iTM-Net: Deep Inverse Tone Mapping Using Novel Loss Function Based on Tone Mapping Operator

Yuma Kinoshita and Hitoshi Kiya
Tokyo Metropolitan University, Tokyo, Japan

***Abstract*—A novel inverse tone mapping network, called "iTM-Net", is proposed in this paper. For training iTM-Net, we also propose a novel loss function considering the pixel distribution of HDR images. In inverse tone mapping with CNNs, we first point out that training CNNs with a standard loss function causes a problem, due to the distribution of HDR images. To overcome the problem, the novel loss function non-linearly tone-maps target HDR images into LDR ones, on the basis of a tone mapping operator, and then the distance between the tone-mapped image and a predicted one is calculated. The proposed loss function enables us not only to normalize HDR images but also to distribute pixel values of HDR images, like LDR ones. Experimental results show that HDR images predicted by the proposed iTM-Net have higher-quality than HDR ones predicted by conventional inverse tone mapping methods including state-of-the-arts, in terms of both HDR-VDP-2.2 and PU encoding + MS-SSIM. In addition, compared with loss functions not considering the HDR pixel distribution, the proposed loss function is shown to improve the performance of CNNs.**

***Index Terms*—Inverse tone mapping, High dynamic range imaging, Loss function, Deep learning, Convolutional neural networks**

I. Introduction

The low dynamic range (LDR) of modern digital cameras is a major factor that prevents cameras from capturing images as well as human vision. This is due to the limited dynamic range that imaging sensors have. For this reason, the interest of high dynamic range (HDR) imaging has been increasing.

To generate an HDR image from a single LDR image, various research works on inverse tone mapping have so far been reported [1]–[12]. Since LDR images are distorted by sensor saturation and a non-linear camera response function (CRF), the objective of inverse tone mapping can be separated into two issues: saturation recovery and linearization. Traditional ways for inverse tone mapping are based on expanding the dynamic range of input LDR images by using a fixed function or a specific parameterized function [1]–[7]. However, inverse tone mapping without prior knowledge is generally an ill-posed problem because of the following two reasons: pixel values might be lost by sensor saturation, and a CRF used for photographing is unknown. Hence, HDR images produced by these methods have limited quality. To obtain high-quality HDR images, inverse tone mapping methods based on deep learning have recently attracted attention.

Several convolutional neural network (CNN) based inverse tone mapping methods have so far been proposed [10]–[12]. The CNN-based methods significantly improve the performance of inverse tone mapping. In [11] and [10], CNNs are utilized for saturation recovery, but these methods employ a linearization method that is not based on deep learning. Although Marnerides et al. [12] tackled the linearization problem by training a CNN with simply normalized HDR images by using the min-max normalization, the performance is still limited because most pixel values of the normalized images distributed in a narrow range. This is due to the non-linear relation between LDR and HDR images.

Thus, in this paper, we propose a novel inverse tone mapping network, called iTM-Net. Similarly to [12], we aim to obtain relative luminance by linearizing LDR images with iTM-Net. To realize this, we also propose a novel loss function considering the non-linear relation between LDR and HDR ones. In the novel loss function, target HDR images are tone-mapped into LDR ones by using an invertible tone mapping operator, and then the distance between the tone-mapped image and a predicted one is calculated. The proposed loss function enables us not only to normalize HDR images but also to distribute pixel values of HDR images, like LDR ones.

In an experiment, the proposed method was compared with state-of-the-art inverse tone mapping methods. Experimental results illustrate that the proposed method outperforms the conventional methods in terms of two objective quality metrics: HDR-VDP-2.2 and PU encoding + MS-SSIM. In addition, the proposed loss function is shown to improve the performance of CNNs, compared with standard loss functions not considering the non-linear relation.

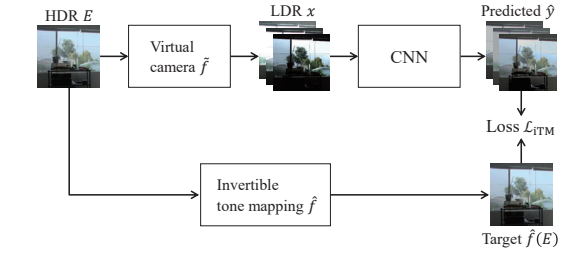
II. Related work

The goal of inverse tone mapping is to restore absolute or relative luminance of a scene, from a single LDR image. Since LDR images are distorted by sensor saturation and a non-linear CRF, this objective can be separated into two issues: saturation recovery and linearization.

