https://doi.org/10.2352/issn.2694-118X.2021.LIM-43
This work is licensed under the Creative Commons
Attribution 4.0 International License.

Content Fidelity of Deep Learning Methods for Clipping and Over-exposure Correction

Mekides Assefa Abebe; NTNU; Gjøvik, Norway

Abstract

Exposure problems, due to standard camera sensor limitations, often lead to image quality degradations such as loss of details and change in color appearance. The quality degradations further hiders the performances of imaging and computer vision applications. Therefore, the reconstruction and enhancement of uder- and over-exposed images is essential for various applications. Accordingly, an increasing number of conventional and deep learning reconstruction approaches have been introduced in recent years. Most conventional methods follow color imaging pipeline, which strongly emphasize on the reconstructed color and content accuracy. The deep learning (DL) approaches have conversely shown stronger capability on recovering lost details. However, the design of most DL architectures and objective functions don't take color fidelity into consideration and, hence, the analysis of existing DL methods with respect to color and content fidelity will be pertinent. Accordingly, this work presents performance evaluation and results of recent DL based over-exposure reconstruction solutions. For the evaluation, various datasets from related research domains were merged and two generative adversarial networks (GAN) based models were additionally adopted for tone mapping application scenario. Overall results show various limitations, mainly for severely over-exposed contents, and a promising potential for DL approaches, GAN, to reconstruct details and appearance.

Introduction

Standard dynamic range (SDR) color images and videos mostly deteriorate from under- and over-exposure problems. Currently, exposure problems arise from camera sensor limitations, unprofessional acquisition setups, post processing for display referred content formats, and related other factors [1, 2, 3]. These situations usually lead to quality degradation such as loss of details, color clipping, saturation, and hue changes, as well as visibility/contrast reduction. Appearance changes due to such quality degradations further disrupt the artistic intent and put additional hurdle on other applications such as computer vision and robotics. Hence, for accurate appearance reproduction as well as scene content analysis, quality enhancement of ill-exposed contents is highly essential [1, 4].

The enhancement of appearance degradation, due to one or two channels camera sensor clipping or application of tone mapping, have been addressed with state-of-the-art techniques such as propagation of data from well-exposed image regions and/or color channels. Statistical and probabilistic corrections of pixel values in different perceptual color spaces were also utilized. In some of the proposed enhancement techniques, for example, luminance or chrominance gradients as well as *RGB* channel spatial variation patters were propagated from the well-exposed parts of the degraded images in a gradient domain and *sRGB* spaces. The methods mostly rely on the assumption of strong spatial and color channel correlations of natural image contents [5, 6, 7]. Other related methods utilized perceptual color spaces to adjust degraded perceptual attributes (such as chroma and hue) for color clipped pixels in ill-exposed image regions [8, 9].

In case of over-exposed contents, where information is lost in all color channels, image reconstruction techniques such as histogram equalization, tone adjustments, inpainting based content hallucination, brightness enhancement using image pyramids, and utilization of *RGB* image channels' spatial correlation were applied [10, 11, 12, 13, 1]. Such techniques also are designed to fill lost details in the over-exposed image regions with information extracted from correlated and well-exposed image regions, and to adjust the intensity distributions of image pixels for better visibility and contrast. However, the conventional computer vision techniques generally showed limited capabilities to recover severely over-exposed contents [1]. Recreating lost details from a single image and with no other well-exposed and correlated image regions requires more advanced machine learning and optimization approaches.

The recent advances of DL based approaches brought more capacity and hope in this regard. Accordingly, few DL based under- and over-exposure enhancement methods were introduced recently. The methods mostly utilized convolutional neural network (CNN) based autoencoder architectures and Laplacian pyramid decompositions techniques [14, 15, 16, 17, 18]. Even if the content reconstruction ability of DL methods (given a proper data set) is robust, the color and content fidelity of the results is not primarily guaranteed. However, for applications like color image reproduction and high dynamic range (HDR) imaging, the reproduction quality of scene appearance and artistic intent is highly relevant, and the results of the DL methods need to be investigated accordingly.

