Neural-Augmented HDR Imaging via Two Aligned Large-Exposure-Ratio Images

Z. G. Li, *Fellow, IEEE*, C. B. Zheng*, *Member, IEEE*, B. Chen and S. Q. Wu, *Senior Member, IEEE*

***Abstract*—Two aligned large-exposure-ratio (LER) images are captured simultaneously by emerging devices for a high dynamic range (HDR) scene to avoid ghosting artifacts from appearing in an image fused from them. However, brightness order reversal usually appears when the two LER images are directly fused together by an existing multi-scale exposure fusion (MEF) algorithm. In this paper, a novel neural augmentation-based exposure interpolation framework is introduced in this paper by seamlessly integrating physics-driven and data-driven approaches. The physics-driven method infers an initial representation while the data-driven one learns the remaining information for the middle-exposure image resulted from the two LER images. They compensate each other well in the proposed framework. These two LER images and the interpolated image are fused together using a novel data-driven MEF algorithm which can further interpolate information for the two LER images. Experimental results show that the proposed framework can indeed solve the brightness order reversal problem for the fusion of two LER images. Real high-frequency information from the two LER images is also preserved well in the fused image.**

***Index Terms*—High dynamic range imaging, Exposure interpolation, Neural augmentation, Physics-driven, Data-driven**

I. Introduction

Due to the gap of dynamic ranges between a real-world scenes and an 8-bit sRGB image image, there are over-exposed and/or under-exposed regions in the 8-bit sRGB. High dynamic range (HDR) imaging is thus highly demanded by many real-world applications [1]–[3]. Merging multi-exposed low dynamic range (LDR) sRGB images to expand the dynamic range is an effective way for the HDR imaging [4], [5]. The differently exposed images are usually captured sequentially while ghosting artifacts caused by moving objects are Achilles' heel for such a capturing method [6]–[8]. Fortunately, this issue can be addressed by using emerging capturing devices [3], [9]–[12] i.e., capture the differently exposed images simultaneously.

Two aligned large-exposure-ratio (LER) sRGB images can be captured so as to save the cost of the new devices while capturing information as much as possible from HDR scenes [11]–[15]. High-light regions in the dark image could be darker than shadow regions in the bright one as shown in Fig. 1(a). Multi-scale exposure fusion (MEF) algorithms such as [16], [17], [19]–[23] could yield unnatural results caused by brightness order reversal (BOR) as in Fig. 1(b). The right display board and trees are darker than the left display board in the original scene (or Fig. 1(a)). However, they are brighter than the left display board in the fused image in Fig. 1(b).

The BOR issue can be addressed by directly fusing the two LER images via a data-driven method such as [26]. However, local contrast and global contrast are not preserved well as demonstrated in Fig. 1(c). This issue can also be efficiently addressed via exposure-interpolated technique as indicated in [13]. The exposure interpolation is based on an observation that there is no BOR artifacts when a few normal-exposure-ration (NER) images are fused together by the existing MEF algorithms. A simple physics-driven method was proposed in [13] to synthesize an image with the middle exposure. Thus, the two LER images and the interpolated image become three NER images and alleviate BOR problem. However, the quality of the interpolated image is poor due to the limited representation capability of the physics-driven method. Subsequently, the quality of finally fused image is also poor. A novel physics-driven deep exposure interpolation framework was proposed to improve the interpolated image in [25]. The image with the middle exposure is first interpolated via a physics-driven method and then refined by a data-driven one. The framework in [25] outperforms the physics-driven method in [13]. Unfortunately, the framework in [25] cannot preserve high-frequency information well. The physics-driven method in [25] still needs to be improved even though it is better than the methods in [13]. As indicated in [38], the accuracy of the physics-driven method is important for the neural augmentation. With a more accurate physics-driven method, the neural augmentation framework would have more chance to outperform the data-driven method. It is thus desired to develop a new exposure interpolation framework for the new HDR imaging devices.

In this paper, a novel neural augmentation-based exposure interpolation framework is introduced for two aligned LER sRGB images Z_1 and Z_3 by combining physics-driven with data-driven methods in this paper. Our physics-driven exposure interpolation includes two parts: physics-driven representation of high-frequency information and physics-driven representation of low-frequency information. The proposed physics-driven representation is different from the one in [27] which only includes the former. A new mapping algorithm is first proposed for the two aligned differently exposed images Z_1 and Z_3. It is then adopted to produce two high-frequency images from the images Z_1 and Z_3 , which are finally merged

* correspondence author

Zhengguo Li is with VI department, the Institute for Infocomm Research, A*STAR, 1 Fusionopolis Way, #21-01, Connexis South Tower, Singapore 138632, (email: ezgli@i2r.a-star.edu.sg). Chaobing Zheng, Bin Chen nand Shiqian Wu are with the School of Electronic Information, Wuhan University of Science and Technology, Wuhan 430081, China. Shiqian Wu is also with the Institute of Advanced Displays and Imaging, Henan Academy of Sciences, Zhengzhou 450016, China. (emails: {zhengchaobing,chenbin, shiqian.wu}@wust.edu.cn).

Fig. 1: *Two solutions to avoid brightness order reversal when two LER images are fused. (a) includes two LER images; (b) is an image fused from these two images in (a) using the MEF algorithm in [16]; (c) is an image fused by the data-driven algorithm in [26]; and (d) is an image fused from these two images in (a) and an interpolated image by the proposed framework.*

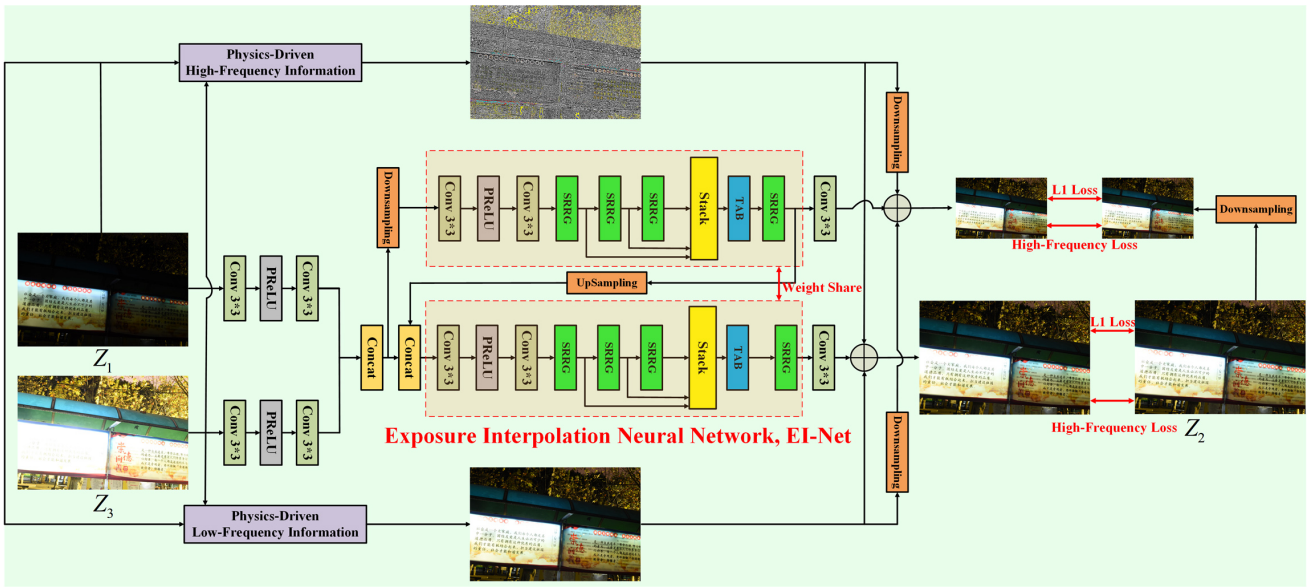


Fig. 2: *The proposed physics-driven deep exposure interpolation framework for the two LER images Z_1 and Z_3. The physics-driven high-frequency information $f_{m,h}(Z_1, Z_3)$ and low-frequency information $f_{m,l}(Z_1, Z_3)$ are first generated. The images Z_1, Z_3, and the physics-driven information $f_m(Z_1, Z_3)$ are fed into the proposed dual-scale EI-Net to learn the information $f_d(Z_1, Z_3)$ (The EI-Net in [27] only has a single-scale). The $f_m(Z_1, Z_3)$ and $f_d(Z_1, Z_3)$ are combined together to generate the image with the middle exposure time Δt_2.*