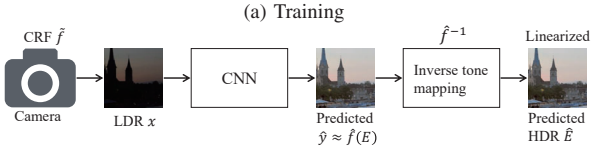
Many inverse tone mapping methods have already been studied [1]–[7], [10]–[12]. Especially in those methods, CNN-based methods [10]–[12] have recently attracted attention because of their effectiveness. Eilertsen et al. [11] aim to reconstruct saturated areas in input LDR images via a CNN. Predicted pixel values are combined with an input LDR image, which is linearized by using a fixed function not considering CRFs, in order to produce an HDR image. Endo et al. [10] have proposed a CNN based method that produces a stack of differently exposed images from a single LDR image. The

This work was supported by JSPS KAKENHI Grant Number JP18J20326.

978-9-0827-9703-9/19/$31.00 ©2019 IEEE



(a) Training



(b) Predicting

Fig. 1. Proposed inverse tone mapping

generated images are linearized and fused by using an existing stack-based method such as Debevec's method [13]. These two methods enable us to recover saturated regions in images, but the linearization problem still remains.

In the work by Marnerides et al. [12], they tackled the linearization problem, and sought to directly produce HDR images by a CNN. For calculating prediction loss in training the CNN, all HDR images are simply normalized into the range of $[0, 1]$, by using the min-max normalization. However, the use of the normalized HDR images for calculating prediction loss causes a problem that most pixel values of the normalized images are distributed in a narrow range. This is due to the non-linear relation between LDR and HDR images, and so the image statistics of LDR and HDR ones differ considerably as pointed out in [11]. Therefore, we aim to improve the performance of CNN-based inverse tone mapping, by using a novel loss function considering the non-linear relation for learning HDR images.

III. Proposed inverse tone mapping

Figure 1 shows an overview of our training procedure and predicting procedure. In the training, all input LDR images x are generated from target HDR images E by using various virtual cameras $\tilde{f}$ [11]. For calculating loss between a predicted image $\hat{y}$ and a target HDR one E, a tone mapping function $\hat{f}$, which is generally a non-linear one, is applied to E.

After the training, various LDR images are applied to the proposed CNN as input images, where the CNN then predicts tone-mapped versions of HDR images. The linearization is done by mapping the predicted images $\hat{y}$ by inverse tone mapping function $\hat{f}^{-1}$. Detailed training conditions are described in Section III-D.

A. Loss function

In literature [12], a loss function for training a CNN is defined by using the L_1-distance $\mathcal{L}_1$ and the cosine similarity $\mathcal{L}_{\cos}$. $\mathcal{L}_1$ and $\mathcal{L}_{\cos}$ are calculated as:

$$\mathcal{L}_1(\hat{y}, E) = \frac{1}{P}\sum_{i,j} \|E_{i,j} - \hat{y}_{i,j}\|_1, \tag{1}$$

$$\mathcal{L}_{\cos}(\hat{y}, E) = 1 - \frac{1}{P}\sum_{i,j} \frac{E_{i,j} \cdot \hat{y}_{i,j}}{\|E_{i,j}\|_2 \|\hat{y}_{i,j}\|_2}, \tag{2}$$

where $E_{i,j}$ and $\hat{y}_{i,j}$ denote an RGB pixel vector at pixel (i, j) in HDR image E and predicted image $\hat{y}$, respectively, and P is the total number of pixels. By using eqs. (1) and (2), the loss function utilized for ExpandNet [12] is given by

$$\mathcal{L}_{\mathrm{Expand}}(\hat{y}, E) = \mathcal{L}_1(\hat{y}, m(E)) + \lambda \mathcal{L}_{\cos}(\hat{y}, m(E)), \tag{3}$$

where $m(E)$ denotes the min-max normalization which simply normalizes E into the range of $[0, 1]$, and λ is a constant factor that adjusts the contribution of the cosine similarity. However, the min-max normalization is unsuitable for learning HDR images because pixel values of HDR images are non-uniformly distributed in an extremely wide range [11], unlike LDR ones.