Consequently, a concise evaluation of the few openly available DL approaches for over-exposure correction is presented in this work. The cycleGAN and Pix2Pix implementation of Jun-Yan Zhu et. al. [19, 20] DL models are also adopted, trained, and evaluated for this scenario, with a newly constructed over-exposure data set. The analysis results indicate the improved performances of generative models compared to the ones with the simple encoder-decoder CNN architectures. Conversely, considering color fidelity, higher inaccuracy was observed for all the assessed models in the extreme cases of scene dynamic ranges and severely over-exposed images. The inaccuracy is mainly resulted from the nature of the network architecture, objective functions and data set related issues. The details of the DL models, the evaluation methodology as well as the evaluation results of this work are provided in the following sections.

Deep Learning Based Over-exposure Enhancement

Considering the enhancement of ill-exposed images applications, most proposed DL methods are intended for under-exposed images [17, 18, 16]. However, in this study, only over-exposed contents are considered, and only limited number of DL over-exposure reconstruction solutions were available at the time of writing. Mahmoud Afifi et. al. [15] recently proposed an over-exposure correction method based on a concatenated encoder-decoder UNet networks [21], processing different levels of the Laplacian pyramid outputs of the input and reference images. Other earlier methods again utilized auto-encoder and GAN based approaches [15, 22, 14]. The proposed models are mainly trained on paired data sets generated from raw-RGB (such as MIT-Adobe FiveK data set [23, 15]) and multiple exposure based images [24, 17, 4]. This work evaluates the two CNN models of Steffens et al. [22, 14] and my adaptations of two GAN architectures [20, 19] for over-exposure enhancement, for their content and color fidelity performances. Brief descriptions of the models are given as follows.

ReExposeNet and UCan

Steffens et. al. proposed two DL models for over-exposure correction with contrast enhancement [14] and luminance and color correction [22]. Both models are designed similarly with a memory optimized encoder-decoder type CNN architecture. For memory optimization, they have used dilated convolution layers with fewer dilation rate and Exponential Linear Unit (ELU) activation layers. To prevent loss of details and efficiency purposes, strided convolution based spatial down- and up-sampling process was applied rather than the common polling operations. The authors also added the Structural Dissimilarity (DSSIM) metric to the loss function (as shown in Eq. 1), which is mostly Pixel-wise Euclidean Distance ($L2$) in other DL applications, for the network to learn more high-level structural differences. The $L2$ distance is also weighted according to the closeness of the image intensity to the saturation point (b) of the camera sensor.

$$L(x,y) = \lambda\,|0.5-b|\,L_2(x,y) + (1-\lambda)DSSIM(x,y) \tag{1}$$

x, y, λ respectively represent the input image, ground truth image, and a user defined parameter.

GANs: Generative adversarial networks

The application of the simple encoder-decoder CNN architectures and optimization processes with minimization of simple and fixed pixel wise differences, usually result in less generalizable models. The processed results of such models will also be mostly blurred. GANs, however, can provide more robust and generalizable adversarial models by utilizing a CNN discriminator network as a learning loss function [25]. Currently, there are many variants of GANs but the basic framework of all mostly consists of two CNN networks: a Generator $G(x,z)$ for generating predictions (z) from the input latent distribution (x) and a Discriminator $D(y)$ to penalize and assess the predictions according to the target ground truth (y). As such, most GANs do not require paired data set training, providing semi-supervised learning opportunities. The training of such frameworks is mostly performed by a min-max optimization formulation of their objective function, which the generator needs to minimize while the discriminator is maximizing it [25]. Since there are not any openly available GANs proposed for over-exposed image enhancement applications, by the time of writing, I have adapted and trained two models (Pix2Pix and CycleGAN), which are originally proposed for image-to-image translation applications [20, 19].

Pix2Pix

This model is a conditional version of GANs proposed by Isola at. al. for image-to-image translation purposes [20]. The model is designed to work with paired images data set with encoder-decoder generator network. Different variations of this network are implemented by Jun-Yan Zhu et. al. [19, 20], by changing the generator architecture and loss functions. For this work, the generator networks with a 256 layers Unet architecture (referred as Pix2Pix1 in the evaluation section and Tab. 1) and with a 128 layers Unet architecture (for Pix2Pix2 and Pix2Pix3 variants) are evaluated [21]. Similarly, two different conditional loss functions used in the vanilla GAN and lsGAN architectures are also evaluated with the Pix2Pix2 and Pix2Pix3 variants, respectively.