to obtain the the interpolated image's high-frequency information. The physics-driven representation of low-frequency information is based on intensity mapping functions (IMFs) among differently exposed images and the proposed new mapping algorithm. Two images are synthesized from the images Z_1 and Z_3, respectively. However, the IMFs are not reliable in under-exposed regions of image Z_1 and over-exposed regions of image Z_3 [42]. The proposed new mapping algorithm is then applied to improve the two synthesized images corresponding to the under-exposed regions of the image Z_1 and the over-exposed regions of the image Z_3 respectively. This is highly demanded for the proposed neural augmentation because the physics-driven method is required to be as accurate as possible by the neural augmentation [38].

A dual-scale exposure interpolation neural network (EI-Net) is then designed to learn the remaining information from the two images Z_1 and Z_3 progressively for the interpolated image as demonstrated in Figs. 2 and 3. This is different from the EI-Net in [27] which only has a single scale. Intuitively, the interpolated image by the physics-driven approach can be regarded as a noisy image for image Z_2 because the proposed physics-driven method has a limited representation capability. As such, the proposed EI-Net is constructed from the CNN in [39]. Short-skip-connection recursive residual group (SRRG) and triple attention block (TAB) are two important components of the EI-Net. The SRRG is constructed from the residual network in [41], and consists of multiple recursive residual groups (RRGs). The proposed SRRG is simpler than the SRRG in [27]. Each RRG has several dual-attention blocks (DABs), and each DAB combines a channel attention block and a spatial attention block. The short-skip-connections among the RRGs can keep and pass shallow layers' information into deep layers. Each TAB consists of one DAB and one unit attention block (UAB). The UAB and DAB can extract features with improved representation abilities.

An unsupervised learning based MEF algorithm is finally pro-

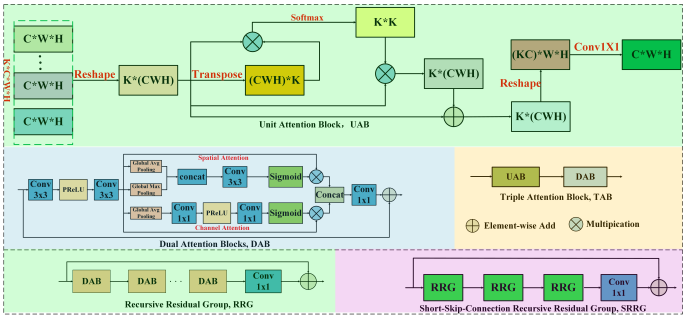


Fig. 3: *Illustration of the key components including SRRG, DAB, UAB, and TAB. The proposed SRRG is simpler than the SRRG in [27].*

posed to fuse the interpolated image and the two LER images. The proposed algorithm is further different from the algorithms in [25], [27] in the sense that these images are fused by using the algorithm in [17]. As shown in [14], the loss function is important for the data-driven exposure fusion algorithm. The loss function and the set of images to be fused are coupled by the algorithm in [14]. They are decoupled in the proposed MEF algorithm. Particularly, the loss function is computed by using the fused image, the two LER images, and two other images from the same HDR scene. The exposure times of the two other images are within the exposure time range defined by the two LER images. Clearly, more information could be kept by using the proposed MEF algorithm. The proposed HDR imaging algorithm wins the state-of-the-art (SOTA) algorithms in [13], [14], [16], [17], [19], [24]–[26], [35], [36], [49] in terms of MEF-SSIM metric [40] and the subjective quality points of view. Experimental results indicate that the MEF algorithms in [16], [17], [19], [24], [35], [36] indeed suffer from the brightness order reversal. Although this problem is overcome by the data-driven algorithms in [14], [26], they have difficulty on preserving high-frequency information [28], [29]. This is also demonstrated in Fig. 1. Experimental results also show that the exposure interpolation-based algorithms outperform the data-driven algorithms from both subjective quality and objective quality points of view. In addition, experimental results verify that the accuracy of the physics-driven method is indeed crucial for the neural augmentation [38]. In summary, the contributions of this paper are highlighted as follows:

1) A novel physics-driven deep exposure interpolation framework is introduced in this paper by fusing physics-driven and data-driven methods. Both theoretical analysis and experiment results show that the proposed framework can improve the quality of the interpolated image and also converges faster than the data-driven approach;

2) A new unsupervised data-driven MEF algorithm is proposed to fuse the two LER images and the interpolated image. The proposed MEF algorithm includes a new type of loss functions which can be defined by the fused image and any set of differently exposed images from the same HDR scene. The proposed MEF algorithm can further interpolate information for the two LER images.

The rest of this paper is organized as follow: a dataset is provided in Section II. The proposed physics-driven deep interpolation framework is provided in Section III. These two LER images and the interpolated images are fused together in Section IV. Experimental result are provided in Section V to verify the proposed framework. Finally, conclusion remarks are given in Section VI.

II. An Exposure Interpolation Dataset

Our exposure interpolation dataset is constructed from two existing datasets on HDR imaging in [11], [48]. The multi-exposed 8-bit sRGB images in these two datasets are captured by VETHDR-Nikon and VETHDR-Canon, respectively. The CRFs of these two cameras are assumed to be available. Actually, they can estimated by using the method in [5].

Fig. 4: *Three sets of (a) short-exposure, (b) middle-exposure, and (c) long-exposure images. Each set of images is aligned.*

Each set in the exposure interpolation dataset includes three differently exposed LDR images without camera movements and moving objects, Z_1, Z_2, and Z_3. These three images are sub-sampled from the corresponding set in [11], [48] by discarding the remaining images. The exposure time of $Z_i (1 \leq i \leq 3)$, is Δt_i. For simplicity, it is assumed that

$$\Delta t_1 \ll \Delta t_3 \; ; \; \frac{\Delta t_2}{\Delta t_1} = \frac{\Delta t_3}{\Delta t_2}. \tag{1}$$

The set of images Z_1, Z_2, and Z_3 is a typical geometric exposure variation [47], and each available exposure setting is a constant factor larger than its subsequent lower setting. The two images in Figs. 4(a) and 4(c) are the inputs to the proposed framework. The images in Fig. 4(b) are the ground-truth ones for training the proposed framework. It is worth noting that the middle set in Fig. 4 is not from the two datasets in [11], [48]. The objective is to evaluate the robustness of the proposed neural augmentation framework.

The resolution of all the images is 720×1280. The interval of two LER inputs is 6-EV. The proposed dataset is composed of 750 training sequences, 80 verifying sequences, and 100 testing sequences.

III. Neural Augmentation-Based Exposure Interpolation

The proposed framework consists of a physics-driven initial representation and a data-driven refinement for the image with the middle exposure. They *compensate* each other in such a neural augmentation [25], [37], [38].

Given two aligned LER images Z_1 and Z_3 with N pixels, the neural augmentation-based exposure interpolation method in [25] represents Z_2 as

$$f_t(Z_1, Z_3) = f_m(Z_1, Z_3) + f_{d,t}(Z_1, Z_3), \tag{2}$$

where $f_m(Z_1, Z_3)$ and $f_{d,t}(Z_1, Z_3)$ are physics-driven representation and data-driven representation of Z_2, respectively. For simplicity, the item t in the equation (2) is omitted in the main body.

