	16.531	0.699	16.643	0.669	17.321	0.691
CLAHE [□] *	13.934	0.568	14.689	0.586	14.453	0.584	15.116	0.593	15.850	0.612
WVM [□] *	12.355	0.624	13.147	0.656	12.748	0.645	14.059	0.669	15.207	0.690
LIME [□] *	9.627	0.549	10.096	0.569	9.875	0.570	10.936	0.597	11.903	0.626
HDR-CNN w/RHT [□]	13.151	0.475	13.637	0.478	13.622	0.497	14.177	0.479	14.625	0.503
HDR-CNN w/PS [□]	14.804	0.651	15.622	0.689	15.348	0.670	16.583	0.685	18.022	0.703
DPED(iPhone) [□]	12.680	0.562	13.422	0.586	13.135	0.581	14.477	0.596	15.702	0.630
DPED(BlackBerry) [□]	15.170	0.621	16.193	0.691	15.781	0.642	17.042	0.677	18.035	0.678
DPED(Sony) [□]	16.398	0.672	17.679	0.707	17.378	0.697	17.997	0.685	18.685	0.700
DPE(HDR) [□]	14.399	0.572	15.219	0.573	15.091	0.593	15.692	0.581	16.640	0.626
DPE(S-5K) [□]	14.314	0.615	14.958	0.628	15.075	0.645	15.987	0.647	16.931	0.667
DPE(U-5K) [□]	14.786	0.638	15.519	0.649	15.625	0.668	16.586	0.664	17.661	0.684
HQEC [□] +	11.775	0.607	12.536	0.631	12.127	0.627	13.424	0.652	14.511	0.675
RetinexNet[□] +	10.149	0.570	10.880	0.586	10.471	0.595	11.498	0.613	12.295	0.635
Deep UPE[□] +	10.047	0.532	10.462	0.568	10.307	0.557	11.583	0.591	12.639	0.619
Zero-DCE[□] +	10.116	0.503	10.767	0.502	10.395	0.514	11.471	0.522	12.354	0.557
MSEC [□]	18.874	0.738	19.569	0.718	19.788	0.760	18.823	0.705	18.936	0.719
Ours	20.787	0.834	23.334	0.893	23.201	0.877	21.158	0.864	20.929	0.863
Underexposed										
HE [□] *	16.158	0.683	16.293	0.669	16.517	0.692	16.632	0.665	17.280	0.684
CLAHE [□] *	16.310	0.619	17.140	0.646	16.779	0.621	15.955	0.613	15.568	0.608
WVM [□] *	17.686	0.728	19.787	0.764	18.670	0.728	18.568	0.729	18.362	0.724
LIME [□] *	13.444	0.653	14.426	0.672	13.980	0.663	15.190	0.673	16.177	0.694
HDR-CNN w/RHT [□]	14.547	0.456	14.347	0.427	14.068	0.441	13.025	0.398	11.957	0.379
HDR-CNN w/PS [□]	17.324	0.692	18.992	0.714	18.047	0.696	18.377	0.689	19.593	0.701
DPED(iPhone) [□]	18.814	0.680	21.129	0.712	20.064	0.683	19.711	0.675	19.574	0.676
DPED(BlackBerry) [□]	19.519	0.673	22.333	0.745	20.342	0.669	19.611	0.683	18.489	0.653
DPED(Sony) [□]	18.952	0.679	20.072	0.691	18.982	0.662	17.450	0.629	15.857	0.601
