Neural Augmentation Based Saturation Restoration for LDR Images of HDR Scenes

Chaobing Zheng,Wenjian Ying, Shiqian Wu*, *Senior Member, IEEE*, and Zhengguo Li, *Fellow, IEEE*

***Abstract*—There are shadow and highlight regions in a low dynamic range (LDR) image which is captured from a high dynamic range (HDR) scene. It is an ill-posed problem to restore the saturated regions of the LDR image. In this paper, the saturated regions of the LDR image are restored by fusing model-based and data-driven approaches. With such a neural augmentation, two synthetic LDR images are first generated from the underlying LDR image via the new model-based approach. It relaxes the requirement of mapping integers to integers and improves the modeling accuracy. One is brighter than the input image to restore the shadow regions and the other is darker than the input image to restore the high-light regions. Both synthetic images are then refined via one single exposedness aware saturation restoration network (EASRN). Finally, the two synthetic images and the input image are combined together via an HDR synthesis algorithm or a multi-scale exposure fusion algorithm. Experimental results indicate that the proposed algorithm outperforms existing algorithms in terms of HDR-VDP-3. The proposed algorithm can be embedded in any smart phones or digital cameras to produce an information-enriched LDR image.**

***Index Terms*—High dynamic range imaging, saturation restoration, model-based, data-driven, neural augmentation.**

I. Introduction

High dynamic range (HDR) imaging provides realistic viewing experience by capturing scene appearance including lighting, contrasts and details. HDR technique has been already used in a variety of applications on realistic image or video synthesis and thus has attracted more and more attention [1], [2], [3], [4], [5]. Most smart phones set the HDR mode as their default mode. Multiple differently exposed low dynamic range (LDR) images are captured to synthesize an HDR image [1]. When such a set is captured by a handhold device with a conventional sensor, there might be cameral movement or moving objects in the captured images. Even though many alignment and ghost removal techniques have been proposed [6], [7], [8], [10], [11], [12], [13], [14], both topics, especially ghost removal are still issues [15]. The ghost artifacts can be solved by specialized high-end imaging devices [16], [17] but their cost could be an issue.

One alternative solution is proposed in this paper to capture one LDR image by using a smart phone or a digital camera and restore the saturated regions of the LDR image to generate an information-enriched LDR image. Generally, there are shadow and highlight regions in the LDR image. It is desirable to restore both the shadow and the highlight regions for the LDR image. To this end, two synthetic images, one with a smaller exposure, the other with a larger exposure are generated to restore the highlight and shadow regions, respectively. Clearly, the proposed method is an indirect method which is different from existing direct methods in [18], [19], [20], [21]. The proposed indirect method allows a user to combine all the three images into an HDR image or an LDR image according to her/his preference. It is noted that the data-driven reverse tone mapping algorithms in [22], [24] are also on top of the indirect method. In this paper, the saturation is restored by fusing data-driven and model-based approaches rather than the data-driven approaches in [22], [24]. As indicated in [25], such a neural augmentation must posses the complete domain knowledge it requires as in [29]. The camera response functions (CRFs) are thus assumed to be known in advance. In other words, the proposed algorithm is supposed to be embedded into a smart phone or a digital camera as the usage in [30].

The synthetic images are first generated by a model-based approach and then refined by using a data-driven approach. The combination of the model-driven and data-driven methods forms a neural augmentation [25], [26], [31]. The simplicity of the model-based method is important for the neural augmentation. The synthetic images are initially generated by using intensity mapping functions (IMFs) to avoid lightness distortion [32], [33]. These two synthetic images are refined by one single exposedness aware saturation restoration network (EASRN) which is proposed by incorporating an exposedness aware guidance branch (EAGB) into a nonlocal recursive residual group (NRRG). Intrinsic properties of differently exposed images are well utilized by the proposed EASRN. The exposedness of the input image is used to define two binary masks. The proposed masks are based on the intrinsic properties of differently exposed images: [9], [10]: 1) IMFs from an image with a large exposure time to an image with a small one are not reliable for all pixels in the overexposed regions of the image with the large exposure time; and 2) IMFs from an image with a small exposure time to an image with a large one are not reliable for all pixels in the underexposed regions of the image with the small exposure time. Clearly, the proposed masks are different from the soft mask in [20]. The objective of the EAGB is to help the proposed EASRN pay more attention to the saturated regions in order to restore them efficiently. The NRRG is inspired by the nonlocal block

Chaobing Zheng and Shiqian Wu are with the Institute of Robotics and Intelligent Systems, School of Information Science and Engineering, Wuhan University of Science and Technology, Wuhan 430081, China(e-mails: {zhengchaobing, shiqian.wu}@wust.edu.cn).

Wenjian Ying is with Naval engineering university of the people's liberation arm of China, (email: 842812952@qq.com).

Zhengguo Li is with the SRO Department, Institute for Infocomm Research, A*STAR, Singapore, 138632, (email: ezgli@i2r.a-star.edu.sg).

in [34], [35] and the recursive residual group module in [37] which explores both local and non-local correlation along both spatial dimension and channel dimension of the synthetic images. As shown in [38], [46], each block in the input image could have nonlocal similar blocks. The self-similarity of the input image could be applied to refine the shadow and highlight regions in the initially synthetic images. The two synthetic images and the input image are combined together via the state-of-the-art multi-scale exposure fusion (MEF) algorithm in [42] or the HDR synthesis algorithm in [1]. Experimental results indicate that the EASRN outperforms the corresponding deep neural network (DNN) without the EAGB. The proposed neural augmentation also outperforms the hybrid learning method in [29]. The proposed algorithm can produce better LDR images than the state-of-the-art algorithms in terms of subjective and objective judgements.

It is worth noting that the single image brightening algorithm in [29] is also based on the neural augmentation. The first difference between the algorithms in [29] and this paper is that two different ResNets [27] are trained to refine two virtual images while only one EASERN is trained to refine the two virtual images. The second difference is that the intrinsic properties of differently exposed images are ignored by the DNN in [29] while they are well utilized by the proposed EASRN. The third difference is that the proposed algorithm restores the shadow and highlight regions from both the neighboring areas and the nonlocal similar regions while the algorithm in [29] restores the shadow regions only from the neighboring areas. In summary, the contributions of this paper are highlighted as follows:

1) Our work is the first to restore both underexposed and overexposed regions of the LDR image by fusing model-based and data-driven approaches. The model-based and data-driven saturation restoration *compensates* each other, and form a neural augmentation. Such a neural augmentation has a potential to achieve learning with few data [25], [26].

2) A new method is proposed to estimate the IMFs. The new method relaxes the requirement of mapping integers to integers. As such, the accuracy of the model-driven method is improved.

3) A novel EASRN is proposed by incorporating the EAGB into the NRNG to improve its performance. The intrinsic properties of differently exposed images are well utilized by the EASRN.

The rest of this paper is organized as below. Relevant works on HDR imaging are reviewed in Section II. The proposed saturation restoration algorithm is presented in Section III. Extensive experimental results are provided in Section IV to verify the proposed algorithm. Finally, conclusion remarks are drawn in Section V.

II. Relevant Works on HDR Imaging

An image captured by an ordinary smart phone or digital camera is typically represented by eight-bit integers which are far from adequate as the visible radiance in a real HDR scene has a dynamic range larger than 10^4. An efficient solution to address the problem is to capture a set of differently exposed images [1] and to integrate them together to generate an image that contains the rich information from all of the input images. There are two kinds of solutions to generate a content rich image. One is to synthesize multi-exposed images into an HDR image [1], and then convert the HDR image into an LDR image [51], [52], [53], [54] such that it can be displayed by conventional LDR digital devices. The other is to fuse all of the differently exposed LDR images into a single LDR image [2], [3], [41], [42], [43], [44], [45]. The latter dominates the HDR solution in smart phones.

Both types of methods require the captured scene to be static and the camera to be stationary. The movement of camera would cause the fused image to be blurred while any movement of objects would contribute to a phenomenon called ghosting artifact in the final synthesized output. It is relative easy to align differently exposed images, for example, by the method in [6]. It is much more challenging to prevent ghosting artifacts from appearing in the final image, and ghost artifacts are thus believed to be Achilles's heel for the HDR imaging. The problem could be addressed by using special capturing systems. One example is a beam splitting based HDR capturing system with few sensors [16]. Another one is a row-wise CMOS HDR video capturing system [17], [55]. Canon also released an innovative global shutter with a specific sensor that reads the sensor twice in an HDR mode [56].