The vanilla GAN conditional loss function ($L_{cGAN}(G,D)$), given in Eq. 2, contains a discriminator loss $D(x,y)$ as a function of both the input and ground truth image. The overall objective function, G^*_{L1} in Eq. 3, then combines the $L1$ norm of the difference between the generated and ground truth image, Eq. 4, as a min-max optimization. On the other hand, to avoid the vanishing gradient problem of such loss functions, lsGAN introduced a least square formulation of the discriminator, Eq. 5, and generator, Eq. 6, loss functions separately. a, b are created labels for fake and real samples, whereas c represents a sample for which the generator wants the discriminator to identify as fake.

$$L_{cGAN}(G,D) = E_{x,y}[\log D(x,y)] + E_{x,y}[\log(1-D(x,G(x,z)))] \tag{2}$$

$$G^*_{L1} = \arg\min_G \max_D L_{cGAN}(G(D)) + \lambda L_{L1}(G) \tag{3}$$

$$L_{L1}(G) = E_{x,y,z}[\|y - G(x,z)\|_1] \tag{4}$$

$$\begin{aligned}\arg\min_D L_{lsGAN}(D) =& 0.5E_{x,y}[(D(x,y)-b)^2] \\ &+ 0.5E_{x,y}[(D(x,G(x,z))-a)^2]\end{aligned} \tag{5}$$

$$\arg\min_G L_{lsGAN}(G) = 0.5E_{x,y}[(D(x,G(x,z))-c)^2] \tag{6}$$

CycleGAN

In cases where there is a lack of paired data sets, the original GAN architecture with semi-supervised learning is more applicable [25]. Accordingly, CycleGAN is proposed for translating information among unpaired image distributions [19]. CycleGAN consists of two consecutive GANs (with similar architecture and objective function as the original GAN [25]) to translate the source to destination and the destination back to the source images. CycleGAN additionally considers the similarity between the source images and the reconstructed images, all the way back through the backward translation network, with an additional cycle consistency loss. This loss helps to insure the preservation of important source image attributes. In this work, we have adapted and evaluated the CycleGAN implementation of Jun-Yan Zhu et.

al. [19, 20] with ResNet generator architectures. For detail understanding of the CycleGAN architecture and the corresponding loss functions, please refer to the original publication [19].

Data set and Training Procedure

As mentioned in the beginning of this section, the ReExposeNet [14] and UCan [22] models are mainly designed for over-exposure enhancement and they are trained on a large data set containing the MIT-Adobe FiveK data set [23, 15]. Therefore, their original pre-trained models, provided by the authors, are used to generate their corresponding results. On the other hand, the Pix2Pix[20] and CycleGAN [19] models have not been tested for over-exposure correction before and they need to be trained on appropriate data set. In this regard, both networks were trained on two sorts of data sets. Firstly, a small scaled but strongly representative data set (containing only severely over-exposed images) was generated from 288 HDR images taken from the RIT Photographic Survey database of HDR photographs [26] and 114 SDR over-exposed image sets generated by Steffens et. al. [14]. The HDR images from the RIT database were in linear OpenEXR format and accompanied by detailed luminance measurements and visual appearance scaling from the original scenes. Hence, due to the required visibility in SDR rendering and compatibility to the Steffens et. al. data sets, the RIT HDR images were slightly tone mapped with $\gamma = 1.5$, as shown in Fig. 1). To create their corresponding over-exposed images, however, Banterle et. al.'s exposure bracketing technique is utilized [3], as described in Abebe et. al. [1]. Since paired images are not required for training the CycleGAN, 402 more over-exposed images were added with no corresponding well-exposed references. The results of the networks trained with this data set are shown as CycleGAN1 and Pix2Pix1 in the evaluation results, Tab. 1.