The physics-driven representation $f_m(Z_1, Z_3)$ in [25] is based on the IMFs, and it cannot preserve high-frequency information. As pointed out in [28], [29], CNNs are usually biased towards learning low-frequency functions because the CNNs correspond to kernels with a rapid frequency falloff [30]. Loss of high-frequency information is one possible issue for the neural augmentation [25]. Another issue on the neural augmentation in [25] is the possible low accuracy of the physics-driven method in [25]. The neural augmentation could outperform the data-driven approach if the physics-driven method is accurate. A new neural augmentation framework will be proposed in this section to address the issues. The representation $f(Z_1, Z_3)$ can be decomposed as in the equation (2) with the following physics-driven $f_m(Z_1, Z_3)$:

$$f_m(Z_1, Z_3) = f_{m,h}(Z_1, Z_3) + f_{m,l}(Z_1, Z_3), \tag{3}$$

where $f_{m,l}(Z_1, Z_3)$ and $f_{m,h}(Z_1, Z_3)$ are physics-driven low-frequency and high-frequency information, respectively. It should be pointed out that the physics-driven interpolation in [27] only includes $f_{m,h}(Z_1, Z_3)$.

Figs. 2 and 3 summarize the proposed neural augmented exposure interpolation framework. The high-frequency information $f_{m,h}(Z_1, Z_3)$ and low-frequency information $f_{m,l}(Z_1, Z_3)$ are first modelled via a physics-driven method. The data-driven information $f_d(Z_1, Z_3)$ is then learnt via a CNN with inputs as the two LER images Z_1 and Z_3 and the physics-driven information $f_m(Z_1, z_3)$. The details are given in the following two subsections.

A. Physics-Driven Exposure Interpolation

The proposed physics-driven representation $f_{m,h}(Z_1, Z_3)$ is derived from the algorithms in [27], [33], [34].

The images $Z_i (i = 1, 3)$ can be normalized to match the image Z_2 as [33], [34]

$$\tilde{Z}_c^{i\to 2}(p) = \alpha_c^{i\to 2}(Z_{i,c}(p) - \mu_{Z_{i,c}}) + \mu_{Z_{2,c}}, \tag{4}$$

where $c \in \{R, G, B\}$. $\mu_{Z_{i,c}}$ and $\sigma_{Z_{i,c}} (i = 1, 2, 3)$ are the mean and standard deviation of all the pixels in the images $Z_{i,c}$, respectively. $\alpha_c^{i\to 2}$ is $\sigma_{Z_{2,c}}/\sigma_{Z_{i,c}}$.

For simplicity, let a square window centered at the pixel p with a radius ρ be denoted as $\Omega_\rho(p)$, and $\mu_{Z_{i,c},\rho}(p)$ be the mean value of the image $Z_{i,c}$ in the window $\Omega_\rho(p)$. The mapping algorithm (4) can be improved as [27]

$$\begin{aligned}\tilde{Z}_{c,\rho}^{i\to 2}(p) =& \alpha_c^{i\to 2}(Z_{i,c}(p) - \mu_{Z_{i,c},\rho}(p)) + \mu_{Z_{2,c},\rho}(p) \\ \doteq& \alpha_c^{i\to 2}\varphi_{h,\rho}(Z_{i,c}(p)) + \mu_{Z_{2,c},\rho}(p) \qquad (5)\\ \doteq& f_{c,h,\rho}^{i\to 2}(p) + \mu_{Z_{2,c},\rho}(p). \qquad (6)\end{aligned}$$

It can be easily verified that the physics-driven representation (5) maximizes the following Pearson correlation:

$$PC(Z_{i,c}, Z_{2,c}) = \frac{1}{|\Theta|}\sum_{p\in\Theta}\frac{\text{cov}(B_{Z_{i,c},\rho}(p), B_{Z_{2,c},\rho}(p))}{\sigma(B_{Z_{i,c},\rho}(p))\sigma(B_{Z_{2,c},\rho}(p))}, \tag{7}$$

where $B_{Z_{i,c},\rho}(p)$ is the square window in the image $Z_{i,c}$ centered at the pixel p of the radius ρ. Θ is a set of pixels and $|\Theta|$ is the cardinality of the set Θ. An important case is that the Θ is the full set. $\text{cov}(B_{Z_{i,c},\rho}(p), B_{Z_{2,c},\rho}(p))$ is the covariance of the blocks $B_{Z_{i,c},\rho}(p)$ and $B_{Z_{2,c},\rho}(p)$, and $\sigma(B_{Z_{i,c},\rho}(p))$ is the standard derivation of the block $B_{Z_{i,c},\rho}(p)$.

However, the choice of $\alpha_c^{i\to 2}$ as $\sigma_{Z_{2,c}}/\sigma_{Z_{i,c}}$ in the physics-driven representation (5) is not the optimal solution to the following problem:

$$\min_{\alpha_c^{i\to 2}}\{\sum_p(\tilde{Z}_{c,\rho}^{i\to 2}(p) - Z_{2,c}(p))^2\}. \tag{8}$$

It can be easily derived that the optimal solution is

$$\alpha_c^{i\to 2,*} = \frac{\sum_p \varphi_{h,\rho}(Z_{i,c}(p))\varphi_{h,\rho}(Z_{2,c}(p))}{\sum_p \varphi_{h,\rho}^2(Z_{i,c}(p))}, \tag{9}$$

and two resultant high-frequency images $f_{c,h,\rho}^{i\to 2,*}$'s are then given as

$$\begin{cases} f_{c,h,\rho}^{1\to 2,*}(p) = \alpha_c^{1\to 2,*}\varphi_{h,\rho}(Z_{1,c}(p)) \\ f_{c,h,\rho}^{3\to 2,*}(p) = \alpha_c^{3\to 2,*}\varphi_{h,\rho}(Z_{3,c}(p)) \end{cases}. \tag{10}$$

Since $\Delta t_1 < \Delta t_2 < \Delta t_3$, $\tilde{Z}_c^{1\to 2}(p)$ is not reliable if $Z_{1,c}(p)$ is underexposed while $\tilde{Z}_c^{3\to 2}(p)$ is not reliable if $Z_{3,c}(p)$ is overexposed [42]. The mapped pixel $\tilde{Z}_2(p)$ is obtained by combining $\tilde{Z}_\rho^{1\to 2}(p)$ and $\tilde{Z}_\rho^{3\to 2}(p)$ as

$$\tilde{Z}_{2,c}(p) = \frac{\sum_{i\in\{1,3\}} W_i(Z_{i,c}(p))f_{c,h,\rho}^{i\to 2,*}(p)}{\sum_{i\in\{1,3\}} W_i(Z_{i,c}(p))} + \mu_{Z_{2,c},\rho}(p), \tag{11}$$

where $W_1(z)$ and $W_3(z)$ are defined as

$$W_1(z) = \begin{cases} \frac{z+1}{56}; & \text{if } z < 55 \\ 1; & \text{otherwise} \end{cases}, \tag{12}$$

$$W_3(z) = \begin{cases} 1; & \text{if } z < 200 \\ \frac{256-z}{56}; & \text{otherwise,} \end{cases}. \tag{13}$$

The new mapping algorithm (11) is verified by using 100 testing image sequences. The PSNR, SSIM and FSIM are computed by using the interpolated image $\tilde{Z}_2$ and the ground-truth image Z_2. Three different choices of ρ as 1, 2, and ∞ (i.e., the whole image) are compared as in Table I. The corresponding results for the choice of $\alpha_c^{i\to 2}$ as $\sigma_{Z_{2,c}}/\sigma_{Z_{i,c}}$ are

TABLE I: *peak signal-to-noise ratio(PSNR) (dB, ↑), structural similarity(SSIM) (↑), Feature Similarity(FSIM) (↑) and Entropy (↑) for three different ρ's (↑ larger is better)*

ρ	**1**	**2**	∞**(the whole image)**
PSNR	36.73	34.43	21.35
SSIM	0.9619	0.9426	0.7713
FSIM	0.9982	0.995	0.8808
Entropy	6.9512	6.9085	6.5843

TABLE II: *PSNR (dB, ↑), SSIM (↑), FSIM (↑) and and Entropy (↑) for the choice of $\alpha_c^{i\to 2}$ as $\sigma_{Z_{2,c}}/\sigma_{Z_{i,c}}$*

ρ	**1**	**2**	∞**(the whole image)**
PSNR	35.03	32.12	20.31
SSIM	0.9425	0.9071	0.7812
FSIM	0.9969	0.9892	0.8707
Entropy	6.9319	6.9079	6.5671

provided in Table II. Obviously, the new mapping algorithm (11) outperforms the one in [27].