Besides the new capturing systems, one more attractive solution is to capture one LDR image by using a smart phone or a digital camera and restore an HDR image from the LDR image. This solution is known as reverse tone mapping [47] and the reverse tone mapping is an ill-posed problem. A simple linear expansion method was proposed in [48]. Psychophysical experiments in [48] show that such simple operation has a potential to outperform a true HDR image. Recently, an indirect approach was proposed in [22] to infer a sequence of synthetic LDR images with different exposures via a deep learning based method, and then reconstruct a pseudo-HDR image by merging the LDR images using the algorithm in [1]. Image formation pipeline including the dequantization, linearization, and hallucination was reversed in [19] to track single-image HDR reconstruction problem. The difficulty of training one single network for reconstructing HDR images is significantly reduced. A data-driven approach was proposed in [20] to reconstruct an HDR image by recovering the saturated pixels of an input LDR image, and a soft feature masking mechanism was introduced to reduce artifacts caused by the invalid information in the saturated regions. The pseudo-HDR image needs to be compressed by using the existing tone mapping algorithms [51] such that it can be displayed by most existing digital devices. The complexity of the pseudo-HDR image could be an issue. It is thus desired to develop a simpler algorithm to restore an LDR image of an HDR scene from the saturation point of view.

Single image brightening also intends to produce an information-enriched LDR image. The input is an image cap-

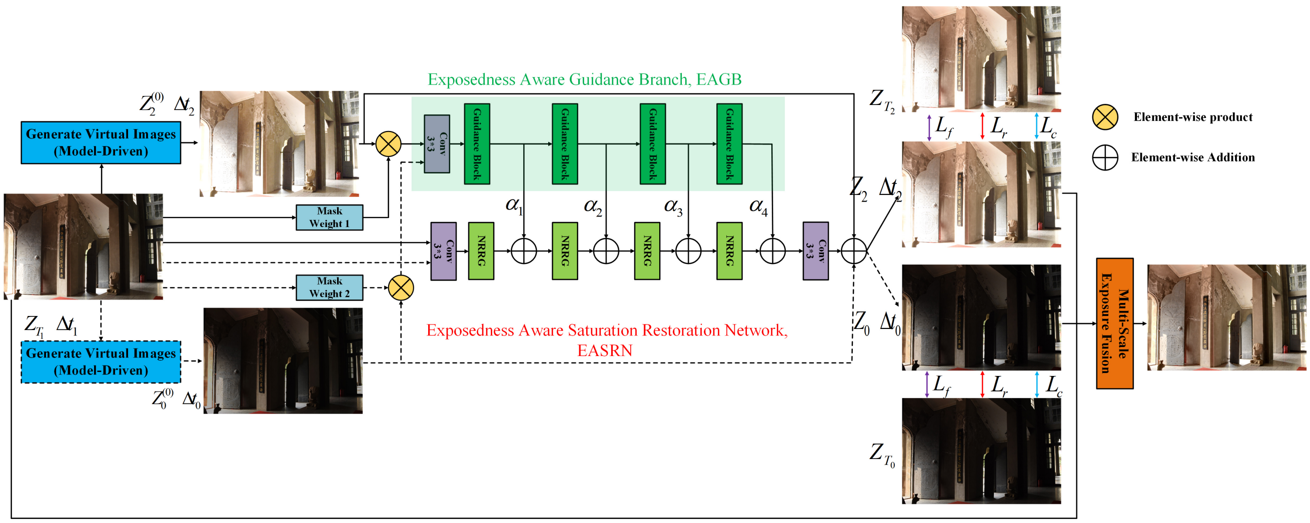


Fig. 1: *The diagram of the proposed saturation restoration algorithm via the neural augmentation. Two synthetic images $Z_0^{(0)}$ and $Z_2^{(0)}$ with Δt_0 and Δt_2 are first generated by the IMF based method. Both $(Z_{T_0} - Z_0^{(0)})$ and $(Z_{T_2} - Z_2^{(0)})$ are then learnt via one single EASRN to enhance the initial synthetic images. The image Z_0, Z_{T_1}, and Z_2 are finally fused to obtain an information enriched LDR image.*

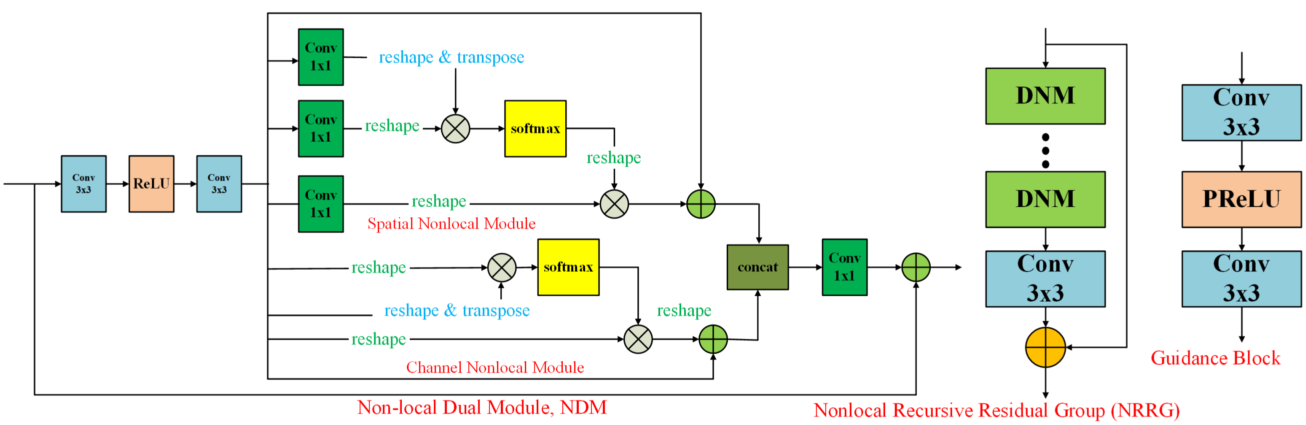


Fig. 2: *Nonlocal recursive residual group (NRRG) contains multiple non-local dual modules (NDMs). Each NDM contains non-local spatial and channel modules. The guidance block includes two $3 * 3$ convolution layers and one activation layer. EAGB is mainly composed of several guidance blocks, which extract more reliable information through weight function to guide the learning of the network.*

tured by a small exposure time and a small ISO value [3]. The single image brightening algorithms can be classified into three categories. One category is model-based methods, such as low lighting image enhancement (LIME) [50], and algorithms in [3], [32]. Another category is deep learning based ones [30], [60]. The other is neural augmented methods which fuse both model-based and data-driven methods [29]. The difference between single image brightening and saturation restoration is that only the shadow regions are restored by the single image brightening while both the shadow and the highlight regions are restored by the saturation restoration. Noise amplification is a challenge for the former while both noise amplification and color distortion are challenges for the latter. Our human eyes are more sensitive to the color distortion in the bright regions. Intrinsic properties of differently exposed images are ignored by the DNN in [29] but they will be well utilized by the proposed framework.

III. Neural Augmented Saturation Restoration

Let Z_{T_1} be an eight-bit image. The key component of the proposed algorithm is to generate two high-quality 8-bit synthetic images Z_0 and Z_2 by combining a model-based method and a data-driven one. The image Z_0 is darker than the input Z_{T_1} to restore the highlight regions of Z_{T_1} while the image Z_2 is brighter than Z_{T_1} to restore the shadow regions of Z_{T_1}. The proposed framework is a neural augmentation [26], [31]. As indicated in [25], such a framework must posses the complete domain knowledge it requires. Thus, the CRFs are assumed to be known as in [30], [29], [31]. In other words, each algorithm is trained for each camera.