Considering the complexity of both Pix2Pix and CycleGAN models, training with a total of 402 and 804 images may be is inadequate and leads to generalization problems. As the result, a second data set is constructed by combining images from the data set provided by Mahmoud Afifi et. al. [15] to the small data set. Their data set is also generated from MIT-Adobe FiveK raw-RGB images [23] and their sRGB images were generated at five different exposure values, by expert photographers. The final combined data set contains 18000 unpaired (for CycleGAN) and 4569 paired (for Pix2Pix) over-exposed images. The evaluated models which are trained on this data set are CycleGAN2, Pix2Pix2, and Pix2Pix3.

All the models were trained on Twin Titan computer (with 2 Nvidia GPUs) for a total of 500 (CycleGAN1), 50 (CycleGAN2), 1000 (Pix2Pix1), 200 (Pix2Pix2, Pix2Pix3) epochs. In all the trainings, the recommended hyper-parameters of the original implementations were followed.

Evaluation and Analysis of Results

As per the main goal of this work, all the trained models (listed in Tab. 1) were assessed for their color and content fidelity. Around 25 over- and well-exposed image pairs are generated for the evaluation, based on the same HDR exposure bracketing technique described in the previous section. Fig. 1b and Fig. 1c also show an example pair, which is created similarly as the evaluation pairs. Only 6 of the evaluation images were taken from Steffens et. al. [14] data set, for fair comparison, and note that none of the 25 images were used for training all the models. The over-exposed images of the 25 pairs are then processed by all the evaluated models for enhancement and content recovery. The evaluation of content fidelity is finally performed by comparing the enhanced images of each model with the well-exposed reference images of the evaluation pairs.

Quality comparisons were done both visually and objectively. To consider both color and structural fidelity, objective quality metrics such as structural similarity index metric (SSIM), improved color image difference (iCID), CIE2000 color difference (CIEDE2000) and visible difference prediction metric for HDR images (HDR-VDP3) are utilized [27, 28].

(a) Reference HDR (b) Input LDR (c) Slightly tone mapped HDR

Figure 1: Example training pare generated from the RIT HDR images. Fig. 1a is the HDR image linearly displayed. Fig. 1b shows the over-exposed version generated with exposure bracketing and Fig. 1c is the slightly tone mapped (with $\gamma = 1.5$) well-exposed version of the training pair.

In the overall visual inspection of results, most of the evaluated models observed good color correction and content recovering capabilities for slightly clipped image regions. For example, the color and details of the slightly clipped sofa, pillow and bunny regions of Fig. 2 (first column) were able to be recovered well almost by all evaluated methods. However, since the focus of this work is mainly on severely over-exposed images, all the 25 evaluation images contain severe over-exposures. In the presence of severely distorted regions, like around the wall and the sky regions of the first and second columns of Fig. 2, the color fidelity performances of most of the models significantly reduced. The models tend to fill some over-exposed image regions with semantically incoherent contents (as can be seen in Fig. 2) or they do not show any improvement at all. Such problems in generative as well as autoencoder based DL methods are very common, mainly with insufficient training data and time. The causes of such infidelity can include the evaluated models' network architecture and loss functions. Both the GAN and CNN based models are constructed using simple CNN components and doesn't consider semantic pixel to pixel (in an image) as well as cross-class (among images) relationships. The various loss functions tested with the evaluated models also not consider color reproduction or appearance quality. It is currently not common to use state of the art image quality assessment metrics as an objective function due to their complexity, convexity, as well as differentiability issues.

For the objective evaluation, the average quality values of the evaluation images set for all the described image quality metrics are computed and given in Tab. 1. The results of all the variants of the evaluated models, described in previous sections, as well as the input over-exposed images are presented. As shown from the results, the two models which are trained on the small-scale data

Table 1: Average quality evaluations of the enhancement results, computed from 25 test images.