For this reason, we utilize an invertible tone mapping operator $\hat{f}(\cdot)$, which is designed to transform HDR images into LDR ones, instead of the min-max normalization $m(\cdot)$. For example, the L_1-distance with $\hat{f}(\cdot)$ is calculated by

$$\mathcal{L}_{\mathrm{iTM}}(\hat{y}, E) = \mathcal{L}_1(\hat{y}, \hat{f}(E)). \tag{4}$$

In this paper, Reinhard's global operator [14] is utilized as $\hat{f}(\cdot)$, where the operator transforms HDR images into high-quality LDR ones, and it has an inverse function. By using the luminance matrix L_E of E, the operator is given by the following equations:

$$\hat{f}(E) = (\hat{g}(L_E) \oslash L_E) \odot E, \tag{5}$$

$$\hat{g}(L_E) = L_X \oslash (1 + L_X), \tag{6}$$

$$L_X = \frac{a}{G(L_E)} L_E. \tag{7}$$

where $\odot$ and $\oslash$ mean pixel-wise multiplication and division, respectively. The parameter $a \in [0, 1]$ determines the brightness of an output image $\hat{f}(E)$, and $G(L_E)$ is the geometric mean of L_E given by

$$G(L_E) = \exp\left(\frac{1}{P}\sum_{i,j} \log\left(\max\left(L_{Ei,j}, \epsilon\right)\right)\right), \tag{8}$$

where ϵ is a small value to avoid singularities at $L_{Ei,j} = 0$. Eq. (7) enables us to calibrate HDR images by adjusting the geometric mean of each HDR image to a, and eq. (6) allows us to distribute pixel values of HDR images, like those of LDR ones. Since $\hat{f}$ is invertible, HDR images can be predicted by using an inverse tone mapping operator $\hat{f}$, as shown in the next Section.

B. Prediction

The proposed CNN generates tone-mapped versions of HDR images E because the CNN is trained by using the loss function shown in eq. (4). Hence, predicting HDR images is

done by applying the inverse tone mapping function $\hat{f}^{-1}$ to predicted image $\hat{y}$, as

$$\hat{E} = \hat{f}^{-1}(\hat{y}) = (\hat{g}^{-1}(L_{\hat{y}}) \oslash L_{\hat{y}}) \odot \hat{y}, \quad (9)$$

$$\hat{g}^{-1}(L_{\hat{y}}) = L_{\hat{y}} \oslash (1 - L_{\hat{y}}), \quad (10)$$

where $L_{\hat{y}}$ is the luminance of $\hat{y}$. Note that eq. (7) can be ignored in the inverse tone mapping since our goal is to obtain relative luminance.

C. iTM-Net architecture

Figure 2 shows the overall network architecture of iTM-Net. The architecture consists of three networks: a local encoder, a global encoder, and a decoder. The local encoder and the decoder in the proposed method are almost the same as those used in U-Net [15]. Concatenated skip connections between the local encoder and the decoder are also utilized like in U-Net. Although U-Net works very well for various image-to-image translation problems, its use for inverse tone mapping often causes distortions in output images due to its network architecture not handling global image information. For this reason, we utilize the global encoder and combine features extracted by both encoders to prevent the distortions. The input for the local encoder is a $P = H \times W$ pixels 24-bit color LDR image. For the global encoder, the input image is resized to a fixed size (128×128).

iTM-Net has five types of layers as shown in Fig. 2:

