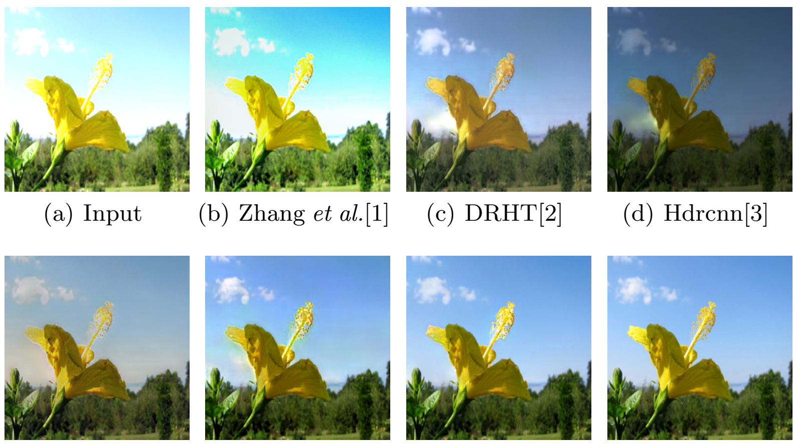
Over-exposure Correction via Exposure and Scene Information Disentanglement

Yuhui Cao[1,2], Yurui Ren[1,2], Thomas H. Li[1,3], and Ge Li✉[1,2]

[1] School of Electronics and Computer Engineering, Peking University
[2] Peng Cheng Laboratory
[3] Advanced Institute of Information Technology, Peking University
{yuhuicao,yrren}@pku.edu.cn tli@aiit.org.cn geli@ece.pku.edu.cn



(a) Input (b) Zhang *et al.*[1] (c) DRHT[2] (d) Hdrcnn[3]

(e) Lu *et al.*[4] (f) Yu *et al.*[5] (g) Ours (h) Ground-truth

Fig. 1. Over-exposure correction results of several image correction methods. As shown in (b)-(f), existing methods have the limitation in recovering the saturated details of the overexposed region. In comparison, as shown in (h), we recover the saturated details and generate a naturalness-preserved result.

Abstract. Over-exposure correction is an important problem of great consequence to social media industries. In this paper, we propose a novel model to tackle this task. Considering that reasonable enhanced results can still vary in terms of exposure, we do not strictly enforce the model to generate identical results with ground-truth images. On the contrary, we train the network to recover the lost scene information according to the existing information of the over-exposure images and generate naturalness-preserved images. Experiments compared with several state-of-the-art methods show the superior performance of the proposed network. Besides, we also verify our hypothesis with ablation studies. Our source code is available at `https://github.com/0x437968/overexposure-correction-dise`.

1 Introduction

In photography, exposure is one of the most important parameters that determine the subjective quality of the captured images. Unreasonable exposure can lead to significant quality degradation. Over-exposure is one of the typical quality degradation phenomenons. Due to the limited dynamic range of digital cameras, relatively bright areas of the scene will be saturated. Therefore, it is important to reasonably recover the saturated information and enhance the quality of over-exposed images. In general, overexposure correction aims to generate alternative contents for saturated regions according to the existing image information while maintaining the global contrast.

Image inpainting also requires generating plausible pixels for corrupted holes according to uncorrupted contents. Along with the rapid progress in deep learning in recent years, inpainting methods [6,7,5,8,9] achieve excellent development. Many methods can generate realistic alternative results. However, it is unreasonable to employ inpainting methods on the over-exposure correction task. There are some important differences between these two tasks. Firstly the missing regions of the images in inpainting are random masks but the overexposed regions in over-exposure correction are correlated. Secondly, to generate reasonable results, the existing contents of the overexposed image are required to be adjusted in the over-exposure correction task.

High Dynamic Range (HDR) images can convey much richer contrasts than conventional Low Dynamic Range (LDR) images. Inverse tone mapping aims to transform the LDR contents into HDR contents, the saturated region in the LDR image has to be recovered in this process. Previous methods of inverse tone mapping employ individual heuristics or optionally use manual intervention to enhance LDR images. Considering the excellent inference ability of convolution neural network, recent works DrTMo[10], Hdrcnn[3], DRHT[2] utilize deep convolutional neural networks to infer HDR results, then they can correct the images exposure by tone mapping, conventional methods or deep learning[11]. Unfortunately, existing inverse tone mapping methods pay more attention to the projection of existing contents but not the recovery of missing contents. The goal of the over-exposure task emphasizes the recovery of the missing contents more than adjusting the existing contents.