	SSIM		iCID		DE2000		VDP-Q	
	mean	δ	**mean**	δ	**mean**	δ	**mean**	δ
Input	0.683661	0.234756	0.339291	0.274037	13.30943	13.15839	6.607415	1.420977
CycleGAN1	**0.714409**	0.180895	**0.325367**	0.203527	**9.771153**	6.824443	**7.094045**	1.151893
CycleGAN2	0.628975	0.20626	0.37248	0.228797	12.12396	9.645018	6.696073	1.272312
Pix2Pix1	**0.799783**	0.13298	**0.242246**	0.185161	**5.956528**	3.658938	**7.161347**	1.096364
Pix2Pix2	0.653193	0.172157	0.381459	0.211426	10.95747	6.703688	6.739119	1.256269
Pix2Pix3	0.643285	0.154942	0.409377	0.178226	11.86516	5.508725	6.204647	1.106786
ReExposeNet	0.608942	0.164575	0.333918	0.197487	10.38835	4.40941	5.78962	1.188603
Ucan	0.525642	0.207551	0.34854	0.21599	11.7023	6.653122	5.857483	1.274583

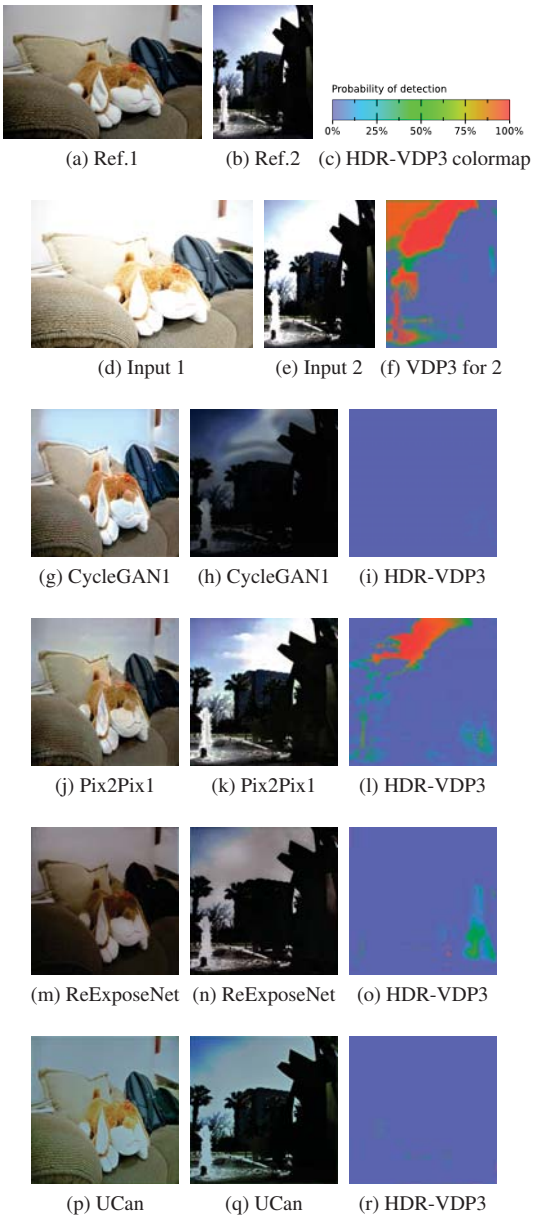
set (CycleGAN1 and Pix2Pix1) resulted in better quality values in terms of both structural and color recovery. The two simple CNN based models (ReExposeNet and Ucan), on the other hand, performed worse in almost all metrics.

The better performance of Pix2Pix1 model is understandable given that it is trained with supervised learning and teachable loss function as discriminator network (compared to the fixed DSSIM and $L2$ norm based loss functions of the ReExposeNet and UCan models). Considering the performances of the CycleGAN models, cycle consistency loss objective function is intended to translation of image attributes with in structurally similar images, with a data set of similar structural distributions like the apple2orange and horse2zebra data sets, proposed by the respective authors [19]. Therefore, the diverse structural distribution of the proposed data sets together with the cycle consistency loss may further hinder and constrain the generator from generating more realistic (well-exposed) results.

The data set improvement made for creating the second data set also seems not to influence and improve the results of the GAN based models (as it can be inferred from the results of CycleGAN2, Pix2Pix2, and Pix2Pix3 in Tab. 1). This issue may be is the result of insufficient training time and less representations of severely over-exposed scenes in the data set. As it is mentioned in the training procedures section, these models were trained on a much larger data set but with much lower number of epochs, due to resource and time limitations (at the time of writing). Therefore, there is a good potential that these models will result in higher performances with longer training time. It is also a good idea to further increase the size of the data set with more representative images for severe over-exposure scenarios.