DPE(HDR) [□]	17.625	0.675	18.542	0.705	18.127	0.677	16.831	0.665	15.891	0.643
DPE(S-5K) [□]	19.130	0.709	19.574	0.674	19.479	0.711	17.924	0.665	16.370	0.625
DPE(U-5K) [□]	20.153	0.738	20.973	0.697	20.915	0.738	19.050	0.688	17.510	0.648
HQEC [□] +	15.801	0.692	17.371	0.718	16.587	0.700	17.090	0.705	17.675	0.716
RetinexNet[□] +	11.676	0.607	12.711	0.611	12.132	0.621	12.720	0.618	13.233	0.637
Deep UPE[□] +	17.832	0.728	19.059	0.754	18.763	0.745	19.641	0.737	20.237	0.740
Zero-DCE[□] +	13.935	0.585	15.239	0.593	14.552	0.589	15.202	0.587	15.893	0.614
MSEC [□]	19.475	0.751	20.546	0.730	20.518	0.768	18.935	0.715	18.756	0.719
Ours	20.841	0.824	22.825	0.870	22.545	0.853	20.216	0.833	19.203	0.815
Underexposed and Overexposed										
HE [□] *	16.148	0.685	16.283	0.671	16.525	0.696	16.639	0.668	17.305	0.688
CLAHE [□] *	14.884	0.589	15.669	0.610	15.383	0.599	15.452	0.601	15.737	0.610
WVM [□] *	14.488	0.665	15.803	0.699	15.117	0.678	15.863	0.693	16.469	0.704
LIME [□] *	11.154	0.591	11.828	0.610	11.517	0.607	12.638	0.628	13.613	0.653
HDR-CNN w/RHT [□]	13.709	0.467	13.921	0.458	13.800	0.474	13.716	0.446	13.558	0.454
HDR-CNN w/PS [□]	15.812	0.667	16.970	0.699	16.428	0.681	17.301	0.687	18.650	0.702
DPED(iPhone) [□]	15.134	0.609	16.505	0.636	15.907	0.622	16.571	0.627	17.251	0.649
DPED(BlackBerry) [□]	16.910	0.642	18.649	0.713	17.606	0.653	18.070	0.679	18.217	0.668
DPED(Sony) [□]	17.419	0.675	18.636	0.701	18.020	0.683	17.554	0.660	17.778	0.663
DPE(HDR) [□]	15.690	0.614	16.548	0.626	16.305	0.626	16.147	0.615	16.341	0.633
DPE(S-5K) [□]	16.240	0.653	16.805	0.646	16.837	0.671	16.762	0.654	16.707	0.650
DPE(U-5K) [□]	16.933	0.678	17.701	0.668	17.741	0.696	17.572	0.674	17.601	0.670
HQEC [□] +	13.385	0.641	14.470	0.666	13.911	0.656	14.891	0.674	15.777	0.692
RetinexNet[□] +	10.759	0.585	11.613	0.596	11.135	0.605	11.987	0.615	12.671	0.636
Deep UPE[□] +	13.161	0.610	13.901	0.642	13.689	0.632	14.806	0.649	15.678	0.667
Zero-DCE[□] +	11.643	0.536	12.555	0.539	12.058	0.544	12.964	0.548	13.769	0.580
MSEC [□]	19.114	0.743	19.960	0.723	20.080	0.763	18.868	0.709	18.864	0.719
Ours	20.809	0.830	23.131	0.884	22.938	0.868	20.781	0.851	20.239	0.844