Since the accuracy of the physics-driven method is crucial for the proposed neural augmentation framework, the proposed physics-driven representation of $f_{m,h}(Z_1, Z_3)$ is derived from the mapping algorithm (11) with ρ being selected as 1. It is worth noting that the mapping algorithm with ρ being selected as 1 is useful to address the ghost removal problem in RGB domain [42].

Since the image $Z_{2,c}$ is not available, the $\alpha_c^{1\to 2,*}$ and $\alpha_c^{3\to 2,*}$ are estimated as

$$\alpha_c^{1\to 2,*} = \sqrt{\frac{\sum_p \varphi_{h,1}(Z_{1,c}(p))\varphi_{h,1}(Z_{3,c}(p))}{\sum_p \varphi_{h,1}^2(Z_{1,c}(p))}}, \tag{14}$$

$$\alpha_c^{3\to 2,*} = \sqrt{\frac{\sum_p \varphi_{h,1}(Z_{1,c}(p))\varphi_{h,1}(Z_{3,c}(p))}{\sum_p \varphi_{h,1}^2(Z_{3,c}(p))}}. \tag{15}$$

The high-frequency information $f_{m,h}(Z_1, Z_3)$ is finally obtained as

$$f_{m,h}(Z_1(p), Z_3(p)) = \frac{\sum_{i\in\{1,3\}} W_i(Z_i(p)) f_{h,1}^{i\to 2,*}(p)}{\sum_{i\in\{1,3\}} W_i(Z_i(p))}. \tag{16}$$

Besides the above high-frequency information $f_{m,h}(Z_1, Z_3)$, the proposed physics-driven interpolation also includes the low-frequency information $f_{m,l}(Z_1, Z_3)$. Details on low-frequency information $f_{m,l}(Z_1, Z_3)$ are provided in the rest of this subsection. The available CRFs are denoted as $F_c(\cdot)$'s. The IMFs from $Z_{i,c}(i = 1, 3)$ to $Z_{2,c}$ are $\Lambda_c^{i\to 2}(\cdot)$ [42] which can be estimated as

$$\Lambda_c^{i\to 2}(z) = \arg \min_{z'\in[0,255]} \{|\frac{\Delta t_2}{\Delta t_i} F_c^{-1}(z) - F_c^{-1}(z')|\}, \tag{17}$$

where $F_c^{-1}(\cdot)$ is the inverse function of the CRF $F_c(\cdot)$. To improve the accuracy of the interpolated image $f_m(Z_1, Z_3)$, the IMFs $\Lambda_c^{i\to 2}(\cdot)$ are not necessary to be integers, and can be relaxed by using the gradients of the inverse CRFs $F^{-1}(\cdot)$.

Same as [13], [25], two images $Z_c^{1\to 2}$ (i.e., $\Lambda_c^{1\to 2}(Z_{1,c})$) and $Z_c^{3\to 2}$ (i.e., $\Lambda_c^{3\to 2}(Z_{3,c})$) are produced. However, the IMFs can neither reliably map a under-exposed pixel of $Z_{1,c}$ to the corresponding pixel of $Z_{2,c}$, nor reliably map an over-exposed pixel of $Z_{3,c}$ to the corresponding pixel of $Z_{2,c}$ [42]. On the other hand, the accuracy of the physics-driven method is important for the neural augmentation framework [38]. Therefore, the image $Z_c^{1\to 2}$ is improved by using the new mapping algorithm (5) as

$$\hat{Z}_c^{1\to 2}(p) = \Phi_c^{1\to 2}(Z_{1,c}(p)), \tag{18}$$

where the function $\Phi_c^{1\to 2}(z)$ is defined as

$$\begin{cases} \Lambda_c^{1\to 2}(z); & \text{if } z \geq 5 \\ \tilde{\alpha}_c^{1\to 2,*}\varphi_{h,1}(z) + \mu_{Z_c^{1\to 2},1}(p); & \text{otherwise} \end{cases}, \tag{19}$$

and the image $Z_c^{3\to 2}$ is also improved by using the new mapping algorithm (5) as

$$\hat{Z}_c^{3\to 2}(p) = \Phi_c^{3\to 2}(Z_{3,c}(p)), \tag{20}$$

where the function $\Phi_c^{3\to 2}(z)$ is defined as

$$\begin{cases} \Lambda_c^{3\to 2}(z); & \text{if } z \leq 250 \\ \tilde{\alpha}_c^{3\to 2,*}\varphi_{h,1}(z) + \mu_{Z_c^{3\to 2},1}(p); & \text{otherwise} \end{cases}, \tag{21}$$

and $\tilde{\alpha}_c^{i\to 2,*}(i \in \{1, 3\})$ is computed as

$$\tilde{\alpha}_c^{i\to 2,*} = \frac{\sum_p \varphi_{h,1}(Z_{i,c}(p))\varphi_{h,1}(Z_c^{i\to 2}(p))}{\sum_p \varphi_{h,1}^2(Z_{i,c}(p))}. \tag{22}$$

It is worth noting that one more method for generating the virtual images $\hat{Z}_c^{i\to 2}(i \in \{1, 3\})$ is to linearize the images using the gamma correction and adjust their exposures to the middle one as [43]

$$\hat{Z}_c^{i\to 2}(p) = Z_{i,c}(p)(\frac{\Delta t_2}{\Delta t_i})^{\frac{1}{2.2}}. \tag{23}$$

This simple method is widely adopted by data-driven ghost removal algorithms [6], [18], [43]–[46]. Unfortunately, this method is not accurate [8], [48], and usually introduces brightness distortion as shown in Fig. 5.

Fig. 5: *Comparison of the proposed physics-driven method and the physics-driven method on top of the gamma correction as in the equation (23).*

An intermediate image $Z_2^{(0)}$ can be interpolated by

$$Z_{2,c}^{(0)}(p) = \frac{\sum_{i\in\{1,3\}} W_i(Z_{i,c}(p))\hat{Z}_c^{i\to 2}(p)}{\sum_{i\in\{1,3\}} W_i(Z_{i,c}(p))}, \tag{24}$$

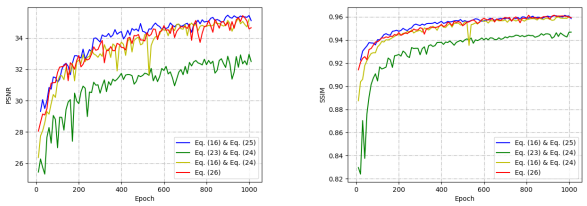
and the physics-driven representation $f_{m,l}(Z_1, Z_3)$ is then computed as

$$f_{m,l}(Z_1, Z_3) = \mu_{Z_2^{(0)},1}. \tag{25}$$

Besides the proposed physics-driven representation by the equations (3), (16), and (25), one more physics-driven representation is given by

$$f_m(Z_1, Z_3) = Z_2^{(0)}. \tag{26}$$

As shown in Fig. 6, the proposed physics-driven representation outperforms the above one (26) from both PSNR and SSIM points of view. This is because the high-frequency information cannot be well preserved by the IMFs $\Lambda_c^{i\to 2}(\cdot)$'s. Both of them outperform the physics-driven representation of $(Z_2^{(0)} + f_{m,h}(Z_1, Z_3))$ (here $\hat{Z}_c^{1\to 2}$ and $\hat{Z}_c^{3\to 2}$ are computed by the equations (18) and (20), respectively). This is also an evidence that the CNN has a fundamental limitation to represent the high-frequency information in natural images. We thus choose the proposed physics-driven representation. All of them outperform the one on top of the equations (23), (24), and (26). Clearly, the accuracy of the gamma correction (23) is indeed an issue.