Conclusion and Future Work

In this work, various DL based over-exposure correction techniques were assessed for their content and color fidelity. The assessment was mainly focused on the color and structural content recovery of severely over-exposed scenarios. The results of the assessment showed better content and color fidelity performances of GAN based solutions. However, for most severely degraded contents, the methods fail to give significant appearance improvements. Advancements on the network architectures and objective functions of the models, to consider the semantic relations of image pixels [29] as well as the representation consistency of images with the same semantic classes [30], can lead to better content fidelity. Creating a more appropriate data set, generated from absolute HDR data with proper inspection on HDR display devices, can also be considered as a next research direction.



(a) Ref.1 (b) Ref.2 (c) HDR-VDP3 colormap

(d) Input 1 (e) Input 2 (f) VDP3 for 2

(g) CycleGAN1 (h) CycleGAN1 (i) HDR-VDP3

(j) Pix2Pix1 (k) Pix2Pix1 (l) HDR-VDP3

(m) ReExposeNet (n) ReExposeNet (o) HDR-VDP3

(p) UCan (q) UCan (r) HDR-VDP3

Figure 2: Sample results shown for two of the evaluation images, representing severe over-exposure. The third column shows the HDR-VDP difference probability of detection maps, for the second column results.

References

[1] M. A. Abebe, A. Booth, J. Kervec, T. Pouli, and M.-C. Larabi, "Towards an automatic correction of over-exposure in photographs: Application to tone-mapping," *Computer Vision and Image Understanding*, vol. 168, pp. 3–20, 2018. Special Issue on Vision and Computational Photography and Graphics.

[2] E. Reinhard, G. Ward, S. Pattanaik, and P. Debevec, *High Dynamic Range Imaging: Acquisition, Display, and Image-Based Lighting (The Morgan Kaufmann Series in Computer Graphics)*. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., 2005.

[3] F. Banterle, A. Artusi, K. Debattista, and A. Chalmers, *Advanced High Dynamic Range Imaging (2nd Edition)*. Natick, MA, USA: AK Peters (CRC Press), July 2017.

[4] C. Steffens, P. L. J. Drews, and S. S. Botelho, "Deep learning based exposure correction for image exposure correction with application in computer vision for robotics," in *2018 Latin American Robotic Symposium, 2018 Brazilian Symposium on Robotics (SBR) and 2018 Workshop on Robotics in Education (WRE)*, pp. 194–200, IEEE, 2018.

[5] M. Rouf, C. Lau, and W. Heidrich, "Gradient domain color restoration of clipped highlights," in *2012 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops*, pp. 7–14, IEEE, 2012.

[6] S. Z. Masood, J. Zhu, and M. F. Tappen, "Automatic correction of saturated regions in photographs using cross-channel correlation," in *Computer Graphics Forum*, vol. 28, pp. 1861–1869, Wiley Online Library, 2009.

[7] M. Assefa, T. Poulie, J. Kervec, and M.-C. Larabi, "Correction of over-exposure using color channel correlations," in *2014 IEEE Global Conference on Signal and Information Processing (GlobalSIP)*, pp. 1078–1082, 2014.

[8] E. Elboher and M. Werman, "Recovering color and details of clipped image regions," *Proc. CGVCVIP*, pp. 1–8, 2010.

[9] D. Guo, Y. Cheng, S. Zhuo, and T. Sim, "Correcting over-exposure in photographs," in *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp. 515–521, 2010.

[10] C. Lee, C. Lee, and C.-S. Kim, "Contrast enhancement based on layered difference representation of 2d histograms," *IEEE Transactions on Image Processing*, vol. 22, no. 12, pp. 5372–5384, 2013.

[11] S. Moran, P. Marza, S. McDonagh, S. Parisot, and G. Slabaugh, "Deeplpf: Deep local parametric filters for image enhancement," in *IEEE CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2020.

[12] J. Park, J.-Y. Lee, D. Yoo, and I. S. Kweon, "Distort-and-recover: Color enhancement using deep reinforcement learning," in *IEEE CVF Conference on Computer Vision and Pattern Recognition*, pp. 5928–5936, 2018.