Table 2: Additional Quantitative comparison on the test set of [□], Expert-A, Expert-B, Expert-D, Expert-E. Methods are compared based on exposure. * denotes non learning-based methods. S and U stand for Supervised and Unsupervised. + denotes under-exposure correction methods

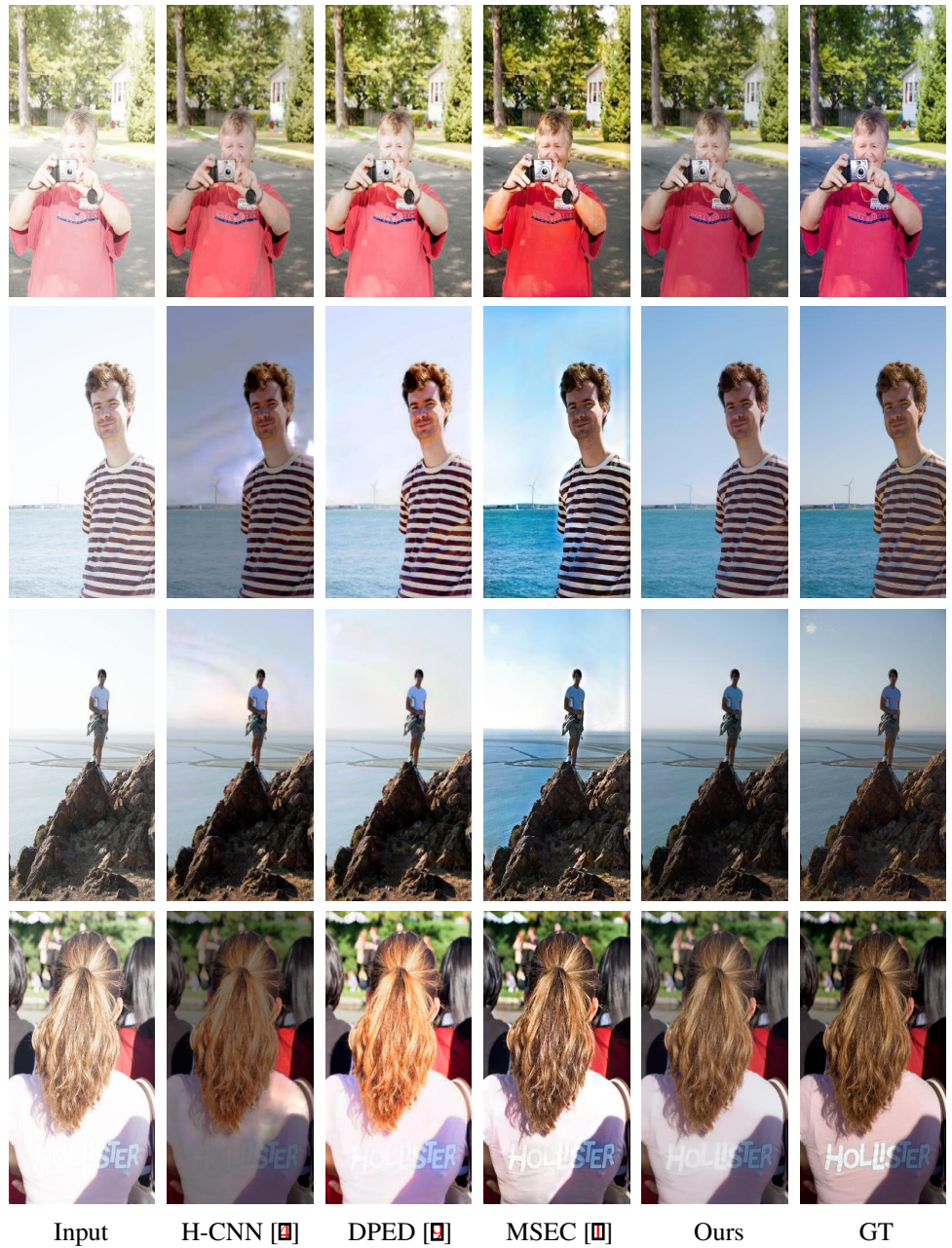


Figure 2: Additional visual comparisons on over-exposure correction Visual . Our method is compared against HDR-CNN [1], DHDR [11] and MSEC [10]. Results produced by our method have a consistent correction compared to the other methods. [10] produces results with color artifacts in row 1 and 2 and inconsistent exposure in row 3 and 4