In this paper, we find that the reasonable correction result of an overexposed image is not unique, *i.e.* they may vary in terms of exposure. Therefore, it is unreasonable to force the network to predict identical results with ground-truth images. The network can not focus on missing information reconstruction. It struggles to predict identical results with ground-truth images. Hence, the over-exposure correction method should merely recover that information which is unrelated to camera exposure. To achieve this goal, we propose a novel method for over-exposure correction via global exposure and scene information disentanglement. We first train a disentanglement network to split the exposure and scene information of images. Then we utilize the pre-trained disentanglement network to constrain the recovery network. We train this network to generate results with the same scene information with ground-truth images. Meanwhile, we use GAN[12] to constrain the results to have the same distribution with ground-truth images. Experiments compared with the state-of-the-art methods show the superior performance of our proposed methods. Ablation studies also prove our hypothesis.

Our main contributions are as follows:

1) We tackle the over-exposure correction problem by disentangling global exposure and scene information.
2) We show that the performance of the network can be largely improved by reconstructing the scene information which is unrelated to the image exposure.
3) Our methods achieve remarkable results compared with state-of-the-art methods.

2 Related Work

We discuss the works that are relevant to the over-exposure correction task in this section, including over-exposure correction in Section2.1, inverse tone mapping in Section2.2, image inpainting in Section2.3.

2.1 Over-exposure Correction

Although overexposed image correction is an important task for many kinds of researches, there is not much previous work directly addressing this problem. Some early works assume that the ratios between different color channels are invariant[13] or gradual[14] in local image regions. However, both methods can only handle pixels which have one or two channels overexposed and all overexposed pixels are left untouched. To deal with this problem. Guo *et al.* [15] separate the input images into lightness and color. Then different smooth operators are performed to these components to correct the inputs. Although this algorithm can generate some color information in overexposed regions, it cannot recover the complicated texture. Considering the excellent performance of the Retinex theory in the under-exposure correction task, SICE[16] proposes a network to respectively recover the reflectance and illumination maps by using the Retinex theory. Then they reconstruct the results by combining these two components. Zhang *et al.*[1] propose a dual illumination estimation to simultaneously process under-exposure and over-exposure images. But the method can not recover vivid textures while dealing with the images which are overexposed in all RGB channels.

2.2 Inverse Tone Mapping

Inverse tone mapping is used to describe the methods that expand Low Dynamic Range (LDR) images for the generation of High Dynamic Range (HDR) images[17]. HDR images contain a broader range of physical information of scenes than LDR images. Therefore, generating HDR images from captured LDR images is an ill-posed problem. It requires the algorithms to recover the lost dynamic information from the over/under-exposed regions in LDR images. Previous methods of inverse tone mapping employ individual heuristics or optionally use manual intervention to enhance LDR images. Rapid progress in deep learning inspired recent learning-based methods. DrTMo[10] and Hdrcnn[3] introduce the learning-based approach by training the LDR and HDR image pairs and infer a reasonable HDR image from an LDR input. DRHT[2] further uses an auto-encoder network to map the generated HDR images back to LDR images. However, existing inverse tone mapping methods always pay more attention to the projection of existing contents but not the recovery of missing information.

2.3 Image Inpainting

Existing inpainting methods can be mainly divided into two groups: conventional methods that use diffusion-based or patch-based methods and learning-based methods that employ convolutional neural networks to infer pixels for the missing regions. The conventional methods such as [18,19,20]. The synthesize pixels by propagating the neighborhood region's appearance to the target holes searching and copying similar image patches from the uncorrupt region. However, the diffusion-based methods can only deal with small holes in background inpainting tasks. The patch-based methods can not generate reasonable results for images with unique structures.

Recently, many learning-based methods[6,5,21,8,22,9] are proposed by formulating inpainting as a conditional image generation problem. A significant advantage of the deep-learning-based methods is that they can infer results by extracting meaningful semantic information. Context Encoder[6] propose an auto-encoder network for image inpainting. However, this method often generates results with visual artifacts. To solve this problem, Iizuka *et al.*[7] use both local and global discriminators to improve the quality of the generated images. In order to make better predictions, Yu *et al.*[5] propose contextual attention to building a remote connection when generated contents are distant with existing information. Liu *et al.*[21] believe the pixels in the masked holes of the inputs introduce artifacts to the results. Therefore, they propose partial convolution to force the network to use uncorrupted pixels only.