The ground-truth images of Z_0 and Z_2 are denoted as Z_{T_0} and Z_{T_2}, respectively. They are captured together with the

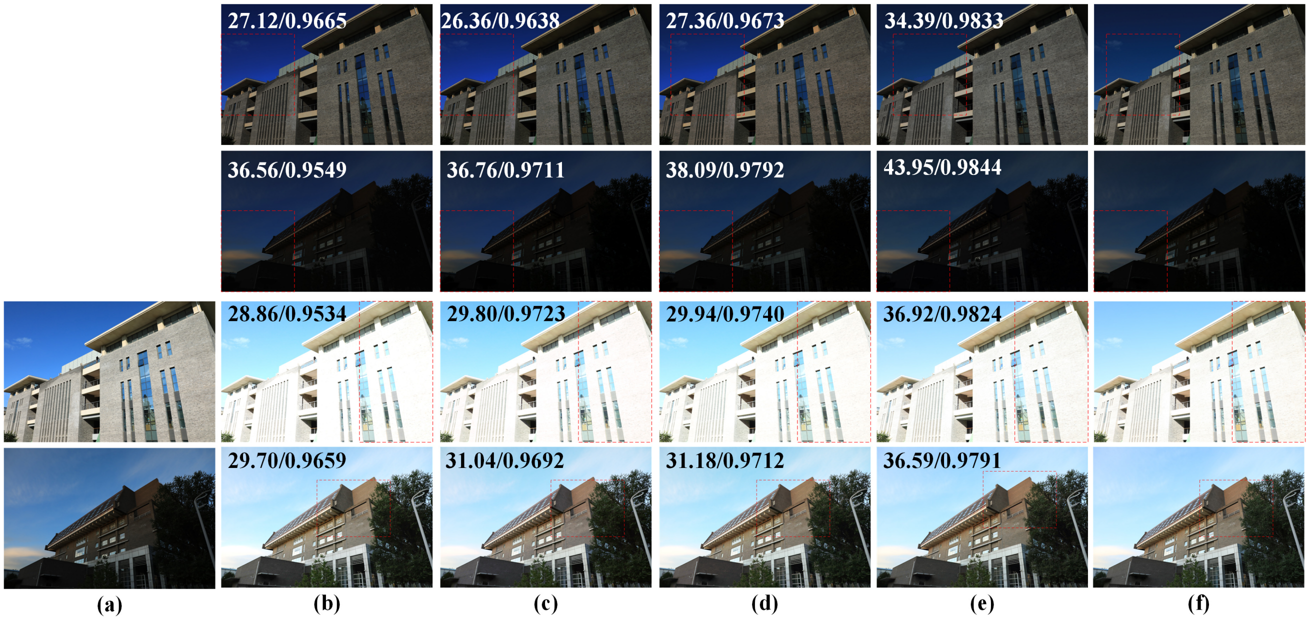


Fig. 3: *The first column includes two input LDR images which are taken with a Nikon 7200 camera. The ISO value is set as 800, and the exposure time Δt_1 is very short. The second column are the set of initial synthetic images $Z_0^{(0)}$'s and $Z_2^{(0)}$'s by the method [44], the exposure time are Δt_0 and Δt_2, $\Delta t_0 < \Delta t_1 < \Delta t_2$. The third column are the set of initial synthetic images $Z_0^{(0)}$'s and $Z_2^{(0)}$'s by the method [31], the fourth are the set of initial synthetic images $Z_0^{(0)}$'s and $Z_2^{(0)}$'s by the proposed IMF method, the fifth are the synthetic images by the proposed method, and the sixth column are the set of ground-truth images.*

image Z_{T_1} by using the method in [1]. Fig. 1 summarizes the pipeline of the proposed saturation restoration algorithm. The synthetic images are first generated by using the IMFs. They are then refined by one single EASRN which is on top of the RRG in [37] and the nonlocal block in [34]. Finally, the input image Z_{T_1} and two synthetic images Z_0 and Z_2 are fused together using the MEF algorithm in [42] or the HDR synthesis algorithm in [1] to produce the final image.

A. Model-Based Saturation Restoration

Two synthetic images $Z_0^{(0)}$ and $Z_2^{(0)}$ are first produced by using the IMFs [33], [13]. Let the CRF be $f_l(\cdot)(l \in \{R, G, B\})$, and the exposure times of Z_{T_0} and Z_{T_1} and Z_{T_2} be Δt_0, Δt_1 and Δt_2, respectively. Without loss of generality, it is assumed that $\Delta t_2 > \Delta t_1 > \Delta t_0$. It has been shown in [44] that there is relative brightness change in the fused image if the exposure ratios are too large. Thus, Δt_2 and Δt_0 are selected as $4\Delta t_1$ and $\Delta t_1/4$ so as to restore the information as much as possible. The IMF between the input image Z_{T_1} and the synthetic image $Z_i (i \in \{0, 2\})$ can be expressed as follows:

$$\Lambda_{1\to i,l}(z) = f_l(\frac{f_l^{-1}(z)\Delta t_i}{\Delta t_1}), \tag{1}$$

where $f_l^{-1}(\cdot)$ is the inverse function of the CRF $f_l(\cdot)$.

The images $Z_0^{(0)}$ and $Z_2^{(0)}$ are generated by using the IMFs as

$$Z_{i,l}^{(0)}(p) = \Lambda_{1\to i,l}(Z_{1,l}(p)). \tag{2}$$

p is the pixel position within the image. Normally, the IMF $\Lambda_{1\to i,l}(z)$ is computed as

$$\Lambda_{1\to i,l}(z) = \arg \min_{z' \in [0,255]} \{|\frac{f_l^{-1}(z)\Delta t_i}{\Delta t_1} - f_l^{-1}(z')|\}. \tag{3}$$

Here, the requirement that $\Lambda_{1\to i,l}(z)$ is an integer is relaxed as in Appendices 1 and 2. As shown in Table I, the images $Z_0^{(0)}$ and $Z_2^{(0)}$ are more accurate. This is very important for the proposed neural augmentation. The model-based method should be as simple as possible and it should be as accurate as possible [26]. This is because that the data-driven method is usually locally convergent. Such an elegant model-based method has a potential to guarantee that the neural augmentation outperforms both the model-based and the data-driven methods.

The resultant synthetic images are shown in Fig. 3. The synthetic images are close to the corresponding ground-truth images if the IMFs are reliable. However, due to limited representation capability of IMFs, the images $Z_0^{(0)}$ and $Z_2^{(0)}$ do not contain all the information in the ground-truth images, especially for those saturated regions. Both intensity and color need to be improved for the underexposed regions of the image $Z_2^{(0)}$ and the overexposed regions of the image $Z_0^{(0)}$.

B. Data-Driven Saturation Restoration

The IMFs can be regarded as a model to represent the correlation among different exposed images, but its representation capability is limited. The unmodelled information $(Z_{T_0} - Z_0^{(0)})$ and $(Z_{T_2} - Z_2^{(0)})$ can be further represented by a data-driven method. The unmodelled information is usually sparse,

TABLE I: *PSNR, SSIM and FSIM of Different IMF Estimation Methods*

IMF	[44]	[31]	the proposed
$\Lambda_{1\to 0}(\cdot)$	28.39 / 0.9392 / 0.9743	28.68 / 0.9389 / 0.9752	29.61 / 0.9426 / 0.9756
$\Lambda_{1\to 2}(\cdot)$	30.09 / 0.9602 / 0.9645	31.81 / 0.9638 / 0.9713	31.61 / 0.9642 / 0.9725

i.e., most values are likely to be zero or small. It can be easily *compensated* by the data-driven method. Subsequently, the saturated regions are further restored. It is worth noting that the modelled information and un-modelled information are analogous to the modelled dynamics and un-modelled dynamics in the field of nonlinear systems [39], [40]. The proposed framework can be explained by leveraging conventional knowledge on nonlinear systems [39], [40]. It should be mentioned that the data-driven approach also becomes a model after it has been trained. There are usually remaining unmodelled dynamics ever after the further representation by the data-driven approach.

One single EASRN is trained to refine the images $Z_0^{(0)}$ and $Z_2^{(0)}$ locally and non-locally as

$$Z_i = Z_i^{(0)} + \text{EASRN}(Z_{T_1}, Z_i^{(0)}, M_i)\ ;\ i = 0, 2, \tag{4}$$

where M_i is the mask for the refinement of the image $Z_i^{(0)}$. This is different from the brightening algorithms in [29] in the senses that two different DNNs are trained to respectively refine the two virtual images and the shadow regions are restored by only using the information from the neighboring areas. Details on the proposed EASRN are presented as below.

As shown in Figs. 1 and 2, the proposed EASRN consists of EAGB and NRRG. The EAGB can help the NRRG pay more attention to saturated areas, which can guide the saturation restoration more effectively. Two binary masks $M_{0,l}$ and $M_{2,l}$ are defined for the image $Z_{T_1,l}$ as

$$M_{0,l}(p) = \begin{cases} 0; & \text{if } Z_{T_1,l}(p) \geq \xi_U \\ 1; & \text{otherwise} \end{cases}, \tag{5}$$

$$M_{2,l}(p) = \begin{cases} 0; & \text{if } Z_{T_1,l}(p) \leq \xi_L \\ 1; & \text{otherwise} \end{cases}. \tag{6}$$

The different masks $M_{0,l}$ and $M_{2,l}$ are defined by using the following intrinsic properties of differently exposed images [9], [10]:

1) The IMF $\Lambda_{1\to 0,l}(\cdot)$ is not reliable for all pixels in the overexposed regions of the image $Z_{T_1,l}$; and 2) The IMF $\Lambda_{1\to 2,l}(\cdot)$ is not reliable for all pixels in the underexposed regions of the image $Z_{T_1,l}$.