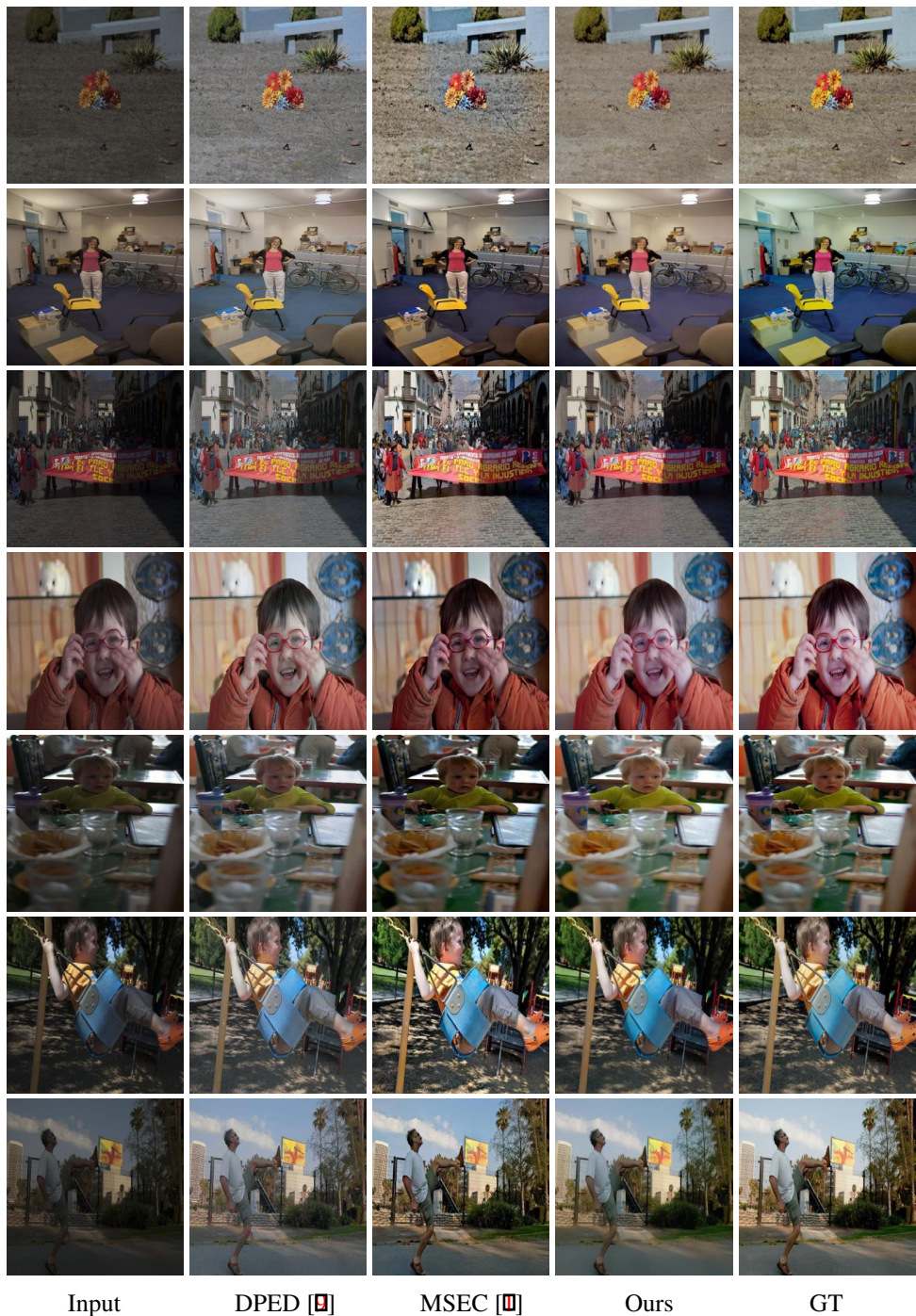