3 Proposed Method

Given an overexposed image, our goal is to generate a naturalness-preserved result with complete scene information. In order to encourage the network to learn the scene information unrelated to image exposure, we first train a disentanglement network to separate the image exposure and scene information. Then we utilize the pre-trained disentanglement network to generate the scene information. Our proposed model consists of two parts: 1) Disentanglement network. 2) Recovery network. In the following subsections, we particularly introduce our model.

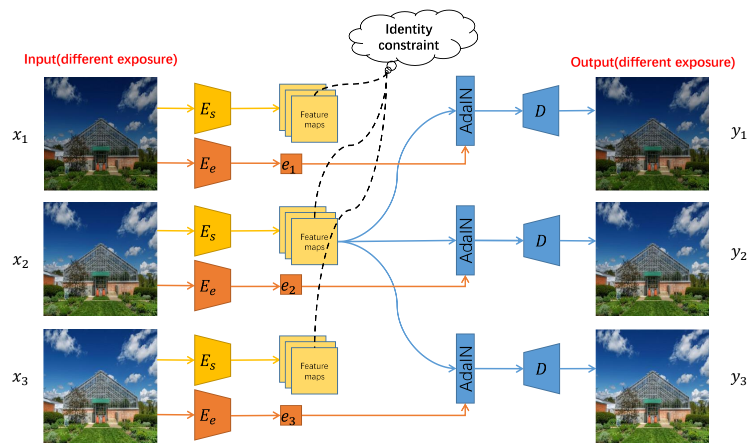


Fig. 2. Overview of the disentanglement network, the network includes scene information encoderE_s, exposure informaiton encoder E_e and decoder D. We set $N = 3$.

3.1 Disentanglement Network

Our disentanglement network can separate the scene information and the exposure information of images. As shown in Fig.2, scene encoder E_s extracts the scene information of the inputs and the exposure encoder E_e extracts the images exposure information. In order to guarantee that the encoders can extract meaningful features. We use decoder D to reconstruct the inputs.

This encoder E_s should extract the scene information that is unrelated to image exposure. To achieve this goal, we use N multi-exposure images $x_1, x_2, ..., x_N$ which are captured with different exposure as the network inputs. Let $s_1, s_2, ..., s_N$ represent the feature maps which are extracted by E_s and $e_1, e_2, ..., e_N$ represent the exposure information vectors of E_s. The scene information in the same scene should be identical, therefore we define the scene loss as follows:

$$\mathcal{L}_{Ds} = \sum_{i=1}^{N} \|s_i - \overline{s}\|_1 \tag{1}$$

where $\overline{s}$ represents the mean of s. This loss ensures that the image exposure will not influence the scene feature maps.

In order to guarantee complete scene information in s_i, we further reconstruct the inputs by decoder D. During the training, we randomly select a feature from $s_1, s_2, ..., s_N$ as the input of decoder D. Meanwhile, we inject the N exposure feature vectors into the picked scene feature via AdaIN[23]. Then we get N results $y_1, y_2, ..., y_N$ that are in different exposure. We define the reconstruction loss as follows:

$$\mathcal{L}_{Dr} = \sum_{i=1}^{N} \|y_i - x_i\|_1 \tag{2}$$

Meanwhile, we add a KL divergence loss to regularize the distribution of the exposure feature vectors $e_1, e_2, ..., e_N$ to be close to normal distribution $p(z) \sim N(0,1)$. The KL divergence loss is defined as follows

$$KL(q(e_i)||p(z)) = -\int q(e_i) \log \frac{p(z)}{q(e_i)} dz \tag{3}$$

As shown in [24], minimizing the KL divergence is equivalent to minimizing the following loss:

$$\mathcal{L}_{KL} = \frac{1}{2} \sum_{i=1}^{N} (\mu_i^2 + \sigma_i^2 - \log(\sigma_i^2) - 1) \tag{4}$$

where μ is the mean of e, and σ is the standard of e. e is sampled as $e = \mu + z \circ \sigma$, where $p(z) \sim N(0,1)$, and $\circ$ represents element-wise multiplication.