(a) (b)

Fig. 6: *Comparisons of PSNR and SSIM for four different physics-driven representations. The proposed one outperforms the other three alternatives from both PSNR and SSIM points of view.*

B. Data-Driven Refinement

The remaining information $f_d(Z_1, Z_3)$ is learnt from the three images $\{Z_1, Z_3, f_m(Z_1, Z_3)\}$ progressively by using the EI-Net in Figs. 2 and 3 which has two scales while the EI-Net in [27] has a single scale. Experiment results indicate the proposed dual-scale EI-Net significantly outperforms the single-scale EI-Net in [27].

Two components of the proposed EI-Net are SRRG and TAB. The structure of the SRRG is inspired by the residual network in [41]. The proposed SRRG is simpler than that in [27], and it consists of several RRGs. Each RRG is composed of multiple DABs [39]. With the short-skip-connections among the RRGs, shallow layers information can be passed into deep layers by the SRRG. As such, the remaining high-frequency information could be well preserved, facilitating the efficient learning of the remaining information $f_d(Z_1, Z_3)$. The TAB consists of a UAB and a DAB [39]. The DAB can suppress those less useful features while allow the propagation of more useful ones [39]. The UAB can assign different attention weights to features of different layers in order to extract features with improved representation abilities.

The loss function of our data-driven approach is defined as

$$L_d = L_r + L_h, \tag{27}$$

where the items L_r and L_h are

$$L_r = \|Z_2 - (f_m(Z_1, Z_3) + f_d(Z_1, Z_3))\|_1, \tag{28}$$

$$L_h = \|\varphi_{h,1}(Z_2) - \varphi_{h,1}(f_m(Z_1, Z_3) + f_d(Z_1, Z_3))\|_1, \tag{29}$$

and the function $\varphi_{h,1}(\cdot)$ is given in the equation (5). The loss function L_h is a new loss function which can be adopted to help the EI-Net preserve the high-frequency information.

IV. Unsupervised Learning Based MEF

The two LER images Z_1 and Z_3 as well as the interpolated image $f(Z_1, Z_3)$ are fused together by using an unsupervised learning based MEF algorithm in this section.

The proposed MEF algorithm is on top of a multi-scale U-Net architecture as illustrated in Fig. 7. Convolutional layers are used to extract features from the three inputs, followed by a fusion process that integrates these features across multiple scales. This multi-scale approach operates in a coarse-to-fine manner, effectively capturing details at different resolutions. To reduce the computational cost, a weight-sharing strategy is adopted throughout the network.

Besides the network structure, the loss function also plays an important role in the proposed exposure fusion algorithm. Same as [14], the MEF-SSIM is also adopted to compute the loss function. For simplicity, the fused image is denoted as Z_F, and the set of differently exposed images for the definition of loss function is denoted as Ω_m. Normally, the set Ω_m is selected as the set of images to be fused, i.e., the images Z_1, Z_3 and $f(Z_1, Z_3)$ [14]. As such, the loss function and the set of images to be fused are coupled. Here, a new type of loss function is introduced by decoupling the loss function and the set of images to be fused. Instead of choosing the set Ω_m as Z_1, Z_3 and $f(Z_1, Z_3)$, it consists of the LER images Z_1 and Z_3 as well as two other differently exposed images from the same HDR scene with the exposure times between the two LER images Z_1 and Z_3. The EV gap between any two successive images in the set Ω_m is 2. Clearly, more information could be further interpolated by the proposed data-driven MEF algorithm. The overall proposed algorithm is shown in Algorithm 1.

Algorithm 1: Neural-augmented HDR imaging via two aligned LER images

Inputs	Two LER images Z_1 and Z_3 .
Output	an information enrich LDR image.
Step 1	Infer the physics-driven high-frequency information $f_{m,h}(Z_1, Z_3)$ via the equation (16) and the low-frequency information $f_{m,l}(Z_1, Z_3)$ via the equation (25).
Step 2	Learn the remaining information $f_d(Z_1, Z_3)$ using the EI-Net in Figs. 2 and 3.
Step 3	Combine the $f_m(Z_1, Z_3)$ and $f_d(Z_1, Z_3)$ to generate the interpolated image $f(Z_1, Z_3)$.
Step 4	Fuse the three differently exposed LDR images via the proposed data-driven MEF algorithm.

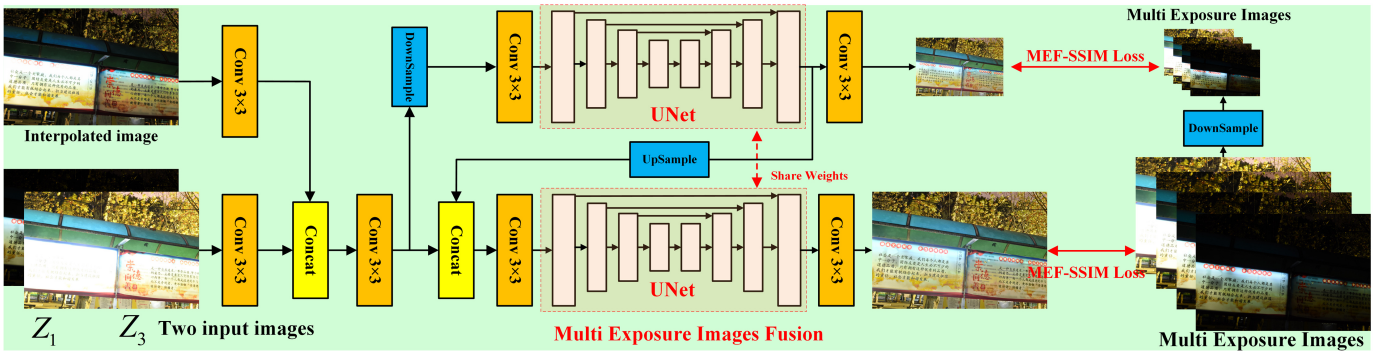


Fig. 7: *Inputs of the proposed multi-scale exposure fusion algorithm are Z_1, Z_3 and $f(Z_1, Z_3)$. The fused image Z_F approaches four differently exposed images from the same HDR scene with the EV gap as 2. Z_1 and Z_3 are included in the set of four differently exposed images.*

It should be mentioned that more images from the same HDR scene with the exposure times larger than Δt_3 and/or smaller than Δt_1 can also be included for the computation of the loss function. More information from the same HDR scene could be extrapolated in this way. Such a loss function will be studied in our future research.

V. Experimental Results

A. Implementation Details

Experimental results are provided to validate the proposed exposure interpolation framework. In our experiments, the size of the window $\Omega_\rho(p)$ is set to 3×3 to extract the high-frequency information efficiently.

Our implementation is built upon the PyTorch framework and executed on four NVIDIA GP100 GPUs. To enhance the diversity of the training data, we apply data augmentation techniques, including mirroring and randomly cropping $512 * 512$ patches from each input image. The proposed EI-Net is trained using the designed loss functions and optimised with the Adam optimiser, employing a batch size of 4. The initial learning rate is set to 10^{-5} and progressively reduced following a cosine annealing schedule. The training process spans 1000 epochs, after which the model is evaluated on the validation dataset to assess its performance. It is particularly noteworthy that the multi-scale fusion network employs an identical training strategy.