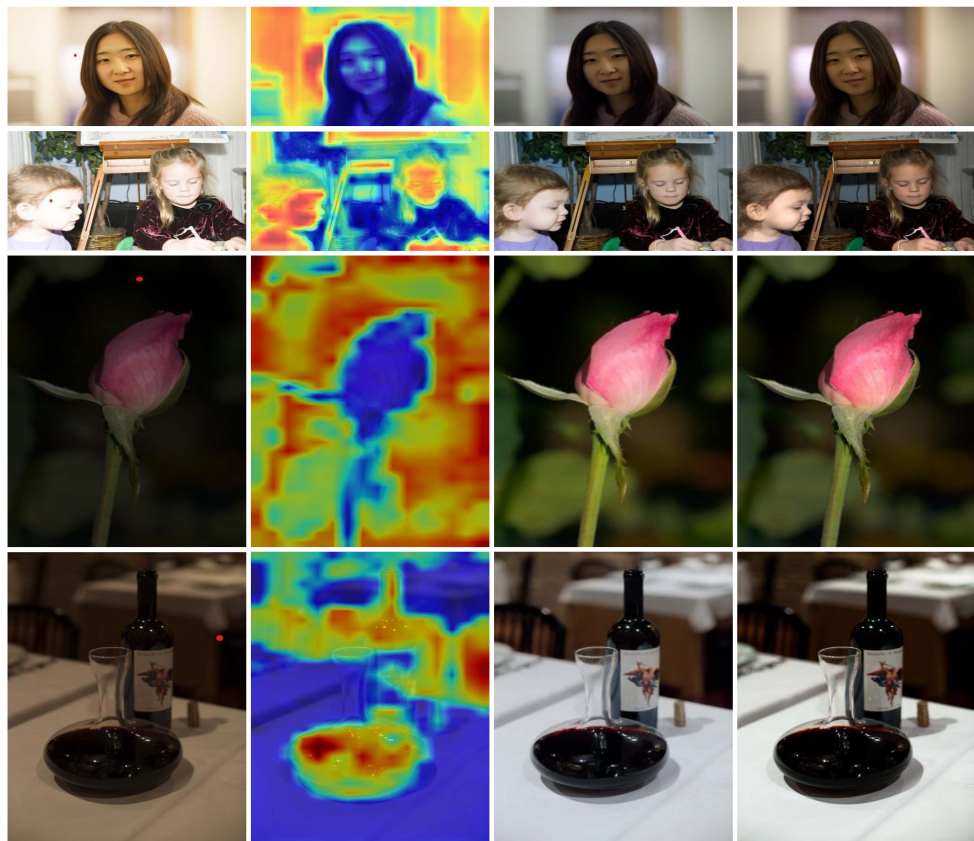