Besides, to help the recovery of more vivid textures. We add the style loss of the perceptual loss[25] between the outputs and ground-truth images:

$$\mathcal{L}_{Dstyle}(x,y) = \left\|G^{\phi}(x) - G^{\phi}(y)\right\|_F^2 \tag{5}$$

where G^{ϕ} represents the output features' Gram matrices of VGG-19 network[14] which is pre-trained on ImageNet[26].

The full objective function of the disentanglement network is a weighted sum of all the losses from (1) to (5)

$$\mathcal{L}_D = \lambda_s \mathcal{L}_{Ds} + \lambda_r \mathcal{L}_{Dr} + \lambda_{KL} \mathcal{L}_{KL} + \lambda_{style} \mathcal{L}_{Dstyle} \tag{6}$$

3.2 Recovery Network

In this section, we introduce the recovery network. As shown in Fig.3, we set the overexposed image x as the input of generator G. Then we train the generator G to recovery the scene information. We compute the *Manhattan distance* of the scene feature maps by using pre-trained E_s. Therefore, our scene information reconstruction loss is defined as follows:

$$\mathcal{L}_{Rsr} = \|E_s(\widehat{y}) - E_s(y)\|_1 \tag{7}$$

Meanwhile, we add the adversarial loss to mimic the distribution of true images:

$$\mathcal{L}_{adv} = \mathbb{E}_{y \sim p(y)}[\log D_e(y)] + \mathbb{E}_{x \sim p(x)}[\log(1 - D_e(G(x)))] \tag{8}$$

where D_e tries to maximize the objective function to distinguish between our recovered results and ground-truth images. On the contrary, G aims to minimize

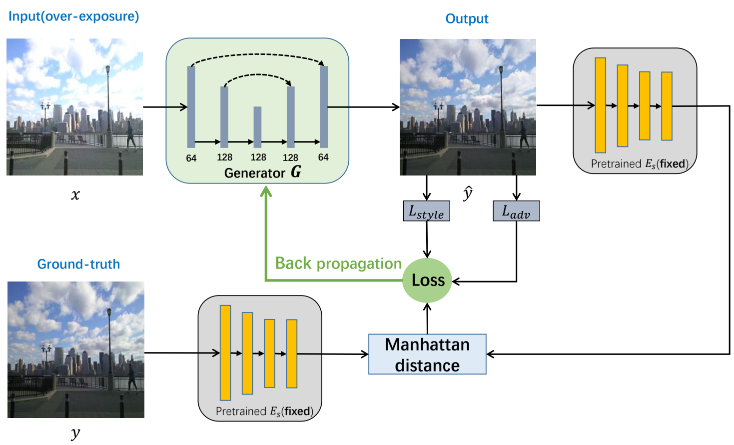


Fig. 3. Overview of the recovery network. We utilize the pre-trained scene information encoder E_s to constrain the recovery network.

the loss to make our recovered results look similar to real samples in ground-truth images.

We also add the style loss of [25] to recover more vivid textures:

$$\mathcal{L}_{Rstyle}(\widehat{y}, y) = \left\| G^{\phi}(\widehat{y}) - G^{\phi}(y) \right\|_F^2 \tag{9}$$

The full loss of the recovery network is defined as follows:

$$\mathcal{L}_R = \gamma_{sr}\mathcal{L}_{Rsr} + \gamma_{adv}\mathcal{L}_{adv} + \gamma_{style}\mathcal{L}_{Rstyle} \tag{10}$$

3.3 Implement Details

Disentanglement network. For the architecture of the disentanglement network, we follow similar structures as the one used[4]. We employ the convolutional layer and the residual block[27] as the basic components of the network. In order to generate multi-exposure images, we use the method[28] proposed by Ying *et al.* to adjust the exposure of the image and guarantee that the exposure changed images won't overexposed. During the training, we use Adam solver to update our encoders and decoder. The learning rate is fixed in 0.0001. In all the experiments, we use 256×256 images with a batch size of 4 for training. For the hyper-parameters, we set $\lambda_s = 1000$, $\lambda_r = 50$, $\lambda_{KL} = 0.01$, $\lambda_{style} = 1000$ and $N = 3$.

Recovery network. For the recovery network, we employ the auto-encoder architecture on the generator G. We use Adam solver to update our generator and discriminator. The learning rate of the generator is fixed in 0.0001 and the learning rate of the discriminator is fixed in 0.00001. In all the experiments, we use 256×256 images with a batch size of 4 for training. For the hyper-parameters, we set $\gamma_{sr} = 10000$, $\gamma_{adv} = 2$, $\gamma_{style} = 2000$.