Obviously, the masks M's are derived from intrinsic properties of differently exposed images, and they are different from the soft mask in [20]. It should be pointed out that the intrinsic properties of differently exposed images are ignored by the brightening algorithms in [29].

The masks (5) and (6) are utilized to design the EAGB. They are first multiplied with the synthetic images to filter out the poor information from the ill-posed regions. The features of the EAGB are embedded into the NRRG in a multilevel manner to realize deep guidance. Subsequently, the proposed EASRN can pay necessary attention to the saturated areas for the saturation restoration. Clearly, the interpretability of the EASRN is improved with the proposed EAGB.

The function of the NRRG is to globally suppress the less useful features and only allow the propagation of more informative ones. The NRRG contains multiple non-local dual modules (NDMs) [34], [35] and each NDM performs both non-local spatial and channel operations. Besides the structure of the proposed network, loss functions also play an important role in the data-driven approach. The proposed overall loss function is defined as

$$L = L_r + w_c L_c + w_f L_f, \tag{7}$$

where w_c and w_f are constant, they are selected as 0.01 and 0.01 in this study, respectively.

Since the l_1 norm usually outperforms the l_2 norm for image restoration [36], the loss function L_r is given as

$$L_r = \sum_{p,l} \Psi_{i,l}(Z_{T_1,l}(p)) |Z_{T_i,l}(p) - Z_{i,l}(p)|, \tag{8}$$

where the weight functions $\Psi_{i,l}(z) (i = 0, 2)$ are defined as

$$\Psi_{0,l}(z) = \begin{cases} 1; & \text{if } z \leq \xi_U \\ \frac{1}{z - \xi_U}; & \text{otherwise} \end{cases}, \tag{9}$$

$$\Psi_{2,l}(z) = \begin{cases} 1; & \text{if } z \geq \xi_L \\ \frac{1}{\xi_L - z}; & \text{otherwise} \end{cases}. \tag{10}$$

As pointed out in [9], [10], color distortion is an issue when the underexposed (overexposed) regions of Z_{T_1} are enhanced by the image Z_0 (Z_2). To minimize the possible color distortion in the two synthetic images, the color loss is defined as

$$L_c = \sum_p \angle(Z_{T_i}(p), Z_i(p)), i \in \{0, 2\} \tag{11}$$

where $\angle(Z_{T_i}(p), Z_i(p))$ is the angle between two 3D (R, G, B) vectors $Z_{T_i}(p)$ and $Z_i(p)$. Since the L_r metric only measures the color difference numerically, it cannot ensure that the color vectors have the same direction [60]. By employing the color loss in Eq. (11), the possible color distortion is reduced, especially for the saturated regions.

The feature-wise loss L_f is defined as

$$L_f = \frac{\sum_{l=1}^{W_{k,j}} \sum_{m=1}^{H_{k,j}} (\phi_{k,j}(Z_{T_i})_{l,m} - \phi_{k,j}(Z_i)_{l,m})^2}{W_{k,j} H_{k,j}}, i \in \{0, 2\} \tag{12}$$

where $W_{k,j}$ and $H_{k,j}$ denote the dimensions of the respective feature maps within the VGG network. $\phi_{k,j}(\cdot)$ is the feature map obtained by the j-th convolution (before activation) before the k-th maxpooling layer within the VGG network. The fine-tuned VGG network for material recognition in [61] is adopted to define the feature-wise loss. The VGG in [61] focuses on textures which is critical for the refinement of the synthetic images.

Clearly, the model-based and data-driven methods *compensate* each other well in the proposed algorithm which has the following advantages:

1) Compared with the model-based method, the final images $Z_i(i=0,2)$ are much closer to the ground-truth images than the initial images $Z_i^{(0)}(i=0,2)$.

2) Compared with the data-driven solution, the proposed framework converges fast and requires fewer training samples, because $(Z_{T_i}-Z_i^{(0)})(i=0,2)$ are sparser than $Z_{T_i}(i=0,2)$.

3) With the masks $M_i(i=0,2)$, the proposed EAGB can assist the network to pay more attention to the saturated regions for the efficient restoration.

And 4) only one single EASRN is trained to refine the two virtual images $Z_i^{(0)}(i=0,2)$ as in the Eq. (4).

C. Combination of Input Image and Two Synthetic Images

The input image $Z_1(=Z_{T_1})$ and the two synthetic images Z_0 and Z_2 can be combined together by using the HDR synthesis algorithm in [1] to produce an HDR image or by using the MEF algorithm in [42] to produce an LDR image. The latter is focused on in this subsection.

As shown in [42], the weighting maps of the differently exposed images play an important role in the MEF algorithm. This subsection focuses on defining the weighting maps for the three images. One function is used to determine amplification factors of all the pixels in the image Z_1. The weight $\psi_1(z)$ is defined as 2. The other is to measure contrast, well exposedness level, and color saturation of each pixel in the three images and it is defined the same as in [42]

$$\psi_2(Z_i(p))=w_c(Z_i(p))\times w_s(Z_i(p))\times w_e(Z_i(p)), \quad (13)$$

where w_c is computed by filtering the gray-scale version of each image via a Laplacian filter. w_s is the standard deviation among the R, G and B channels. w_e is derived by first using a Gauss curve with the sigma as 0.2 on each color channel separately and then multiplying the results.

Let Y_1 be the luminance component of the image Z_1 in YUV color space. The weight of the pixel $Z_1(p)$ is given as $\psi_1(Y_1(p))\psi_2(Z_1(p))$, and the weight of the pixel $Z_i(p)(i=0,2)$ is given as $\psi_2(Z_i(p))$. With the above weighting maps, all the images $Z_i(i=0,1,2)$ are fused together via the state-of-the-art MEF algorithm in [42]. The number of pyramid levels is increased by one to reduce the halo artifacts. The MEF-SSIM is also slightly improved due to the small size of the images.

IV. Experimental Results

Extensive experiments are carried out in this section to demonstrate the rationality and effectiveness of the proposed neural augmentation framework. The emphasis is to show how the model-based method and the data-driven one *compensate* each other. Readers are invited to view to the electronic version of the full-size figures and zoom in these figures in order to better appreciate the differences among images.

A. Dataset

The exposure dataset is constructed from two existing datasets on HDR imaging in [28], [29]. All the images are captured by two different cameras which are Nikon and Canon. Only exposure times are changed. Camera shaking and object movement are strictly controlled so as to mimic images captured by the new device [49]. The interval between the two adjacently exposed images is 2-EV. Thus, the interval of two inputs is 4-EV. The dataset is randomly split into three parts: 650 sequences for training, 50 ones for verifying, and the rest 100 ones for testing.

Firstly, the CRF is calculated using the algorithm in [1], then the initial virtual images are generated by the proposed IMF constructed using the CRF. The ξ_U and ξ_L are set to 250 and 5 respectively according to [29]. The number of NRRGs and guidance blocks are set as 4, and each NRRG contains 4 NDMs. We randomly crop a $128*128$ patch from each input during training. The proposed network is trained using the proposed loss functions and Adam optimizer with $\beta_1=0.9$ and $\beta_2=0.99$. We set the batch size to 8. The learning rate is initially set to 10^{-5} because the initial virtual images are already very close to the ground truth, and then decreased using a cosine annealing schedule. All the experiments are implemented using PyTorch on NVIDIA GP100 GPUs.

B. Evaluation of The EAGB

The features from the EAGB are embedded into the NRRG to guide the NRRG. In order to validate the effectiveness of the EAGB, the two modes with and without the EAGB are compared as shown in Fig. 5. The mode with the EAGB is more stable than the mode without EAGB, and can achieve higher PSNR values in different epoches. Clearly, the EAGB indeed guides the NRRG to exploit reliable information for restoring the synthetic images. The intrinsic properties of differently exposed images are indeed helpful for the data-driven approach.