[13] A. G. Rempel, M. Trentacoste, H. Seetzen, H. D. Young, W. Heidrich, L. Whitehead, and G. Ward, "Ldr2hdr: on-the-fly reverse tone mapping of legacy video and photographs," *ACM transactions on graphics (TOG)*, vol. 26, no. 3, pp. 39–es, 2007.

[14] C. R. Steffens, L. R. V. Messias, P. L. J. Drews-Jr, and S. S. C. Botelho, "Contrast enhancement and image completion: A cnn based model to restore ill exposed images," in *2019 IEEE 17th International Conference on Industrial Informatics (INDIN)*, p. 8, IEEE, July 2019.

[15] M. Afifi, K. G. Derpanis, B. Ommer, and M. S. Brown, "Learning multi-scale photo exposure correction," 2021.

[16] C. Guo, C. Li, J. Guo, C. C. Loy, J. Hou, S. Kwong, and R. Cong, "Zero-reference deep curve estimation for low-light image enhancement," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2020.

[17] C. Wei, W. Wang, W. Yang, and J. Liu, "Deep retinex decomposition for low-light enhancement," in *British Machine Vision Conference*, 2018.

[18] R. Yu, W. Liu, Y. Zhang, Z. Qu, D. Zhao, and B. Zhang, "Deepexposure: Learning to expose photos with asynchronously reinforced adversarial learning," in *Proceedings of the 32nd International Conference on Neural Information Processing Systems*, NIPS'18, (Red Hook, NY, USA), p. 2153–2163, Curran Associates Inc., 2018.

[19] J.-Y. Zhu, T. Park, P. Isola, and A. A. Efros, "Unpaired image-to-image translation using cycle-consistent adversarial networks," in *Computer Vision (ICCV), 2017 IEEE International Conference on*, 2017.

[20] P. Isola, J.-Y. Zhu, T. Zhou, and A. A. Efros, "Image-to-image translation with conditional adversarial networks," *CVPR*, 2017.

[21] O. Ronneberger, P.Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, vol. 9351 of *LNCS*, pp. 234–241, Springer, 2015.

[22] C. Steffens, V. Huttner, L. Messias, P. Drews, S. Botelho, and R. Guerra, "Cnn-based luminance and color correction for ill-exposed images," in *2019 IEEE International Conference on Image Processing (ICIP)*, pp. 3252–3256, IEEE, 2019.

[23] V. Bychkovsky, S. Paris, E. Chan, and F. Durand, "Learning photographic global tonal adjustment with a database of input / output image pairs," in *CVPR 2011*, pp. 97–104, 2011.

[24] R. Wang, Q. Zhang, C. Fu, X. Shen, W. Zheng, and J. Jia, "Underexposed photo enhancement using deep illumination estimation," in *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 6842–6850, 2019.

[25] I. J. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial nets," in *Proceedings of the 27th International Conference on Neural Information Process-*

ing Systems - Volume 2, NIPS'14, (Cambridge, MA, USA), p. 2672–2680, MIT Press, 2014.

[26] M. D. Fairchild, "The hdr photographic survey," *Color and Imaging Conference*, vol. 2007, no. 1, pp. 233–238, 2007.

[27] R. Mantiuk, S. Daly, K. Myszkowski, and H.-P. Seidel, "Predicting visible differences in high dynamic range images - model and its calibration," in *Human Vision and Electronic Imaging X, IS and T/SPIE's 17th Annual Symposium on Electronic Imaging (2005)* (B. E. Rogowitz, T. N. Pappas, and S. J. Daly, eds.), vol. 5666, pp. 204–214, 2005.

[28] M. Pedersen and J. Y. Hardeberg, "Full-reference image quality metrics: Classification and evaluation," vol. 7, p. 1–80, Jan. 2012.

[29] F. Yang, H. Yang, J. Fu, H. Lu, and B. Guo, "Learning texture transformer network for image super-resolution," in *CVPR*, June 2020.

[30] Z. Yin, K. Liang, Z. Ma, and J. Guo, "Duplex contextual relation network for polyp segmentation," *ArXiv*, vol. abs/2103.06725, 2021.