4 Experiments

In this section, we first discuss the experiments deploys. Then, we compare our method against several state-of-the-art methods including Zhang *et al.*[1], DRHT[2], Hdrcnn[3], Lu *et al.*[4], Yu *et al.*[5]. Finally, we analyze the ablation study.

4.1 Experiments Deploys

Datasets. We implement our experiments on the place365 dataset[29]. Considering that the images containing sky, human face are very easily overexposed in photography. We elaborately pick 5000 outdoor images and 2400 images of the human portrait as our ground-truth images in the place365 dataset[29] and all of the images are in the normal exposure. Then we use the method[28] proposed by Ying *et al.* to adjust the exposure of images and obtain overexposed images from the ground-truth images. For each image, the ratio of exposure between the overexposed image and the ground-truth image is randomly selected in [1.8,2.4]. All images are resized to 256×256. For the outdoor dataset, we use 4000 images for training and 1000 images for testing. In the portrait dataset, we use 2000 images for training and 400 images for testing.

Evaluation metrics. For the evaluation of experiments, considering that the reasonable correction results of an overexposed image are not unique in the overexposure correction task, it is unreasonable to use metric which is sensitive about image exposure such as PSNR. We first use *Frchet Inception Distance*(FID)[31] and *Kernel Inception Distance*(KID)[32] to measure the performance. Considering the outperformance of deep features compared with classic metrics, we also use LPIPS[33] to evaluate the performance. Besides, considering that the output of the scene encoder E_s can represent the scene information of the inputs. We formulate Scene Information Identity(SII) as the scene information evaluation metric:

$$SII(y, \widehat{y}) = \|E_s(y) - E_s(\widehat{y})\|_1 \tag{11}$$

4.2 Comparisons

We compare the proposed method with Hdrcnn[3], DRHT[2], Zhang *et al.*[1], Yu *et al.*[5] and Lu *et al.*[4]. Both Hdrcnn[3] and DRHT[2] are the inverse tone mapping methods that use an auto-encoder network to infer the HDR image from

(a) Inputs (b) Zhang *et al.*[1] (c) DRHT[2] (d) Hdrcnn[3]

(e) Lu *et al.*[4] (f) Yu *et al.*[5] (g) Ours (h) Ground-truth

Fig. 4. Visual portrait comparisons of our method with Zhang *et al.*[1], DRHT[2], Hdrcnn[3], Lu *et al.*[4] and Yu *et al.*[5]. The size of images on the bottom row is 512×512 and the others are 256×256.

Table 1. Quantitative comparisons of our method with other methods. SII is defined in Eq.11. The results show the superior performance of the proposed method.

Datasets	Outdoor				Portrait			
Metrics / Methods	FID	KID	LPIPS	SII	FID	KID	LPIPS	SII
Ours	**4.1477**	**-6.1942**	**0.0210**	**1.1110**	**19.2002**	**-8.9204**	**0.0252**	**0.6594**
Yu *et al.*[5]	10.8996	-5.6884	0.0398	3.5771	28.3579	-8.4790	0.0321	0.7399
Zhang *et al.*[1]	13.1616	-5.6805	0.1407	16.2549	41.3969	-7.8780	0.1252	4.3884
DRHT[2]	15.7146	-5.1977	0.1001	16.3488	47.3493	-7.4641	0.1430	7.2226
Hdrcnn[3]	14.9932	-5.3118	0.1737	13.7195	59.1835	-6.5646	0.2350	7.9029
Lu *et al.*[4]	17.8173	-5.1606	0.0826	5.1949	41.0575	-7.9727	0.0732	3.0559

an LDR input. Zhang *et al.*[1] is a Retinex-based conventional method which can process both over-exposure and under-exposure by inverting the inputs. Yu *et al.*[5] is an excellent inpainting method and Lu *et al.*[4] is a deblurring method.

For DRHT[2] and Hdrcnn[3]. we use the pre-trained model[4] provided by the authors to predict the overexposed images. For Zhang *et al.*[1], we test the overexposed images on the code provided by the authors. We retrain [5] and [4] on our datasets.