- 3×3 **Conv.+ BN + ReLU** which calculates a 3×3 convolution with a stride of 1 and a padding of 1. After convolution, batch normalization [16] and the rectified linear unit activation function [17] (ReLU) are applied. In the local encoder and the decoder, two adjacent 3×3 Conv.+ BN + ReLU layers will have the same number K of filters. From the first two layers to the last ones, the numbers of filters are $K = 32$, 64, 128, 256, 512, 256, 128, 64, and 32, respectively. In the global encoder, all layers have 64 filters.
- 2×2 **Max pool** which downsamples feature maps by max pooling with a kernel size of 2×2 and a stride of 2.
- 4×4 **Transposed Conv. + BN + ReLU** which calculates a 4×4 convolution with a stride of $1/2$ and a padding of 1. After convolution, BN and ReLU are applied. From the first layer to the last one, the numbers of filters are $K = 256$, 128, 64, and 32, respectively.
- 1×1 **Conv. + ReLU** which calculates a 1×1 convolution with a stride of 1 and a padding of 1. After convolution, ReLU is applied. The number of filters in the layer is 3.
- 4×4 **Conv. + BN + ReLU (w/o padding)** which calculates a 4×4 convolution without padding. The number of filters in the layer is 64.

D. Training

Numerous LDR images taken under various conditions, x, and corresponding HDR images, E, are needed to train iTM-Net. To prepare a sufficient amount of training data, we utilize various virtual cameras to generate x from HDR images E [11]. For training, 831 HDR images were collected from online available databases [18]–[23].

The training procedure of our CNN is shown as follows.

i Select 16 HDR images from the 831 HDR images at random.

ii Generate 16 pairs of an input image and its target one (x, $\tilde{E}$) from each HDR image. Each pair is generated in accordance with the following steps.

(a) Crop HDR image E to an image patch $\tilde{E}$ at $N \times N$ pixels. The size N is given as a product of a uniform random number in the range $[0.2, 0.6]$ and the length of the short side of E. In addition, the position of the patch in E is also determined at random.

(b) Resize $\tilde{E}$ to 256×256 pixels.

(c) Flip $\tilde{E}$ horizontally or vertically with a probability of 0.5.

(d) Calculate exposure X from $\tilde{E}$ by $X = \Delta t(v) \cdot \tilde{E}$, where pixel values larger than 1 are clipped. Shutter speed Δt is calculated as $\Delta t(v) = 0.18 \cdot 2^v / G(L_{\tilde{E}})$ as in [14] by using a uniform random number v in the range $[-2, 2]$. $G(L_{\tilde{E}})$ is the geometric mean of luminance of $\tilde{E}$.

(e) Generate an input LDR image x from X by using a virtual camera $\tilde{f}$, as

$$x = \tilde{f}(X) = (\tilde{g}(L_X) \oslash L_X) \odot X, \quad (11)$$

$$\tilde{g}(L_X) = (1 + \eta)(L_X^\gamma \oslash (L_X^\gamma + \eta)), \quad (12)$$

where η and γ are random numbers that follow normal distributions with a mean of 0.6 and a variance of 0.1 and with a mean of 0.9 and a variance of 0.1, respectively. L_X is luminance of X, and exponentiation L_X^γ is calculated as a pixel-wise operation.

iii Predict 16 LDR images $\hat{y}$ from 16 input LDR images x by using iTM-Net.

iv Evaluate loss between predicted images $\hat{y}$ and target images $\tilde{E}$ by using eq. (4) with Reinhard's global operator $\hat{f}$. Here, $a = 0.18$ is used in (7).

v Update filter weights ω and biases b in the CNN by back-propagation.

In our experiments, iTM-Net were trained with 500 epochs, where the above procedure was repeated 51 times in each epoch. In addition, each HDR image had only one chance to be selected, in Step i in each epoch. He's method [24] was used for initializing of iTM-Net. In addition, the Adam optimizer [25] was utilized for optimization, where parameters in Adam were set as $\alpha = 0.002$, $\beta_1 = 0.9$, and $\beta_2 = 0.999$.

IV. Simulation

We evaluated the effectiveness of the proposed method by using two objective quality metrics.