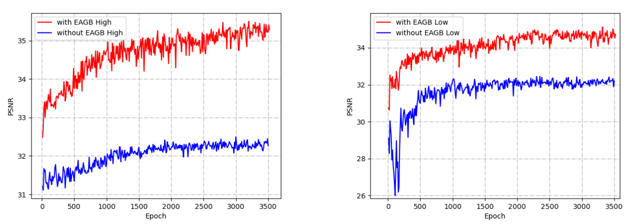
C. Ablation Study of Different Methods

The proposed method is compared with the model-based method and the data-driven one in Fig. 4 as well as the method in [29]. The structure of the data-driven method is the same as the proposed method except the EAGB which requires the initial synthetic images. The average structural similarity index (SSIM) and Peak Signal-to-Noise Ratio (PSNR) for the 100 sets of test images are shown in Table. II. The proposed method can improve the SSIM, PSNR, and the visual quality

Fig. 4: *The diagram of a data-driven saturation restoration algorithm. Two synthetic images Z_0 and Z_2 with Δt_0 and Δt_2 are generated directly by data-driven method. The input image and the two synthetic images are finally fused to obtain an information enriched LDR image.*

TABLE II: *SSIM and PSNR of synthetic images generated by different methods.*

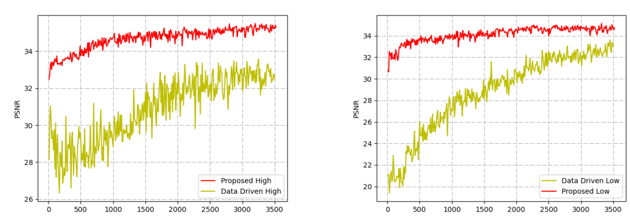
	SSIM		PSNR		FSIM	
	Low	Hig	Low	High	Low	High
model-based	0.9426	0.9642	29.61	31.61	0.9756	0.9725
Hybrid learning in [29]	0.9476	0.9679	32.21	33.56	0.9759	0.9792
data-driven in Fig. 4	**0.9589**	0.9735	33.58	33.58	0.9800	**0.9796**
proposed (no EAGB)	0.9384	0.9651	31.85	32.49	0.9787	0.9759
proposed	0.9575	**0.9738**	**34.85**	**35.49**	**0.9805**	0.9779



(a) (b)

Fig. 5: *(a) Comparisons of PSNR between the model with and without EAGB for high exposure images. (b) Comparisons of PSNR between the model with and without EAGB for low exposure images. The red curve represents the model with EAGB, and the blue curve represents the model without EAGB. The model with EAGB can achieve higher PSNR for both high-exposure and low-exposure images.*

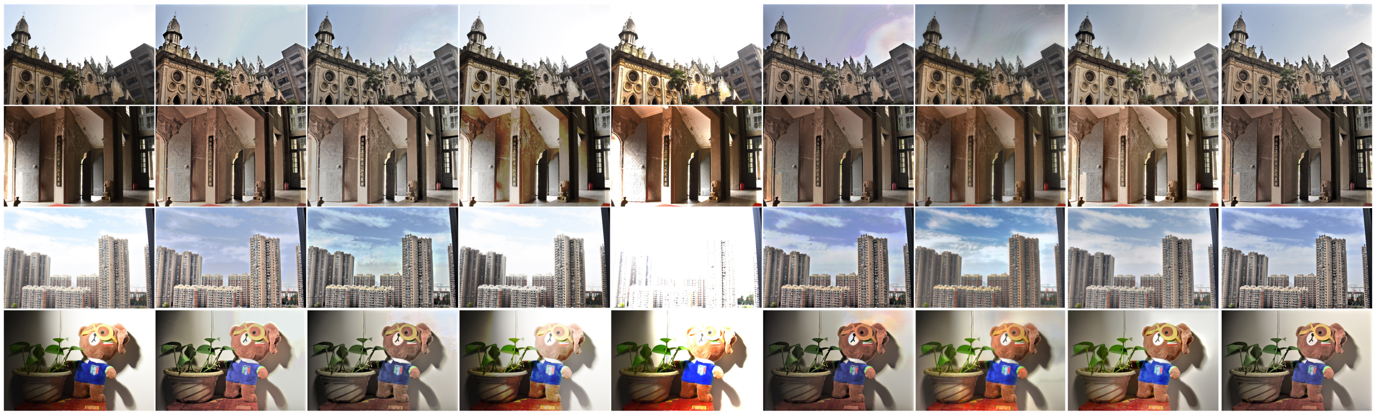
of the synthetic images. Clearly, the proposed method can make the synthetic images much closer to the ground-truth images. The results of MEF-SSIM [59] are also shown in Table. III. The proposed algorithm has been tested on 100 sets of data. To save space for display, Tables III,IV, V, VI and VII shows the test results on 20 randomly selected sets. The quality of the finally fused images is also improved. The proposed EASRN outperforms the DNN in [29]. This also shows that the intrinsic properties of differently exposed images are important for the data-driven method on the saturation restoration. In addition, the EAGB plays an important role in guiding the proposed network by extracting reliable information. The model with the EAGB can achieve higher PSNR as shown in Fig. 5 and higher MEF-SSIM values as illustrated in Table III.



(a) (b)

Fig. 6: *(a) Comparisons of PSNR between the proposed framework and data-driven method for high exposure images. (b) Comparisons of PSNR between the proposed framework and data-driven method for low exposure images. The red curve represents the proposed framework, and the yellow curve represents the data-driven method. The proposed framework is more stable than the data-driven method, and can also obtain higher PSNR.*

The convergence speed is also analyzed. There is large difference between the input image Z_1 and the ground-truth low/high exposure images Z_{T_i}. The low exposure images



(a) (b) (c) (d) (e) (f) (g) (h) (i)

Fig. 7: *The first column shows the LDR images, the second to seventh columns illustrate the results of (b) ExpandNet [18], (c) DrTMO [22], (d) Histogram HDR [57], (e) IRA HDR [58], (f) SingleHDR [19], (g) MEG [23], and (h) the proposed algorithm, respectively, as well as the ground truth LDR images which are fused by three captured images with different exposures*

.

TABLE III: *MEF-SSIM of Five Different Methods*

image	model -based	Hybrid learning	data -driven	proposed (no EAGB)	proposed
set1	0.9544	0.9476	0.9540	**0.9576**	0.9565
set2	0.9567	0.9569	0.9590	0.9592	**0.9614**
set3	0.9249	0.9424	**0.9485**	0.9483	0.9484
set4	0.9514	0.9417	0.9526	0.9563	**0.9558**
set5	0.9158	0.9155	0.9156	0.9154	**0.9165**
set6	0.9382	0.9387	0.9372	0.9404	**0.9409**
set7	**0.9758**	0.9732	0.9744	0.9740	0.9736
set8	0.9275	0.9403	**0.9478**	0.9457	0.9452
set9	0.8946	**0.9138**	0.9073	0.8979	0.9045
set10	0.9255	0.9377	**0.9510**	0.9467	0.9483
set11	0.9844	0.9837	0.9846	**0.9849**	0.9835
set12	0.8848	0.8892	**0.9026**	0.8870	0.8851
set13	**0.9830**	0.9798	0.9803	0.9820	0.9814
set14	0.9565	0.9563	**0.9606**	0.9592	0.9591
set15	0.9434	0.9401	**0.9440**	0.9424	0.9437
set16	0.9853	0.9844	0.9848	**0.9862**	0.9855
set17	0.9220	0.9375	0.9375	0.9399	**0.9431**
set18	**0.9056**	0.8867	0.8702	0.9002	0.8963
set19	0.9073	0.9130	**0.9225**	0.9161	0.9165
set20	**0.9447**	0.9396	0.9332	0.9410	0.9387
average	0.9391	0.9409	0.9443	0.9440	**0.9442**

contain random noise, which is difficult to learn for the input image Z_1. On the other hand, the synthetic images Z_i which are generated by the model-based method are already close to Z_{T_i}. It is easy for the residual CNNs to represent $(Z_{T_i} - Z_i)$, and the networks can converge much faster as shown in Fig. 6. Clearly, the model-based method and the data-driven one indeed *compensate* each other in the proposed framework.

D. Comparison of Six Different Algorithms

In this subsection, the proposed saturation restoration algorithm is compared with three state-of-the-art (SOTA) data-driven algorithms including ExpandNet [18], DrTMO [22], and SingleHDR [19], as well as two SOTA model-based algorithms including Histogram HDR [57] and IRA HDR [58]. HDR images are generated by these three data-driven algorithms. The proposed algorithm and the data-driven three algorithms are compared by using the mean squared error (MSE), the latest HDR-VDP-3, average PU2-SSIM and PU2-PSNR of HDR images [62]. The ground-truth HDR image is synthesized from the three differently exposed images Z_{T_0}, Z_1, and Z_{T_2} by using the method in [1] and the CRFs $F_c(\cdot)$. The HDR image of the proposed algorithm is generated from Z_0, Z_1, Z_2, and the CRFs $F_c(\cdot)$. LDR images are produced by the two model-based algorithms. The proposed algorithm and the model-based algorithms are compared by using the MEF-SSIM with the reference images as Z_{T_0}, Z_1, and Z_{T_2}. It is shown in Tables IV, V, VI, and VII that the proposed algorithm usually outperforms the four data-driven algorithms from the MSE, HDR-VDP-2, PU2-SSIM, and PU2-PSNR points of view and the two model-based algorithms from the MEF-SSIM point of view. The model-driven algorithms have limited representation capability and cannot represent complex lighting environments well. Data-driven algorithms have stronger representation capability, but they rely heavily on data and have the risk of overfitting. The proposed algorithm adopts a complementary form of model-based and data-driven methods, reducing the dependence on data and enhancing the robustness of the algorithm.