The quantitative results are shown in Table 1 and the visual results are shown Fig.4 and Fig.5. For the visual results, we also add 512×512 size results in the bottom row of Fig.4 to show the applicability of the proposed method. By comparing the quantitative results in Table 1, the proposed method achieves the remarkable results in the experiments. As shown in Fig.5(b), Zhang *et al.*[1] has the limitation in recovering missing information when all RGB channels are overexposed. We can realize that the inverse tone mapping methods pay more attention to the projection of existing contents via observing Fig.4(c) and Fig.4(d). The results of Yu *et al.*[5] are relatively well in the mid-row of Fig.5(f). That may is because of the coarse-to-fine architectures. But as shown in both the mid and bottom rows of Fig.4(f), the results of [5] also exist unreasonable artifacts and still have a big gap with our results. Some meaningful results are shown in the bottom row of Fig.5. Our method recovers the missing contents on the road and the textures of the overexposed cloud. Specifically, we also recover the details at the end of double amber lines(*i.e.*, red box in Fig.5(h)).

To further prove the generalization ability and applicability of the proposed method, we also test our pre-trained model on the CelebA dataset[30] and the SICE dataset[16]. Some results on the CelebA dataset[30] are shown in Fig.6 and the overexposed inputs are obtained in the same way that is mentioned in Section4.1. The results on the SICE dataset[16] are shown in Fig.7 and the inputs are from the real world. Considering that the large size of the original images in the SICE dataset[16] can make the testing difficult, we resize the inputs to 768×512.

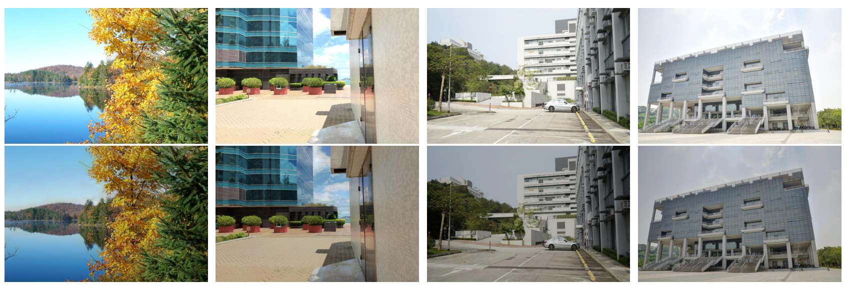
[4] HDR images are required in DRHT[2] and Hdrcnn[3], therefore we can not retrain these two methods.

(a) Inputs (b) Zhang *et al.*[1] (c) DRHT[2] (d) Hdrcnn[3]

(e) Lu *et al.*[4] (f) Yu *et al.*[5] (g) Ours (h) Ground-truth

Fig. 5. Visual outdoor comparisons of our method with Zhang *et al.*[1], DRHT[2], Hdrcnn[3], Lu *et al.*[4] and Yu *et al.*[5].

Fig. 6. Visual results on the CelebA dataset[30]. Images on the top row are the over-exposed inputs. The bottom row images are our results.



(a)

(b)

Fig. 7. Visual results in the real world. For each subfigure, images on the top row are the inputs which are from the SICE dataset[16] and are resized to 768×512. Images on the bottom row are our results.

4.3 Ablation studies.

In this paper, we believe that the reasonable correction result of an overexposed image is not unique. Therefore it is unreasonable to force the network to generate identical results with ground-truth images. In the ablation studies, we first relace the scene information reconstruction loss with L_1 distance between the generated results and the ground-truth images and maintain the others unchanged to retrain the recovery network. Then we remove one of the Style, GAN, Scene loss and maintain the others unchanged to retrain the recovery network in turn for determining the role of each loss. Qualitative results are shown in Fig.8. We can see that figures in Fig. 8(e) and 8(b) which are trained without the scene loss exists severe artifacts, but the results in Fig.8(c), 8(d), 8(f) which are trained with the scene loss are under a good condition. Besides, the comparisons between Fig. 8(c), 8(d), 8(f) also denote that Scene loss has a major contribution to the proposed method. Quantitative results are shown in Table2. It also could be seen that the comparisons between the results trained with scene loss and the results trained without scene loss denote that the scene loss can largely improve the performance of the recovery network.