B. Comparison with Different Exposure Interpolation Algorithms

TABLE III: *Comparison of different physics-driven intensity mapping algorithms.*

Method	PSNR↑	SSIM↑	FSIM↑	Entropy↑
physics-driven method in [13]	25.17	0.8982	0.9638	6.6683
physics-driven method in [25]	26.42	0.9039	0.9703	6.7729
our physics-driven method (24)	27.58	0.9186	0.9772	6.8006

Since one major contribution of this paper is the physics-driven method (24), the proposed neural augmentation-based exposure interpolation is then compared with the physics-driven method in [13] and the neural augmentation-based exposure interpolation in [25]. As demonstrated in Table III,IV and Fig. 8, the proposed neural augmentation-based exposure interpolation can achieve much higher PSNR and SSIM values than the neural augmentation-based exposure interpolation in [25]. The impact of the physics-driven method is also evaluated by using the same data-driven method as the proposed EI-NET. As shown in Fig. 9, the accuracy of the physics-driven method is indeed important for the neural augmentation which will be theoretically analyzed in the appendix. The data-driven method is locally convergent.

TABLE IV: *Comparison of different neural augmentation intensity mapping algorithms.*

Method	PSNR↑	SSIM↑	FSIM↑	Entropy↑
data-driven method in Figs. 2 and 3	33.81	0.9516	0.9878	6.8806
neural augmentation in [25]	31.73	0.9361	0.9837	6.9057
our neural augmentation	35.34	0.9615	0.9895	6.9131

C. Comparison with Eleven SOTA Algorithms

The proposed algorithm is compared with five conventional MEF algorithms: ExposureFusion [16], Kou_2017 [17], WQT_2020 [19], Yang_2018 [13], and PAS-MEF [36], five deep exposure fusion algorithms: DeepFusion [14], DeepGuide [24], FusionDN [35], DEP-MEF [26], and HoLoCo [49], as well as the neural augmentation-based algorithm in [25]. The model sizes of DeepFusion [14], DeepGuide [24], FusionDN [35], and DEP-MEF [26], the framework in [25], HoLoCo in [49] and the proposed framework are 333 KB, 341.9 KB, 1.7 GB, 54.5 MB, 10.9 MB, 33 MB and 21.3 MB, respectively. The inputs to all the algorithms are the two LER images Z_1 and Z_3.

Fig. 10 includes fused images for three pairs of LER images in Figs. 4(a) and 4(c). Readers can zoom in digital version of all the images to distinguish differences among them. As indicated in the introduction, the right display board and trees are darker than the left display board in the top pair of LER images. However, they are brighter than the left display board in the fused images by the algorithms in [16], [17], [19]. The PAS-MEF [36] cannot preserve information in the brightest regions. Neither local contrast nor global contrast (such as Chinese characters) is preserved well by the algorithms in [13],

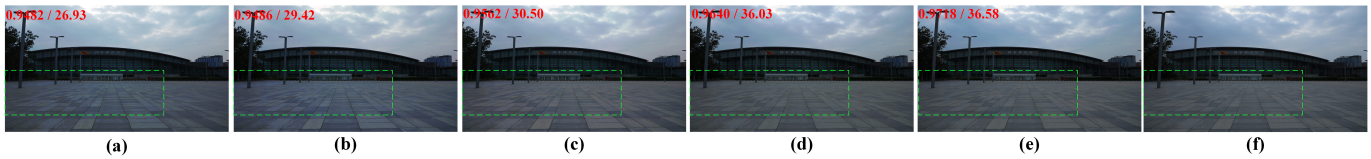


Fig. 8: *Interpolated images via (a) the physics-driven method in [13]; (b) the physics-driven method in [25]; (c) the proposed physics-driven method (24); (d) the neural augmentation in [25]; (e) the proposed neural augmentation; and (f) the ground truth image. As highlighted in the green box, the result in (e) is noticeably closer to the ground truth image, whereas (a) and (b) exhibit significant colour distortion. Furthermore, in the sky region, there is a considerable colour discrepancy between the ground truth and the results in (a), (b), and (c).*

TABLE V: *MEF-SSIM of 12 Different Algorithms (↑: larger is better))*

MEF [16]	Kou_2017 [17]	WQT_2020 [19]	DeepFusion [14]	DeepGuide [24]	Yang_2018 [13]
0.8927	0.8922	0.8926	0.9041	0.8828	0.9118
FusionDN [35]	**PAS-MEF [36]**	**DEP-MEF [26]**	**EA-Net [25]**	**HoLoCo [49]**	**Prpposed]**
0.8756	0.8878	0.9202	0.9091	0.7857	0.9473

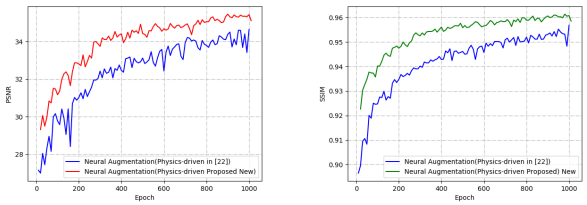


Fig. 9: *Comparisons between two neural augmentations on top of the proposed physics-driven representation (3), (16), and (25) and the physics-driven method in [25].*

[14], [24], [26], [36] even though they can somewhat preserve the relative brightness. There is also visible color distortion by FusionDN [35], DeepGuide [24] and DeepFusion [14]. Similar observations can be obtained for the second and third pairs of LER images. For example, the sky looks darker than the floor in the fused images by the algorithms in [16], [17], [19] for the second pair even though it is actually brighter than the floor. In addition, it is shown by the experimental results that CNNs are indeed poor at learning high-frequency functions [28], [29].

Besides the above subjective evaluation, the popular MEF-SSIM in [40] is also applied to study all the algorithms. The MEF-SSIM is calculated by using the fused image and the three captured images $Z_i (1 \leq i \leq 3)$ which are the reference images [40]. As demonstrated in Table V, the proposed algorithm significantly outperforms all the ten algorithms from the MEF-SSIM point of view. This is mainly own to the proposed physics-driven method and the dual-scale of the proposed EI-Net. It should be pointed out that the MEF-SSIM values of the algorithms in [16], [19], [24], [36] are respectively improved as 0.9186, 0.9207, 0.9139, and 0.8891 when they are used to fuse the two input images and the interpolated image. It can be illustrated that the algorithms in [13], [26] and the proposed one outperform other seven algorithms in Fig. 10 subjectively and objectively from the MEF-SSIM point of view in Table V. It is shown that the MEF-SSIM and the subjective evaluation are consistent with each other. Since the exposure interpolation-based algorithms in [13], [25] and the proposed one outperform the data-driven algorithms in [14], [26], [35] subjectively and objectively, the exposure interpolation is indeed required by the HDR imaging on top of two LER images. On the other hand, it should be pointed out that there are many other ways to implement exposure interpolation.

D. Ablation Study of Key Components

Key components of the proposed framework are: 1) the proposed physics-driven representation (3), (16), and (25), 2) different choices of ρ in the equation (16), 3) the loss function L_h (29), 4) the UAB, and 5) the dual-scale. Their performances are evaluated in this subsection. As illustrated in Table VI, the choice of ρ as 1 can indeed obtain a better performance than 2 for the proposed framework. Since the physics-driven method with ρ as 1 is more accurate than that with ρ as 2, the accuracy of the physics-driven method is indeed important for the neural augmentation framework as indicated in [38]. The dual-scale EI-Net also significantly outperforms the single-scale EI-Net. Both the short-skip connections and UAB are helpful for the proposed dual-scale EI-Net for the PSNR, SSIM, and FSIM points of view.

This subsection also compares the proposed neural augmentation framework with the data-driven approach in Figs. 2 and 3 without the physics-driven representation f_m. The proposed framework improves both the PSNR and the SSIM as shown in Table III and Fig. 11. This is because the CNN is usually locally convergent. With an accurate physics-driven representation, the proposed neural augmentation converges to a more optimal solution.