A. Simulation conditions

The quality of generated HDR images $\hat{E}$ by using iTM-Net was evaluated by two objective quality metrics: HDR-VDP-2.2 [26], and PU encoding [27] with MS-SSIM [28] which

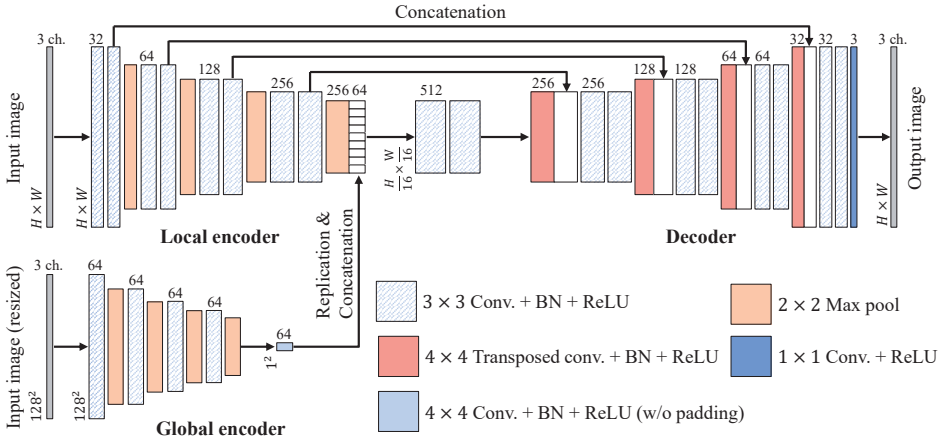


Fig. 2. Network architecture. Architecture consists of local encoder, global encoder, and decoder. Each box denotes multi-channel feature map produced by each layer. Number of channels is denoted above each box. Feature map resolutions are denoted to the left of boxes.

utilize an original HDR image E as a reference. Literature [29] have shown that these metrics are suitable for evaluating the quality of HDR images.

The two metrics are designed for evaluating the difference between two HDR images that have absolute luminance of a scene. Such HDR images, namely HDR ones having absolute luminance, are only in the dataset [20]. Hence, 20 HDR images selected from the dataset [20] were used for the experiment. Note that they were not used for training. Input LDR images x were generated in accordance with Steps ii(d) and ii(e) in Section III-D. In addition, predicted HDR images $\hat{E}$ were scaled to match the range of $\hat{E}$ with that of the original HDR image E because inverse tone mapping methods can only predict HDR images having relative luminance.

The proposed method is compared with four methods including the state-of-the-art ones: direct inverse tone mapping operator (Direct ITMO) [7], pseudo-multi-exposure-based tone fusion (PMET) [5], ExpandNet [12], iTM-Net trained by the standard L_1-loss $\mathcal{L}_1(\hat{y}, m(E))$ without tone mapping (Proposed w/o TM in Tables I and II), where the third and fourth methods are CNN-based ones, but the other methods are not based on machine learning.

B. Results

Tables I and II illustrate experimental results in terms of HDR-VDP and PU encoding + MS-SSIM, respectively. A larger value for both metrics means higher similarity between a predicted HDR image and its original HDR image. As shown in Tables I and II, the proposed iTM-Net provided the highest scores for the two metrics in the five methods. Therefore, the proposed method outperformed the conventional methods in terms of both HDR-VDP and PU-encoding + MS-SSIM. These results indicate that HDR images predicted by iTM-Net were more similar to original HDR images than those predicted by the other methods. Since all predicted HDR images scaled to match original HDR images, the results illustrate that the proposed method can linearize images with high-quality. By comparing with iTM-Net trained with the standard L_1-loss, the proposed iTM-Net produced higher scores for both metrics.

TABLE I
HDR-VDP-2.2 SCORES. "SD" DENOTES THE STANDARD DEVIATION.

Scene	Direct ITMO [7]	PMET [5]	ExpandNet [12]	Proposed w/o TM	Proposed
DelicateArch	40.94	43.85	59.67	60.58	**71.87**
GoldenGate(1)	31.31	31.31	46.95	48.70	**49.05**
HooverDam	28.50	28.50	46.93	47.94	**48.52**
JesseBrownsCabin	31.89	31.89	39.44	36.79	**39.65**
WestBranchAusable(2)	33.18	33.18	**52.19**	48.30	51.77
Average (20 images)	34.26	34.67	48.47	49.69	**52.10**
SD (20 images)	5.86	6.83	5.93	8.43	9.94