(a) Inputs (b) No Scene (c) No Style (d) No GAN (e) L_1 (f) Scene

Fig. 8. Visual results of the ablation studies. Figures in (b)-(d) respectively are the ablation results without corresponding loss(*e.g.* results in (b) are trained with GAN and Style loss).Figures in (e) are the results trained with GAN, Style, L_1 loss and results in (f) are trained with Style, GAN, Scene loss. The comparisons between the different ablation results show the significant impacts of the scene loss.

4.4 Failure Cases

Despite the aforementioned success, our method contains limitations in recovering the details of a large continuously overexposed region. Fig.9 shows two

Table 2. Quantitative results of the ablation studies. SII is defined in Eq.11. The results verify the significant impacts of Scene loss.

Datasets	Outdoor				Portrait			
Loss \ Metrics	FID	KID	LPIPS	SII	FID	KID	LPIPS	SII
Style,GAN,Scene	**4.1477**	**-6.1942**	**0.0210**	1.1110	19.2002	-8.9204	0.0252	0.6594
Style,GAN,L_1	8.8833	-5.8798	0.0719	1.8312	20.5921	-8.8673	0.0270	0.7792
Style,GAN	9.1331	-5.8819	0.0767	1.6240	19.8518	-8.8826	0.0262	0.7630
GAN,Scene	5.4939	-6.1175	0.0249	**1.0205**	18.2920	-8.9328	**0.0222**	**0.6107**
Style,Scene	4.6370	-6.1759	0.0250	1.6615	**17.9326**	**-8.9596**	0.0235	0.6914

examples. Although our method can recover the most textures of the overexposed region, there are limitations in recovering the center of the overexposed region. This is because limited information is given in the input images.

(a)

(b)

Fig. 9. Failure examples. For each subfigure, left is the input and right is the result.

5 Conclusion

In this paper, we find that the reasonable correction result of an overexposed image is not unique. To tackle the over-exposure correction task, we propose a novel method via disentangling the image exposure and the scene information. In order to force the recovery network to focus on the scene information recovery, we first train a network to disentangle the image exposure and scene information, and then we utilize the pre-trained scene information encoder to constrain the recovery network. Our method achieves remarkable results in comparisons with other state-of-the-art methods. The ablation studies also verify that the proposed method can largely improve the performance by forcing the network to reconstruct the scene information.

Acknowledgements. This work was supported by Key-Area Research and Development Program of Guangdong Province(No.2019B121204008), National Natural Science Foundation of China and Guangdong Province Scientific Research on Big Data(No. U1611461). In addition, we thank the anonymous reviewers for their time and valuable comments.

References

1. Zhang, Q., Nie, Y., Zheng, W.S.: Dual Illumination Estimation for Robust Exposure Correction. Computer Graphics Forum **38** (2019) 243–252
2. Yang, X., Xu, K., Song, Y., Zhang, Q., Wei, X., Lau, R.W.: Image correction via deep reciprocating hdr transformation. In: The IEEE Conference on Computer Vision and Pattern Recognition (CVPR). (2018)
3. Eilertsen, G., Kronander, J., Denes, G., Mantiuk, R.K., Unger, J.: HDR image reconstruction from a single exposure using deep CNNs. ACM Transactions on Graphics **36** (2017)
4. Lu, B., Chen, J.C., Chellappa, R.: Unsupervised domain-specific deblurring via disentangled representations. Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition **2019-June** (2019) 10217–10226
5. Yu, J., Lin, Z., Yang, J., Shen, X., Lu, X., Huang, T.S.: Generative Image Inpainting with Contextual Attention. Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (2018) 5505–5514
6. Pathak, D., Krahenbuhl, P., Donahue, J., Darrell, T., Efros, A.A.: Context Encoders: Feature Learning by Inpainting. Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition **2016-December** (2016) 2536–2544
7. Iizuka, S., Simo-Serra, E., Ishikawa, H.: Globally and locally consistent image completion. ACM Transactions on Graphics **36** (2017)
8. Ren, Y., Yu, X., Zhang, R., Li, T.H., Liu, S., Li, G.: Structureflow: Image inpainting via structure-aware appearance flow. In: The IEEE International Conference on Computer Vision (ICCV). (2019)
9. Zhang, R., Ren, Y., Qiu, J., Li, G.: Base-detail image inpainting. In: BMVC. (2019)
10. Endo, Y., Kanamori, Y., Mitani, J.: Deep reverse tone mapping. ACM Transactions on Graphics **36** (2017)
11. Gharbi, M., Chen, J., Barron, J.T., Hasinoff, S.W., Durand, F.: Deep bilateral learning for real-time image enhancement. CoRR **abs/1707.02880** (2017)
12. Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., Bengio, Y.: Generative adversarial nets. In Ghahramani, Z., Welling, M., Cortes, C., Lawrence, N.D., Weinberger, K.Q., eds.: Advances in Neural Information Processing Systems 27. Curran Associates, Inc. (2014) 2672–2680
13. Zhang, X., Brainard, D.H.: Estimation of saturated pixel values in digital color imaging. J. Opt. Soc. Am. A **21** (2004) 2301–2310
14. Masood, S., Zhu, J., Tappen, M.: Automatic correction of saturated regions in photographs using cross-channel correlation. Comput. Graph. Forum **28** (2009) 1861–1869
15. Guo, D., Cheng, Y., Zhuo, S., Sim, T.: Correcting over-exposure in photographs. In: 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. (2010) 515–521
16. Cai, J., Gu, S., Zhang, L.: Learning a deep single image contrast enhancer from multi-exposure images. IEEE Transactions on Image Processing **27** (2018) 2049–2062
17. Banterle, F., Ledda, P., Debattista, K., Chalmers, A.: Inverse tone mapping. (2006) 349–356
18. Takahashi, T.: Image inpainting. Kyokai Joho Imeji Zasshi/Journal of the Institute of Image Information and Television Engineers **71** (2017) 503–504