Fig. 7 includes the final images that are generated by the seven algorithms. The pseudo-HDR images of the former three are tone mapped by using the state-of-the-art algorithm in [54]. The proposed algorithm restores saturated regions better than those algorithms in [18] [19] [22] [23] [57] [58] as shown in the sky regions of the first and third rows. The algorithm in [18] and the proposed algorithm produces slight color distortion. There are also visible color distortions in the highlighted region by the algorithms in [19], [22] and [23] as shown in the last row. The images generated by the algorithm in [22] and the proposed one look more natural and realistic.

All the algorithms in [18], [22], [19], [23] and the proposed algorithm include deep neural networks. Their sizes are 1.8 MB, (Up: 647.8 MB, Down: 648.4 MB), 804.3 MB, 821.7 MB, and 20.6MB, respectively. The proposed EA-Net is larger than those DNN in [18] but smaller than those DNNs in [22],

TABLE IV: *MEF-SSIM of three Different Algorithms*

image	Histogram HDR [57]	IRA HDR [58]	Proposed
set1	0.9246	0.8479	**0.9565**
set2	0.8855	0.5773	**0.9614**
set3	0.7945	0.5956	**0.9484**
set4	0.9021	0.8182	**0.9558**
set5	0.8695	0.7522	**0.9165**
set6	0.8921	0.7973	**0.9409**
set7	0.9300	0.8972	**0.9736**
set8	0.8710	0.7244	**0.9452**
set9	0.8712	0.8467	**0.9045**
set10	0.8473	0.7379	**0.9483**
set11	0.9338	0.9094	**0.9835**
set12	0.8609	0.8084	**0.8851**
set13	0.8730	0.7948	**0.9814**
set14	0.9401	0.8440	**0.9591**
set15	0.9305	0.9033	**0.9437**
set16	0.9545	0.7676	**0.9855**
set17	0.8917	0.8225	**0.9431**
set18	0.8549	0.5026	**0.8963**
set19	0.8668	0.8402	**0.9165**
set20	0.9119	0.7102	**0.9387**
average	0.8903	0.7749	**0.9442**

TABLE V: *MSE of Four Different Algorithms*

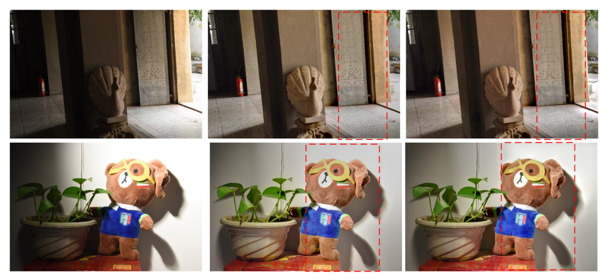
image	ExpandNet [18]	DrTMO [22]	SingleHDR [19]	MEG [23]	Proposed
set1	**0.0025**	0.0141	0.0243	0.0077	0.0158
set2	0.0615	0.0044	0.0495	**0.0061**	0.0124
set3	0.0082	0.0130	0.0481	0.0168	**0.0074**
set4	0.0151	0.0121	0.0978	0.0108	**0.0092**
set5	0.0125	0.0295	$\mathbf{3.6}\times10^{-4}$	0.0016	5.1×10^{-4}
set6	0.0038	0.0286	0.0149	0.0030	**0.0023**
set7	0.0421	0.0505	0.0216	0.0576	**0.0023**
set8	0.0143	0.0092	0.0443	**0.0065**	0.0069
set9	0.0137	0.0104	0.0098	**0.0055**	0.0059
set10	**0.0108**	0.0282	0.0673	0.0097	**0.0092**
set11	0.0242	0.0428	0.0826	0.0062	**0.0018**
set12	0.0133	0.0306	0.0081	0.0035	**0.0017**
set13	0.2199	0.2020	0.0097	**0.0088**	0.0548
set14	0.0439	0.0475	**0.0044**	0.0056	0.0164
set15	0.0700	0.0094	**0.0057**	0.0081	0.0344
set16	0.0618	0.1055	0.0054	0.0096	**0.0052**
set17	0.0121	0.0223	0.0580	**0.0082**	0.0104
set18	$\mathbf{2.4}\times10^{-4}$	0.0013	7.2×10^{-4}	0.0013	3.0×10^{-4}
set19	0.0066	0.0161	0.0310	0.0070	**0.0042**
set20	0.0152	0.0113	$\mathbf{3.5}\times10^{-4}$	0.0036	7.5×10^{-4}

TABLE VI: *HDR-VDP-3 of Four Different Algorithms*

image	ExpandNet [18]	DrTMO [22]	SingleHDR [19]	MEG [23]	Proposed
set1	8.21	8.79	7.47	8.51	**8.91**
set2	8.17	8.54	8.39	7.98	**8.94**
set3	7.89	7.77	6.02	7.48	**8.54**
set4	7.62	**8.01**	5.78	5.51	7.74
set5	8.01	7.88	9.60	8.71	**9.77**
set6	8.30	7.14	7.50	8.17	**9.49**
set7	8.26	8.96	9.26	6.75	**9.44**
set8	7.62	8.41	7.52	8.35	**9.03**
set9	7.02	7.60	8.34	7.78	**8.50**
set10	7.96	7.42	6.47	8.15	**8.94**
set11	9.01	8.99	8.19	8.34	**9.52**
set12	7.69	7.53	7.16	7.59	**9.11**
set13	7.23	8.07	9.03	5.52	**8.78**
set14	7.44	8.21	9.01	8.29	**8.90**
set15	7.57	8.45	8.86	6.54	**8.70**
set16	7.33	7.71	7.82	7.95	**9.27**
set17	7.84	7.54	6.81	5.52	**8.84**
set18	8.41	8.14	7.74	7.18	**9.88**
set19	7.70	7.43	7.06	7.65	**8.82**
set20	6.85	8.25	**9.65**	8.34	9.58

TABLE VII: *Average PU2-SSIM and PU2-PSNR of Four different Methods.*

	PU2-SSIM	PU2-PSNR
ExpandNet [18]	0.9109	38.05
DrTMO [22]	0.8538	35.29
SingleHDR [19]	0.9147	34.91
MEG [23]	**0.9475**	40.09
Proposed	0.9404	**43.28**



(a) (b) (c)

Fig. 8: *Comparison of the information enriched LDR images by using inaccurate CRFs and accurate CRFs. (a) shows the input images; (b) illustrates the results by using inaccurate CRFs; (c) demonstrates the images by using accurate CRFs.*

[19], [23]. The running times of these five different algorithms are given in Table VIII. The proposed algorithm is ranked the third. All the DNNs can be pruned in real applications.

E. Limitation of the Proposed Algorithm

Although the proposed algorithm outperforms the other six algorithms, there is space for further study. For example, the proposed algorithm assumes that the accurate CRFs are available. For images from unknown sources, it is quite challenging to estimate the CRFs. If the estimated CRFs are not accurate, the difference between the synthetic images and their ground-truth images become large, and the quality of the brightened images drops a little bit, as shown in Fig. 8.

V. Conclusion Remarks and Discussions

A new framework is introduced to restore saturated regions for a low dynamic range (LDR) image of a high dynamic range (HDR) scene. Two synthetic images are first generated by using the model-based method, and are then enhanced by using the data-driven method. They *compensate* each other very well in such a neural augmentation. The proposed framework has a potential to achieve learning with few data. All the input image and the two synthetic images are finally combined together to produce an LDR or HDR image. The proposed algorithm is also applicable to study saturation restoration of LDR videos. Flicking artifacts could be well addressed because of the model-based method. It is important to produce an information enriched panoramic LDR image without visual artifacts for an HDR scene through stitching multiple geometrically synchronized LDR images with different exposures and pairwise overlapping fields of views. One more interesting problem is to study joint HDR and 3D imaging [63]. Most existing works on monocular and stereo depth estimation [63], [64], depth completion, structure from motion ignore possible saturations

TABLE VIII: *Average Running Times of Five Different Algorithms including Deep Neural Networks*

Methods	ExpandNet [18]	DrTMO [22]	SingleHDR [19]	MEG [23]	Ours
Times	**0.2644**	4.3771	0.5586	1.5921	1.2454

in input image(s). Unfortunately, this is not always true in real-world applications. All these problems will be studied in our future research.