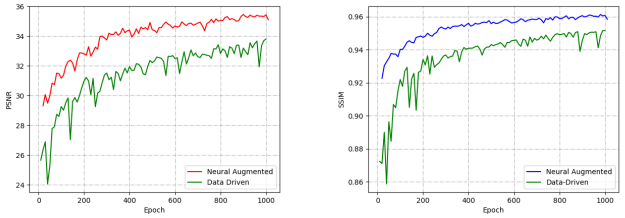
E. Limitation of the Proposed Framework

The image with the middle exposure $f(Z_1, Z_3)$ is firstly interpolated for the two LER images Z_1 and Z_3. All the three

Fig. 10: *Fused images by twelve different algorithms for three pairs of LER images in Figures 4(a) and 4(c). (a) MEF [16]; (b) Kou_2017 [17]; (c) WQT_2020 [19]; (d) DeepFusion [14]; (e) DeepGuide [24]; (f) Yang_2018 [13]; (g) FusionDN [35]; (h) PAS-MEF [36] ; (i) DEP-MEF [26]; (j) EA-Net [25], (k) HoLoCo [49] and (l) the proposed one.*

TABLE VI: *Ablation study of the data-driven approach on generating high exposure images from low exposure ones (↑: larger is better)*

Case	physics-driven	L_h	ρ	UAB	Scale	SSIM (↑)	PSNR (dB: ↑)	FSIM (↑)	Entropy↑
1	N	Y	NIL	Y	2	0.9516	33.81	0.9878	6.8806
2	Y	Y	2	Y	2	0.9612	35.05	0.9892	6.9042
3	Y	N	1	Y	2	0.9589	35.07	0.9891	6.9008
4	Y	Y	1	N	2	0.9613	35.18	0.9893	6.8923
5	Y	Y	1	Y	1	0.9614	35.04	0.9894	6.8946
6	Y	Y	1	Y	2	**0.9615**	**35.34**	**0.9895**	**6.9131**



(a) (b)

Fig. 11: *Comparisons of PSNR and SSIM between the proposed neural augmentation and the data-driven method.*

images form a set of NER images. The three NER images are then fused together via the unsupervised learning based MEF algorithm in Section IV. Even though the quality of fused image Z_F is well improved, the complexity of the proposal algorithm could be an issue for mobile devices with limited computational resources.

VI. Conclusion Remarks and Discussion

A novel physics-driven deep exposure interpolation framework is proposed to interpolate an image with the middle exposure for two aligned large-exposure-ratio (LER) images in this paper. Experimental results show that the exposure interpolation-based algorithms outperform the data-driven algorithms subjectively and objectively. Therefore, the exposure interpolation is indeed required by the high dynamic range (HDR) imaging on top of two aligned LER images which are captured by the

emerging HDR imaging devices. The proposed algorithm can be extended to the case of three or even more aligned LER images.

The proposed framework can also be applied to study HDR and 3D imaging simultaneously [53]. Existing works on 3D imaging include stereo matching [53] ignoring saturations in input(s). Unfortunately, there are saturated regions in real-world applications. PSNR, SSIM, and FSIM are commonly used metrics for assessing image quality, but they do not effectively capture the specific issue of brightness order reversal. It is desired to introduce a new evaluation metrics which can more directly quantify the success of the proposed method in addressing this problem and provide evidence of its advanced nature. These problems will be studied in our future research.

APPENDIX: ANALYSIS OF THE PROPOSED FRAMEWORK

θ is the vector of all network parameters, and θ_t is the time-varying parameters, and η_t is the learning rate. $f_t(Z_1, Z_3)$ also denotes the corresponding vector. Similar to [51], it can be derived that

$$\begin{cases} \dot{\theta}_t = -\eta_t (\nabla_\theta f_t(Z_1, Z_3))^T \nabla_{f_t(Z_1,Z_3)} L \\ \dot{f}_t(Z_1, Z_3) = -\eta_t \Theta_t \nabla_{f_t(Z_1,Z_3)} L \end{cases}, \tag{30}$$

where the matrix Θ_t is $\nabla_\theta f_t(Z_1, Z_3)(\nabla_\theta f_t(Z_1, Z_3))^T$.

Without loss of generality, the loss function L is defined by the l_2 norm [51] and the learning rate η is fixed, it can be derived that

$$\dot{f}_t(Z_1, Z_3) = -\eta \Theta_t (f_t(Z_1, Z_3) - Z_2). \tag{31}$$

Suppose that the physics-driven method $f_m(Z_1, Z_3)$ is accurate. It then corresponds to the data-driven approach $f_{d,t}(Z_1, Z_3)$ with an initial θ_0 which is in a small region of the optimal θ^*. With the first-order Taylor expansion, the proposed neural augmentation framework (2) and (3) can be approximated by

$$\begin{aligned} f_t(Z_1, Z_3) =& f_m(Z_1, Z_3) + f_{d,t}(Z_1, Z_3) \\ \approx& f_m(Z_1, Z_3) + \nabla_\theta f_{d,t}(Z_1, Z_3)|_{\theta=\theta_0}(\theta - \theta_0). \end{aligned} \tag{32}$$

Let the matrix $\nabla_\theta f_{d,t}(Z_1, Z_3)(\nabla_\theta f_{d,t}(Z_1, Z_3))^T|_{\theta=\theta_0}$ be denoted as Θ_0 which is constant and symmetric positive semidefinite. It follows that

$$\begin{aligned} f_{d,t}(Z_1, Z_3) =& \exp(-\eta \Theta_0 t)(f_m(Z_1, Z_3) - Z_2) \\ &+ Z_2 - f_m(Z_1, Z_3). \end{aligned} \tag{33}$$

Since $\|f_m(Z_1, Z_3) - Z_2\|$ is smaller than $\|Z_2\|$, the proposed neural augmentation framework converges to the optimal solution faster than the corresponding data-driven approach [52].

The complexity of the model $f_m(Z_1, Z_3)$ is also required to be as low as possible so as to reduce the complexity of the proposed framework. Overall, the model $f_m(Z_1, Z_3)$ is required to be *as accurate as possible while as simple as possible*. Therefore, elegant physics-driven methods are highly demanded by the proposed neural augmentation framework.

REFERENCES

[1] G. Somanath and D. Kurz, "HDR environment map estimation for real-time augmented reality", in *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 11293-11301, 2021.

[2] A. R. Varkonyi-Koczy, A. Rovid, and T. Hashimoto, "Gradient-based synthesized multiple exposure time color HDR image", IEEE Trans. Instrum. Meas., vol. 57, no. 8, pp. 1779–1785, Aug. 2008.

[3] X. Tan, H. Chen, K. Xu, C. Xu, Y. Jin, and C. Zhu, "High dynamic range imaging for dynamic scenes with large-scale motions and severe saturation", IEEE Trans. Instrum. Meas., vol. 71, pp. 1–15, 2022.

[4] Y. Chen, G. Jiang, M. Yu, et al, "Learning stereo high dynamic range imaging from a pair of cameras with different exposure parameters", IEEE Trans. on Computational Imaging, vol. 6, pp. 1044-1058, 2020.

[5] P. E. Debevec and J. Malik. "Recovering high dynamic range radiance maps from photographs", in *Special Interest Group on Graphics and Interactive Techniques (SIGGRAPH)*, pp. 369-378 1997.

[6] Z. Liu, W. J. Lin, X. P. Li, Q. Rao, T. Jiang, M. Y. Han, H. Q. Fan, J. Sun, and S. C. Liu, "ADNet: attention-guided deformable convolutional network for high dynamic range imaging", in *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 463-470, Jun. 2021.

[7] S. Catley-Chandar, T. Tanay, L. Vandroux, A. Lenoardis, G. Slabaugh, and E. Perez-Pellitero, "FlexHDR: modeling alignment and exposure uncertainties for flexible HDR imaging", IEEE Trans. on Image Processing, vol. 31, pp. 5923-5935, 2022.

[8] 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, vol. 33, no. 10 , pp. 5332-5344, Oct. 2023.