Figure 3: Additional visual comparisons on under-exposure correction. Our method is compared against DPED [9] and MSEC [10]. Results produced by our methods compete on par with these methods.



Input

Attention

Ours

GT

Figure 4: Attention Maps visualization. (a) Input image with query pixel (red dot). (b) Our global attention block attends to all the pixels in the image, and attends more to pixels that are similar to the query pixel either in terms of brightness(illumination) or color.

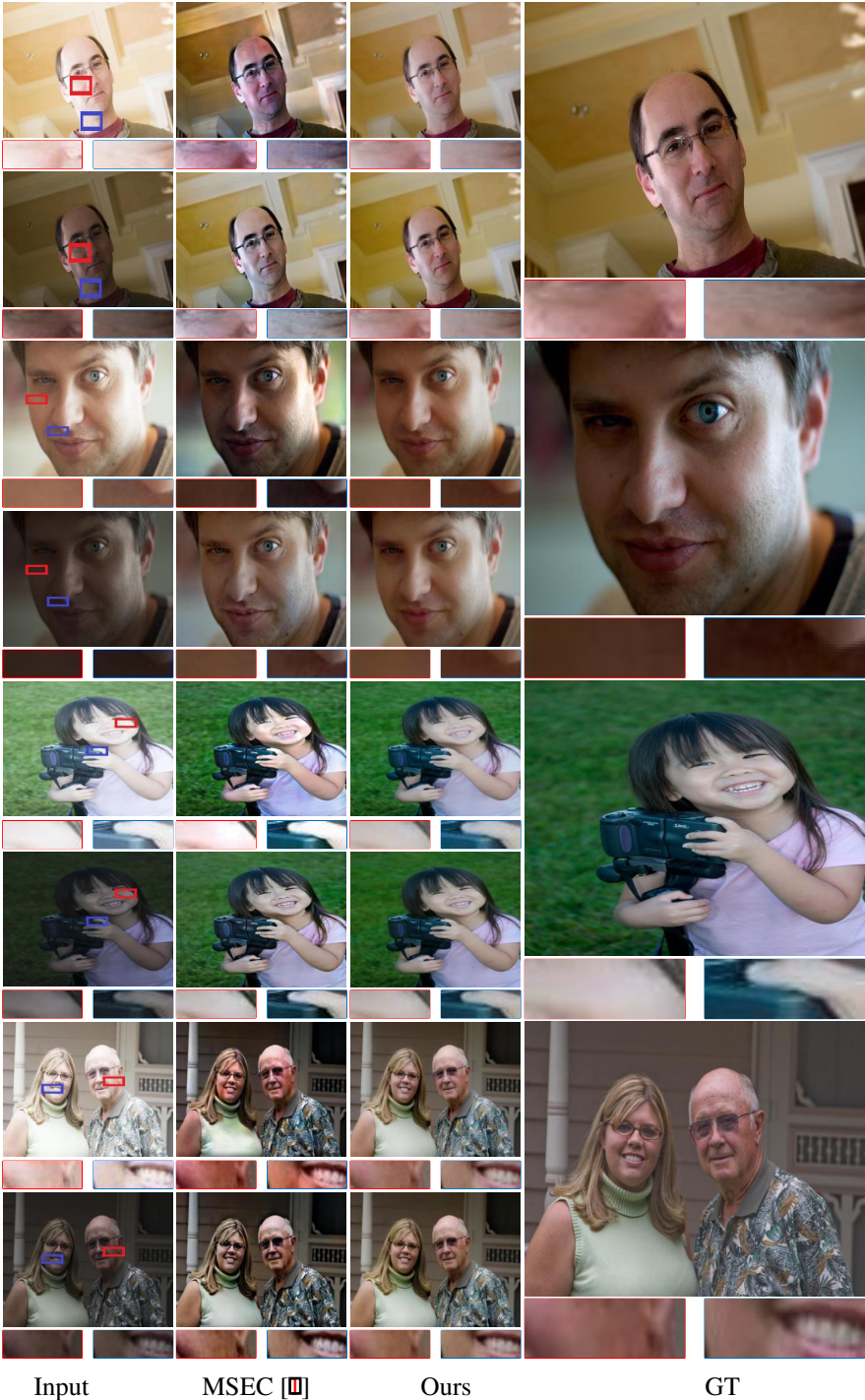


Figure 5: Additional qualitative comparison on the test set of [10] in terms of Exposure Consistency. Given two images with the same content but different exposures, our method tends to generate images with consistent exposure.

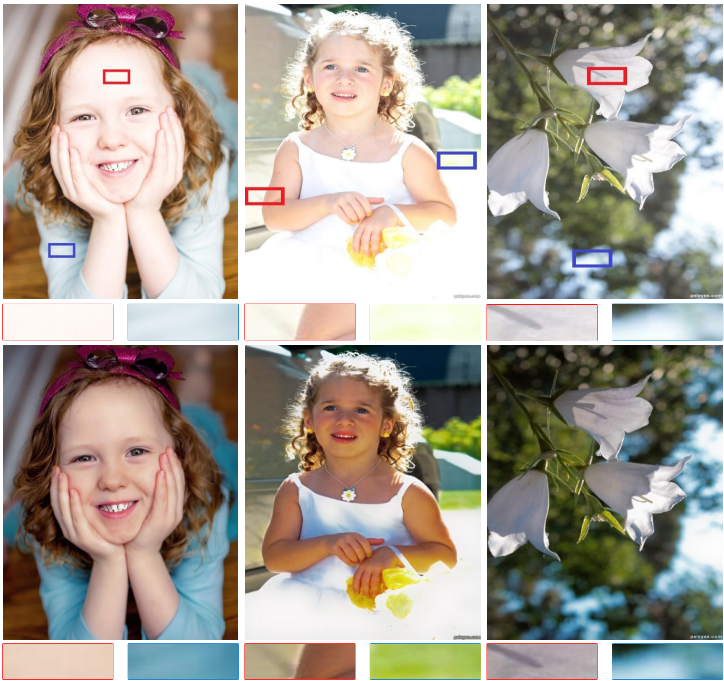


Figure 6: Results of our methods on internet images. The top row shows over-exposed images downloaded from the internet. The bottom row shows correction results by our method.