19. Barnes, C., Shechtman, E., Finkelstein, A., Goldman, D.B.: PatchMatch: A randomized correspondence algorithm for structural image editing. ACM Transactions on Graphics (Proc. SIGGRAPH) **28** (2009)
20. Hays, J., Efros, A.A.: Scene completion using millions of photographs. ACM Transactions on Graphics **26** (2007) 1–7
21. Liu, G., Reda, F.A., Shih, K.J., Wang, T., Tao, A., Catanzaro, B.: Image inpainting for irregular holes using partial convolutions. CoRR **abs/1804.07723** (2018)
22. Li, J., Wang, N., Zhang, L., Du, B., Tao, D.: Recurrent feature reasoning for image inpainting. In: The IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). (2020)
23. Huang, X., Belongie, S.: Arbitrary Style Transfer in Real-Time with Adaptive Instance Normalization. Proceedings of the IEEE International Conference on Computer Vision **2017-October** (2017) 1510–1519
24. Kingma, D.P., Welling, M.: Auto-encoding variational bayes. 2nd International Conference on Learning Representations, ICLR 2014 - Conference Track Proceedings (2014) 1–14
25. Johnson, J., Alahi, A., Fei-Fei, L.: Perceptual losses for real-time style transfer and super-resolution. Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics) **9906 LNCS** (2016) 694–711
26. Deng, J., Dong, W., Socher, R., Li, L., Kai Li, Li Fei-Fei: Imagenet: A large-scale hierarchical image database. In: 2009 IEEE Conference on Computer Vision and Pattern Recognition. (2009) 248–255
27. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. CoRR **abs/1512.03385** (2015)
28. Ying, Z., Li, G., Ren, Y., Wang, R., Wang, W.: A New Low-Light Image Enhancement Algorithm Using Camera Response Model. Proceedings - 2017 IEEE International Conference on Computer Vision Workshops, ICCVW 2017 **2018-January** (2017) 3015–3022
29. Zhou, B., Lapedriza, A., Khosla, A., Oliva, A., Torralba, A.: Places: A 10 million image database for scene recognition. IEEE Transactions on Pattern Analysis and Machine Intelligence (2017)
30. Liu, Z., Luo, P., Wang, X., Tang, X.: Deep learning face attributes in the wild. In: Proceedings of International Conference on Computer Vision (ICCV). (2015)
31. Heusel, M., Ramsauer, H., Unterthiner, T., Nessler, B., Hochreiter, S.: GANs trained by a two time-scale update rule converge to a local Nash equilibrium. Advances in Neural Information Processing Systems **2017-December** (2017) 6627–6638
32. Binkowski, M., Sutherland, D., Arbel, M., Gretton, A.: Demystifying mmd gans. ArXiv **abs/1801.01401** (2018)
33. Zhang, R., Isola, P., Efros, A.A., Shechtman, E., Wang, O.: The Unreasonable Effectiveness of Deep Features as a Perceptual Metric. Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (2018) 586–595