Appendix 1: Matlab Implementation of $\Lambda_{1\to 0,l}(\cdot)$

```
%%% Initialization
Λ_{1→0,l} = zeros(1,256);
Λ_{1→0,l}(0) = 0;
Δ = 1/8;%%% Step size
%%% Iteration
for i=1:255
    a = (Δt_0/Δt_1) f_l^{-1}(i);
    z_2 = Λ_{1→0,l}(i-1);
    while(f_l^{-1}(z_2) < a)
        z_2 = z_2 + Δ;
    end
    z_1 = z_2 - Δ;
    if 2a ≤ f_l^{-1}(z_1) + f_l^{-1}(z_2)
        Λ_{1→0,l}(i) = z_1;
    else
        Λ_{1→0,l}(i) = z_2;
    end
end
```

Appendix 2: Matlab Implementation of $\Lambda_{1\to 2,l}(\cdot)$

```
%%% Initialization
Λ_{1→2,l} = zeros(1,256);
Λ_{1→2,l}(255) = 255;
Δ = 1/8;%%% Step size
%%% Iteration
for i=254:0
    a = (Δt_2/Δt_1) f_l^{-1}(i);
    z_1 = Λ_{1→2,l}(i+1);
    while(f_l^{-1}(z_1) > a)
        z_1 = z_1 - Δ;
    end
    z_2 = z_1 + Δ;
    if 2a ≤ f_l^{-1}(z_1) + f_l^{-1}(z_2)
        Λ_{1→2,l}(i) = z_1;
    else
        Λ_{1→2,l}(i) = z_2;
    end
end
```

Acknowledgement

This work was supported in part by A*STAR Singapore under Robotics Horizontal Technology Coordinating Office (HTCO) Project C221518005 and Nature Science Foundation of Hubei Province, China (Grant No. 2022CFB676).

References

[1] P. E. Debevec and J. Malik, "Recovering high dynamic range radiance maps from photographs," In *Proceedings of SIGGRAPH*, pp. 369-378, USA, May 1997.

[2] T. Mertens, J. Kautz, and F. V. Reeth, "Exposure fusion: a simple and practical alternative to high dynamic range photography," Computer Graphics Forum, vol. 28, no. 1, pp. 161-171, Jan. 2009.

[3] Z. G. Li, Z. Wei, C. Y. Wen, and J. H. Zheng, "Detail-enhanced multi-scale exposure fusion," IEEE Trans. on Image Processing, vol. 26, no. 3, pp. 1243-1252, Mar. 2017.

[4] A. R. Varkonyi-Koczy, A. Rovid and T. Hashimoto, "Gradient-Based Synthesized Multiple Exposure Time Color HDR Image," IEEE Transactions on Instrumentation and Measurement, vol. 57, no. 8, pp. 1779-1785, Aug. 2008.

[5] X Tan, H. A. Chen, K. XU, et al. "High Dynamic Range Imaging for Dynamic Scenes With Large-Scale Motions and Severe Saturation," IEEE Transactions on Instrumentation and Measurement, vol. 71, pp. 1-15, 2022.

[6] S. Wu, Z. Li, J. Zheng, and Z. Zhu, "Exposure robust method for aligning differently exposed images," IEEE Signal Processing Letter, vol. 21, no. 7, pp. 885-889, Jul. 2014.

[7] J. Hu, O. Gallo, and K. Pulli, "Exposure stacks of live scene with hand-held cameras," In *Proceedings of the European Conference of Computer Vision*, pp.499-512, Italy, Oct. 2012.

[8] P. Sen, N. Khademi Kalantari, M. Yaesoubi, S. Darabi, D. Goldman, and E. Shechtman, "Robust patch-based HDR reconstruction of dynamic scenes," ACM Trans. on Graphics, vol. 31, no. 6, Art. 203, Nov. 2012.

[9] J. H. Zheng, Z. G. Li, Z. J. Zhu, S. Q. Wu, and S. Rahardja, "Hybrid patching for a sequence of differently exposed images with moving objects," IEEE Trans. on Image Processing, vol. 22, no. 12, pp.5190-5201, Dec. 2013.

[10] Z. G. Li, J. H. Zheng, Z. J. Zhu, and S. Q. Wu, "Selectively detailenhanced fusion of differently exposed images with moving objects," IEEE Trans. on Image Processing, vol. 23, no. 10, pp. 4372-4382, Oct. 2014.

[11] N. K. Kalantari and R. Ramamoorthi, "Deep high dynamic range imaging of dynamic scenes," ACM Trans. on Graphics, vol. 36, no. 4, pp. 1-12, Jul. 2017.

[12] S. Wu, J. Xu, Y. W. Tai, and C. K. Tang, "Deep high dynamic range imaging with large foreground motions," in *European Conference on Computer Vision (ECCV)*, 2018.

[13] Z. Y. Liu, Z. G. Li, Z. Liu, X. M. Wu and W. H. Chen, "Unsupervised optical flow estimation for differently exposed images in LDR domain," IEEE Trans. on Circuits and Systems for Video Technology, DOI: 10.1109/TCSVT.2023.3252007, 2023.

[14] H. Li, K. Ma, H. Yong, and L. Zhang, "Fast multi-scale structure patch decomposition for multi-exposure image fusion," IEEE Trans. on Image Processing, vol. 29, pp.5808-5816, Jan. 2020.

[15] X. Tan, H. Chen, K. Xu, C. M. Xu, Y. Jin, C. G. ZHu, abd J. J. Zheng, "High Dynamic Range Imaging for Dynamic Scenes With Large-Scale Motions and Severe Saturation," IEEE Transactions on Instrumentation and Measurement, vol. 71, pp. 1-15, Aug. 2022.

[16] M. D. Tocci, C. Kiser, N. Tocci, and P. Sen, "A versatile HDR video production system," in *Proc. SIGGRAPH*, pp. 1-9, USA, 2011.

[17] J. Gu, Y. Hitomi, T. Mitsunaga, and S. Nayar, "Coded rolling shutter photography: flexible space-time sampling," In *IEEE International Conference on Computational Photography*, pp. 1-8, USA, 2010.

[18] D. Marnerides, T. Bashford-Rogers, J. Hatchett, and K. Debattista, "ExpandNet: A deep convolutional neural network for high dynamic range expansion from low dynamic range content," Computer Graphics Forum, vol. 37, no. 2, pp. 37-49, Mar. 2018.

[19] Y. L. Liu, W. S. Lai, Y. S. Chen, Y. L. Kao, M. H. Yang, Y.Y. Chuang and J. B. Huang, "Single image HDR reconstruction by learning to reverse the camera pipeline," in *IEEE Conference on Computer Vision and Pattern Recognition*, Jun. 2020.

[20] M. Santos, I. R. Tsang, and K. K. Nima, "Single image HDR reconstruction using a CNN with masked features and perceptual loss," ACM Trans. on Graphics, vol. 39, no. 4, 80-1, Aug. 2020.

[21] H. Cui, J. Li, Z. Hua and L. Fan, "Progressive Dual-Branch Network for Low-Light Image Enhancement," IEEE Transactions on Instrumentation and Measurement, vol. 71, pp. 1-18, Oct. 2022.