[9] D. Tocci, C. Kiser, N. Tocci, and P. Sen, "A versatile HDR video production system", in *Special Interest Group on Graphics and Interactive Techniques (SIGGRAPH)*, pp. 1-9, 2011.

[10] 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, 2010.

[11] 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. 4255-4270, Aug. 2022.

[12] S. Tanaka, T. Otaka, K. Mori, N. Yoshimura, S. Matsuo, H. Abe, N.Yasuda, K. Ishikawa, S. Okura, S. Ohsawa, T. Akutsu, W. Fu, H. Chien, K. Liu, Y. Tsai, S. Chen, L. Teng, and I. Takayanagi, "Single exposure type wide dynamic range CMOS image sensor with enhanced nir sensitivity", ITE Trans. on Media Technology and Applications, vol. 6, no. 7, pp. 195-201, Jul. 2018.

[13] 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.

[14] K. Ram Prabhakar, V. Sai Srikar, and R. Venkatesh Babu, "Deepfuse: a deep unsupervised approach for exposure fusion with extreme exposure image pairs", in *IEEE International Conference on Computer Vision*, pp. 4724-4732, 2017.

[15] C. Zheng, W. Ying, S. Wu and Z. Li, "Neural Augmentation-Based Saturation Restoration for LDR Images of HDR Scenes", IEEE Trans. on Instrumentation and Measurement, vol. 71, pp. 1–11, 2023.

[16] T. Mertens, J. Kautz, and F. Van 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.

[17] F. Kou, Z. Li, C. Wen, and W. Chen, "Edge-preserving smoothing pyramid based multi-scale exposure fusion", Journal of Visual Communication and Image Representation, vol. 58, no. 4, pp. 235-244, Apr. 2018.

[18] G. W. Xu, Y. J. Wang, J. W. Gu, T. F. Xue, and X. Yang, "HDRFlow: Real-Time HDR Video Reconstruction with Large Motions", in *IEEE Conference on Computer Vision and Pattern Recognition*, 2024.

[19] Q. Wang, W. Chen, X. Wu and Z. Li, "Detail-enhanced multi-scale exposure fusion in YUV color space", IEEE Trans. on Circuits and Systems for Video Technology, vol. 30, no. 8, pp. 2418-2429, Aug. 2020.

[20] K. Ma, Z. Duanmu, H. Yeganeh, and Z. Wang, "Multi-exposure image fusion by optimizing a structural similarity index", IEEE Trans. on Computational Imaging, vol., no. 1, pp. 60-72, Jan 2017.

[21] H. Naila and M. Imran, "Ghost-free multi exposure image fusion technique using dense SIFT descriptor and guided filter", Journal of Visual Communication and Image Representation, vol. 62, pp. 295-308, 2019.

[22] Z. Yang,Y. Chen, Z. Le, and Y. Ma, "GANFuse: a novel multi-exposure image fusion method based on generative adversarial networks", Neural Computing and Applications, vol. 33, pp. 6133-6145, 2021.

[23] O. Ulucan, D. Ulucan, and M. Turkan, "Ghosting-free multi-exposure image fusion for static and dynamic scenes", Signal Processing, vol. 202. 108774, 2023.

[24] 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, 2019.

[25] 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.

[26] D. Han, L. Li, X. Guo, and J. Ma, "Multi-exposure image fusion via deep perceptual enhancement", Information Fusion, vol. 79, pp. 248-262, 2022.

[27] Z. G. Li, C. B. Zheng, J. H. Zheng, and S. Q. Wu, "Neural augmented exposure interpolation for HDR imaging", in *IEEE International Conference on Image Processing*, pp. 171-175, Malaysia, Oct. 2023.

[28] N. Rahaman, A. Baratin, D. Arpit, F. Draxler, M. Lin, F. Hamprecht, Y. Bengio, and A. Courville, "On the spectral bias of neural networks", in *International Conference on Machine Learning*, pp. 5301-5310, 2019.

[29] R. Basri, D. Jacobs, Y. Kasten, and S. Kritchman, "The convergence rate of neural networks for learned functions of different frequencies", in *Annual Conference on Neural Information Processing Systems*, 2019.

[30] A. Jacot, F. Gabriel, and C. Hongler, "Neural tangent kernel: convergence and generalization in neural networks", in *Annual Conference on Neural Information Processing Systems*, 2018.

[31] Ian J. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial nets", in *Advances in neural information processing systems*, pp. 2672-2680, 2014.

[32] S. Wang, O. Wang, R. Zhang, A. Owens, and Alexei A. Efros, "CNN-generated images are surprisingly easy to spot... for now", in *2020 IEEE Conference on Computer Vision and Pattern Recognition*, pp. 8692-8701, 2020.

[33] E. Reinhard, M. Ashikhmin, B. Gooch, and P. Shirley, "Color transfer between images", IEEE Computer Graphics and Applications, vol. 21, no. 5, pp. 34–41, May 2001.

[34] D. Ulyanov, A. Vedaldi, and V. Lempitsky, "Instance normalization: the missing ingredient for fast stylization", in *IEEE Conference on Computer Vision and Pattern Recognition*, 2018.

[35] H. Xu, J. Ma, Z. Le, J. Jiang, and X. Guo, "FusionDN: a unified densely connected network for image fusion", in *Thirty-Fourth AAAI Conference on Artifical Intelligence*, 2020.

[36] D. Karakaya, O. Ulucan and M. Turkan, "PAS-MEF: multi-exposure image fusion based on principal component analysis, adaptive well-exposedness and saliency map", in *the International Conference on Acoustics, Speech, and Signal Processing*, pp. 2345-2349, 2022.

[37] N. Shlezinger, J. Whang, C. Eldar, and G. Dimakis, "Model-based deep learning: key approaches and design guidelines", pp. 1-6, in *IEEE Data Science Workshop*, 2021.

[38] 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.

[39] S. Waqas Zamir, A. Arora, S. Khan, M.Hayat, F. Khan, M. Yang, and L. Shao, "CycleISP: real image restoration via improved data synthesis", in *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 2693-2702, 2020.

[40] K. Ma, K. Zeng, and Z. Wang, "Perceptual quality assessment for multiexposure image fusion", IEEE Trans. on Image Processing, vol. 24, no. 11, pp. 3345-3356, Nov. 2015.

[41] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition", in *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 770–778, 2019.

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

[43] N. Khademi 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.

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

[45] Q. Yan, D. Gong, Q. Shi,A. Hengel, C. Shen, I. Reid, and Y. Zhang, "Attention-guided network for ghost-free high dynamic range imaging", in *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1751-1760, 2019.

[46] 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, 2020.

[47] N. Barakat, A. Nicholas Hone, and E. Darcie, "Minimal-bracketing sets for high-dynamic-range image capture", IEEE Trans. on Image Processing, vol. 17, no. 10, pp. 1864–1875, Oct. 2008.

[48] C. Zheng, Z. Li, Y. Yang, and S. 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.

[49] J. Y. Liu, G. Y. Wu, J. S. Luan, Z. Y. Jiang, R. S. Liu, and X. Fan, "HoLoCo: Holistic and local contrastive learning network for multi-exposure image fusion", Information Fusion, vol. 95, pp. 237-249, 2023.

[50] J. Cai, S. Gu, and L. Zhang, "Learning a deep single image contrast enhancer from multi-exposure images", IEEE Trans. on Image Processing, vol. 27, pp. 2049-2062, 2018.

[51] J. Lee, L. Xiao, S. Schoenholz, Y. Bahri, R. Novak, J. SohlDickstein, and J. Pennington, "Wide neural networks of any depth evolve as linear models under gradient descent", in *Conference on Neural Information Processing Systems*, 2019.

[52] 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.

[53] G. Xu, X. Wang, X. Ding, and X. Yang, "Iterative geometry encoding volume for stereo matching", in *IEEE Conference on Computer Vision and Pattern Recognition*, Jun. 2023.