TABLE II
PU ENCODING + MS-SSIM SCORES. "SD" DENOTES THE STANDARD DEVIATION

Scene	Direct ITMO [7]	PMET [5]	ExpandNet [12]	Proposed w/o TM	Proposed
DelicateArch	0.7289	0.7702	0.9896	0.9741	**0.9960**
GoldenGate(1)	0.6335	0.6335	0.9489	0.9576	**0.9759**
HooverDam	0.5350	0.5350	0.8910	0.8969	**0.9114**
JesseBrownsCabin	0.2577	0.2577	**0.8851**	0.6490	0.8508
WestBranchAusable(2)	0.2264	0.2264	0.8942	0.8908	**0.9760**
Average (20 images)	0.4215	0.4256	0.9017	0.8964	**0.9569**
SD (20 images)	0.2127	0.2207	0.0593	0.0843	0.0410

Hence, the proposed loss function is effective to train CNNs for inverse tone mapping.

Figure 3 shows examples of HDR images generated by the four methods. Here, these images were tone-mapped from predicted HDR images because HDR images cannot be displayed in commonly-used LDR devices. From Fig. 3, it is confirmed that the proposed method produced a higher-quality HDR image which similar to an original HDR one $\tilde{E}$, than other methods. These figures also show that methods not based on machine learning caused images to be significantly distorted.

For these reasons, it is showed that the proposed method is effective to generate high-quality HDR images from single LDR images. Especially, the use of the proposed loss function enables us to improve the performance of CNNs for inverse tone mapping.

V. CONCLUSION

In this paper, a novel inverse tone mapping network, called iTM-Net, was proposed. For training iTM-Net, a novel loss

(a) Input x (b) Direct ITMO [7] (c) PMET [5] (d) ExpandNet [12] (e) Proposed (f) Ground truth $\tilde{E}$

Fig. 3. Experimental Results [WestBranchAusable(2)]. HDR images (b)-(f) were tone-mapped for visualizing.

function considering the non-linear relation between HDR and LDR images was also proposed. In the proposed loss function, target HDR images are tone-mapped into LDR images by an invertible tone mapping operator. The use of the proposed loss function enables us not only to normalize HDR images, but also to distribute pixel values of HDR images, like those of LDR ones. As a result, the performance of CNNs for inverse tone mapping can be improved. Experimental results showed that HDR images predicted by iTM-Net trained with the proposed loss function have higher-quality than HDR ones predicted by conventional methods including the state-of-the-arts, in terms of HDR-VDP-2.2 and PU encoding + MS-SSIM. In addition, it was also confirmed that the proposed loss function improves the performance of CNNs, compared with loss functions not considering the non-linear relation.

REFERENCES

[1] F. Banterle, P. Ledda, K. Debattista, and A. Chalmers, "Inverse tone mapping," in *Proc. Int. Conf. Comput. Graph. Interact. Tech. Australas. Southeast Asia*, Nov./Dec. 2006, pp. 349–356.

[2] A. G. Rempel, M. Trentacoste, H. Seetzen, H. D. Young, W. Heidrich, L. Whitehead, and G. Ward, "Ldr2Hdr: on-the-fly reverse tone mapping of legacy video and photographs," *ACM Trans. Graph.*, vol. 26, no. 3, pp. 39:1–39:6, Jul. 2007.

[3] P.-H. Kuo, C.-S. Tang, and S.-Y. Chien, "Content-adaptive inverse tone mapping," in *Proc. Vis. Commun. Image Process.*, Nov. 2012, pp. 1–6.

[4] H. Youngquing, Y. Fan, and V. Brost, "Dodging and burning inspired inverse tone mapping algorithm," *J. Comput. Inf. Syst.*, vol. 9, no. 9, pp. 3461–3468, May 2013.

[5] T.-H. Wang, C.-W. Chiu, W.-C. Wu, J.-W. Wang, C.-Y. Lin, C.-T. Chiu, and J.-J. Liou, "Pseudo-multiple-exposure-based tone fusion with local region adjustment," *IEEE Trans. Multimed.*, vol. 17, no. 4, pp. 470–484, Apr. 2015.

[6] Y. Kinoshita, S. Shiota, and H. Kiya, "Fast inverse tone mapping with Reinhard's global operator," in *Proc. IEEE Int. Conf. Acoust. Speech Signal Process.*, Mar. 2017, pp. 1972–1976.