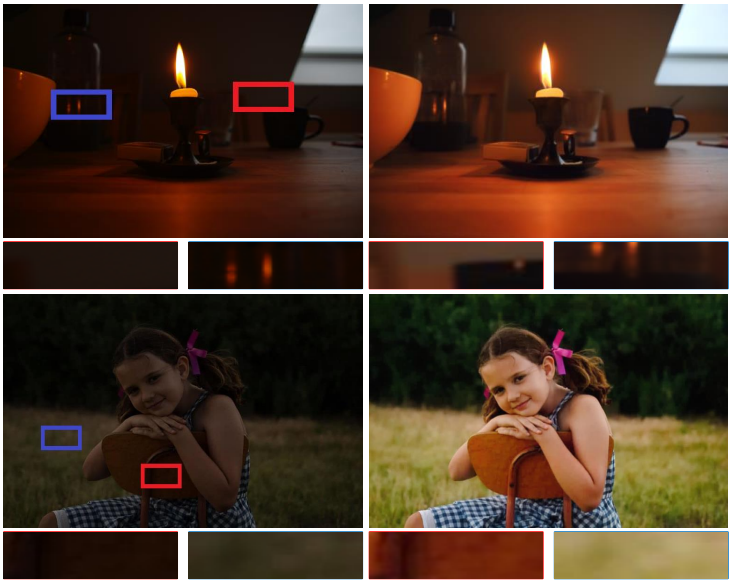


Figure 7: Results of our methods on internet images. The left column shows under-exposed images downloaded from the internet. The right column shows correction results by our method.



Figure 8: Failure cases. In row 1 we attempt to correct an extremely overexposed image. In row two we attempt to correct an underexposed image with extremely dark region.

References

- [1] Mahmoud Afifi, Konstantinos G Derpanis, Björn Ommer, and Michael S Brown. Learning multi-scale photo exposure correction. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2021.
- [2] Y. Chen, Yu-Ching Wang, Man-Hsin Kao, and Yung-Yu Chuang. Deep photo enhancer: Unpaired learning for image enhancement from photographs with gans. *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 6306–6314, 2018.
- [3] F. Drago, K. Myszkowski, T. Annen, and N. Chiba. Adaptive logarithmic mapping for displaying high contrast scenes. *Computer Graphics Forum*, 22, 2003.
- [4] G. Eilertsen, Joel Kronander, G. Denes, R. Mantiuk, and J. Unger. Hdr image reconstruction from a single exposure using deep cnns. *ACM Transactions on Graphics (TOG)*, 36:1 – 15, 2017.
- [5] Xueyang Fu, Delu Zeng, Y. Huang, X. Zhang, and Xinghao Ding. A weighted variational model for simultaneous reflectance and illumination estimation. *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 2782–2790, 2016.
- [6] R. González and R. Woods. Digital image processing. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-3:242–243, 1981.
- [7] C. Guo, Chongyi Li, J. Guo, Chen Change Loy, J. Hou, S. Kwong, and Runmin Cong. Zero-reference deep curve estimation for low-light image enhancement. *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1777–1786, 2020.
- [8] Xiaojie Guo. Lime: A method for low-light image enhancement. *Proceedings of the 24th ACM international conference on Multimedia*, 2016.
- [9] A. Ignatov, Nikolay Kobyshev, Kenneth Vanhoey, R. Timofte, and L. Gool. Dslr-quality photos on mobile devices with deep convolutional networks. *2017 IEEE International Conference on Computer Vision (ICCV)*, pages 3297–3305, 2017.
- [10] Marcel Santana Santos, Ing Ren Tsang, and N. Kalantari. Single image hdr reconstruction using a cnn with masked features and perceptual loss. *ACM Transactions on Graphics (TOG)*, 39:80:1 – 80:10, 2020.
- [11] Ashish Vaswani, Noam M. Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, and Illia Polosukhin. Attention is all you need. *ArXiv*, abs/1706.03762, 2017.
- [12] R. Wang, Q. Zhang, Chi-Wing Fu, Xiaoyong Shen, W. Zheng, and J. Jia. Underexposed photo enhancement using deep illumination estimation. *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 6842–6850, 2019.
- [13] Chen Wei, W. Wang, W. Yang, and Jiaying Liu. Deep retinex decomposition for low-light enhancement. In *BMVC*, 2018.

- [14] Qing Zhang, Ganzhao Yuan, Chunxia Xiao, L. Zhu, and W. Zheng. High-quality exposure correction of underexposed photos. *Proceedings of the 26th ACM international conference on Multimedia*, 2018.
- [15] Yulun Zhang, Yapeng Tian, Yu Kong, Bineng Zhong, and Yun Fu. Residual dense network for image super-resolution. *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 2472–2481, 2018.
- [16] Yulun Zhang, Yapeng Tian, Yu Kong, Bineng Zhong, and Yun Fu. Residual dense network for image restoration. *IEEE transactions on pattern analysis and machine intelligence*, 2020.
- [17] K. Zuiderveld. Contrast limited adaptive histogram equalization. In *Graphics Gems*, 1994.