[22] Y. Endo, Y. Kanamori, and J. Mitani, "Deep reverse tone mapping," ACM Transactions on Graphics, vol. 36, no. 6, pp. 1-10, Nov. 2017.
[23] P. H. Le, Q. Le, R. Nguyen, B. S. Hua, "Single-Image HDR Reconstruction by Multi-Exposure Generation," in *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, January, 2023.
[24] J. Kim, S. Lee, and S.Kang, "End-to-end differentiable learning to HDR image synthesis for multi-exposure Images," in *AAAI, 2021*.
[25] N. Shlezinger, J. Whang, Y. Eldar, and A. Dimakis, "Model-based deep learning," arxiv Preprint arXiv: 2012.08405v2, 27 Jun 2021.
[26] Z. G. Li, C. B. Zheng, H. Y. Shu, and S. Q. Wu, "Dual-scale single image dehazing via neural augmentation," IEEE Trans. on Image Processing, vol. 31, pp. 6213-6223, Sept 2022.
[27] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 770-778.
[28] Y. Xu, Z. Liu, X. Wu, W. Chen, C. Wen, and Z. Li, "Deep joint demosaicing and high dynamic range imaging within a single shot," IEEE Trans. on Circuits and Systems for Video Technology, vol. 32, no. 8, pp.1-16, Aug. 2022.
[29] C. B. Zheng, Z. G. Li, Y. Yang, and S. Q. Wu, "Single image brightening via multi-scale exposure fusion with hybrid learning," IEEE Trans. on Circuits and Systems for Video Technology, vol. 31, no. 4, pp.1425-1435, Apr. 2021.
[30] J. Xu, C. Chen, Q. Chen, and V. Koltun, "Learning to see in the dark," in *IEEE Conference on Computer Vision and Pattern Recognition*, Jun. 2018.
[31] C. Zheng, W. Jia, S. Wu, and Z. Li, "Neural augmented exposure interpolation for two large-exposure-ratio images", IEEE Trans. on Consumer Electronics, vol. 69, no. 1, pp. 87-97, Feb. 2023.
[32] Y. Ren, Z. Ying, T. Li, and G. Li, "LECARM: Low-light image enhancement using camera response model," IEEE Trans. on Circuits and Systems for Video Technology, vol. 29, no. 4, pp. 968-981, Apr. 2019.
[33] Y. Xu, Z. Li, W. Chen, and C. Wen, "Novel intensity mapping functions: weighted histogram averaging," in *the 17th IEEE Conference on Industrial Electronics and Applications*, pp. 1157-1161, China, Dec. 2022.
[34] X. Wang, R. Girshick, A. Gupta, and K. He, "Non-local neural networks," in *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 7794-7803, Jun. 2018.
[35] J. Fu, J. Liu, H. Tian, Y. Li, T. Bao, Z. Fang, and H. Lu, "Dual attention network for scene segmentation," in *IEEE Conference on Computer Vision and Pattern Recognition*, Jun. 2019.
[36] H. Zhao, O. Gallo, I. Frosio, J. Kautz, "Loss Functions for Image Restoration With Neural Networks," IEEE Trans. on Computational Imaging, vol. 3, no. 1, pp.47-57, Mac. 2017.
[37] S. Wa. Zamir, A. Arora, S. Khan, M. Hayat, F. S. Khan, M. Yang, and L. Shao "CycleISP: real image restoration via improved data synthesis," in *IEEE Conference on Computer Vision and Pattern Recognition*, Jun. 2020.
[38] A. Buades, B. Coll, and J.-M. Morel, "A non-local algorithm for image denoising," in *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 60-65, Jun. 2005.
[39] Z. Li, Y. C. Soh, and C. Wen, "Robust stability of a class of hybrid nonlinear systems," IEEE Trans. on Automatic Control, vol. 46, no. 6, pp.897-903, Jun. 2001.
[40] H. K. Khalil and J. W. Grizzle, Nonlinear system, Prentice Hall, 2002.
[41] C. O. Ancuti, C. Ancuti, C. De Vleeschouwer, and Alan C. Bovik, "Single-scale fusion: an effective approach to merging images," IEEE Trans. on Image Processing, vol. 26, no. 1, pp. 65-78, Jan. 2017.
[42] F. Kou, Z. G. Li, C. Wen, and W. H. Chen, "Multi-scale exposure fusion via gradient domain guided image filtering," in *IEEE International Conference on Multimedia and Expo.*, Hong Kong, China, pp. 1105-1110, Jul. 2017.
[43] F. Kou, Z. Wei, W. H. Chen, X. M. Wu, C. Y. Wen, and Z. G. Li, "Intelligent detail enhancement for exposure fusion," IEEE Trans. on Multimedia, vol. 20, no. 2, pp. 484-495, Feb. 2018.
[44] Y. Yang, W. Cao, S. Wu, and Z. Li, "Multi-scale fusion of two large-exposure-ratio images," IEEE Signal Processing Letters, vol. 25, no. 12, pp. 1885-1889, Dec. 2018.
[45] K. Ma, Z. Duanmu, H. Zhu, Y. Fang, and Z. Wang, "Deep guided learning for fast multi-exposure image fusion," IEEE Trans. on Image Processing, vol. 29, pp. 2808-2819, Dec. 2019.
[46] K. Dabov, A. Foi, V. Katkovnik, and K. Egiazarian, "Image denoising by sparse 3D transform-domain collaborative filtering," IEEE Trans. Image Process., vol. 16, no. 8, pp. 2080-2095, Aug. 2007.
[47] F. Banterle, P. Ledda, K. Debattista, and A. Chalmers, "Inverse tone mapping," in *Proc. of GRAPHITE'06*, pp. 349-356, 2006.
[48] A. Akyüz, R. Fleming, B. E. Riecke, E. Reinhard, and H. H. Bülthoff, "Do HDR displays support LDR content?: a psychophysical evaluation," ACM Trans. Graph., vol. 26, no. 3, Article 38, Jul. 2007.
[49] S. Tanaka, T. Otaka, et al., "Single exposure type wide dynamic range CMOS image sensor with enhanced nir sensitivity," ITE Transactions on Media Technology and Applications, vol. 6, no. 7, pp. 195-201, Jul. 2018.
[50] X. Guo, Y. Li, and H. Ling, "LIME: Low-light image enhancement via illumination map estimation," IEEE Trans. on Image Processing, vol. 26, no. 2, pp. 982-993, Feb. 2017.
[51] E. Reinhard, M. Stark, P. Shirley, and J. Ferwerda, "Photographic tone reproduction for digital images," ACM Trans. Graph., vol. 21, no. 3, pp. 267-276, Jul. 2002.
[52] F. Durand and J. Dorsey, "Fast bilateral filtering for the display of high-dynamic-range images," ACM Trans. Graphics, vol. 21, no. 3, pp. 257-266, Jul. 2002.
[53] Z. Farbman, R. Fattal, D. Lischinski, and R. Szeliski, "Edge-preserving decompositions for multi-scale tone and detail manipulation," ACM Trans. Graphics, vol. 21, no. 3, pp. 249–256, Aug. 2008.
[54] Z. Liang, J. Xu, D. Zhang, Z. Cao, and L. Zhang, "A hybrid L1-L0 layer decomposition model for tone mapping," in *IEEE Conference on Computer Vision and Pattern Recognition*, Jun. 2018.
[55] Y. L. Xu, Z. Y. Liu, J. Liu, W. H. Chen, and Z. G. Li, "Restoration of HDR images for SVE-based HDRI via a novel DCNN," in *IEEE 2021 International Conference on Multimedia and Expo*, P. R. China, 2021.
[56] M. Kobayashi, H. Sekine, T. Miki, T. Muto, T. Tsuboi, Y. Onuki, Y. Matsuno, H. Takahashi, T. Ichikawa, and S. Inoue, "A 3.4 μm pixel pitch global shutter CMOS image sensor with dual in-pixel charge domain memory," Japanese of Applied Physics, 58 SBBL02, 2019.
[57] R. Wang, Z. Qing, C. Fu, X. Shen, W. Zheng, and J. Jia, "Single image-based ghost-free high dynamic range imaging using local histogram stretching and spatially-adaptive denoising," IEEE Transactions on Consumer Electronics, vol. 57, no. 4, pp. 1478-1484, Apr. 2011.
[58] P. J. Sung, J. W. Soh, and N. I. Cho, "Generation of high dynamic range illumination from a single image for the enhancement of undesirably illuminated images," Multimedia Tools and Applications, vol. 78, pp. 20263-20283, Feb. 2019.
[59] K. Ma, K. Zeng, and Z. Wang, "Perceptual quality assessment for multi-exposure image fusion," IEEE Trans. on Image Processing, vol. 24, no. 11, pp. 3345-3356, Nov. 2015.
[60] R. Wang, Z. Qing, C. Fu, X. Shen, W. Zheng, and J. Jia, "Underexposed photo enhancement using deep illumination estimation," in *IEEE Conference on Computer Vision and Pattern Recognition*, Jun. 2019.
[61] S. Bell, P. Upchurch, N. Snavely, and K. Bala, "Material recognition in the wild with the materials in context database," in *IEEE Conference on Computer Vision and Pattern Recognition*, Jun. 2015.
[62] R. Mantiuk, K. J. Kim, Allan G. Rempel, and W. Heidrich, "HDR-VDP-2: A calibrated visual metric for visibility and quality predictions in all luminance conditions," ACM Trans. Graphics, vol. 24, no. 4, pp. 40:1-14, Jul. 2011.
[63] J. Wang, D. Scharstein, A. Bapa, et al. "A practical stereo depth system for smart glasses," in *IEEE Conference on Computer Vision and Pattern Recognition*, Jun. 2023.
[64] S. Shao, Z. Pei, W. Chen, X. Wu, and Z. Li, "NDDepth: normal-distance assisted monocular depth estimation," submitted to *International Conference on Computer Vision*, Oct. 2023.