[7] Y. Kinoshita, S. Shiota, and H. Kiya, "Fast inverse tone mapping based on Reinhard's global operator with estimated parameters," *IEICE Trans. Fundam. Electron. Commun. Comput. Sci.*, vol. E100.A, no. 11, pp. 2248–2255, Nov. 2017.

[8] Y. Kinoshita and H. Kiya, "Automatic exposure compensation using an image segmentation method for single-image-based multi-exposure fusion," *APSIPA Trans. Signal Inf. Process.*, vol. 7, no. e22, Dec. 2018.

[9] Y. Kinoshita and H. Kiya, "Scene Segmentation-Based Luminance Adjustment for Multi-Exposure Image Fusion," *IEEE Trans. Image Process.*, to be published. doi: 10.1109/TIP.2019.2906501

[10] Y. Endo, Y. Kanamori, and J. Mitani, "Deep reverse tone mapping," *ACM Trans. Graph.*, vol. 36, no. 6, pp. 177:1–177:10, Nov. 2017.

[11] G. Eilertsen, J. Kronander, G. Denes, R. K. Mantiuk, and J. Unger, "HDR image reconstruction from a single exposure using deep CNNs," *ACM Trans. Graph.*, vol. 36, no. 6, pp. 1–15, Nov. 2017.

[12] 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," *Comput. Graph. Forum*, vol. 37, no. 2, pp. 37–49, May 2018.

[13] P. E. Debevec and J. Malik, "Recovering high dynamic range radiance maps from photographs," in *Proc. ACM SIGGRAPH*, Aug. 1997, pp. 369–378.

[14] 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.

[15] O. Ronneberger, P.Fischer, and T. Brox, "U-Net: convolutional networks for biomedical image segmentation," in *Med. Image Comput. Comput. Interv.*, ser. LNCS, vol. 9351. Springer, Nov. 2015, pp. 234–241.

[16] S. Ioffe and C. Szegedy, "Batch normalization: accelerating deep network training by reducing internal covariate shift," *arXiv Prepr. arXiv1502.03167*, pp. 1–11, Feb. 2015. [Online]. Available: http://arxiv.org/abs/1502.03167

[17] X. Glorot, A. Bordes, and Y. Bengio, "Deep sparse rectifier neural networks," in *Proc. Int. Conf. Artif. Intell. Stat.*, Apr. 2011, pp. 315–323.

[18] "GitHub - openexr." [Online]. Available: https://github.com/openexr/

[19] "High Dynamic Range Image Examples." [Online]. Available: http://www.anyhere.com/gward/hdrenc/pages/originals.html

[20] "The HDR Photographic Survey." [Online]. Available: http://rit-mcsl.org/fairchild/HDRPS/HDRthumbs.html

[21] "Max Planck institut informatik." [Online]. Available: http://resources.mpi-inf.mpg.de/hdr/gallery.html

[22] P. Zolliker, Z. Barańczuk, D. Küpper, I. Sprow, and T. Stamm, "Creating HDR video content for visual quality assessment using stop-motion," in *Proc. Eur. Signal Process. Conf.*, Sep. 2013, pp. 1–5.

[23] H. Nemoto, P. Korshunov, P. Hanhart, and T. Ebrahimi, "Visual attention in LDR and HDR images," in *Proc. 9th Int. Work. Video Process. Qual. Metrics Consum. Electron.*, Feb. 2015, pp. 1–6.

[24] K. He, X. Zhang, S. Ren, and J. Sun, "Delving deep into rectifiers: surpassing human-level performance on ImageNet classification," in *Proc. IEEE Int. Conf. Comput. Vis.*, Dec. 2015, pp. 1026–1034.

[25] D. P. Kingma and J. Ba, "Adam: a method for stochastic optimization," *arXiv Prepr. arXiv1412.6980*, pp. 1–15, Dec. 2014. [Online]. Available: http://arxiv.org/abs/1412.6980

[26] M. Narwaria, R. K. Mantiuk, M. P. Da Silva, and P. Le Callet, "HDR-VDP-2.2: a calibrated method for objective quality prediction of high-dynamic range and standard images," *J. Electron. Imaging*, vol. 24, no. 1, p. 010501, Jan. 2015.

[27] T. O. Aydın, R. Mantiuk, and H.-P. Seidel, "Extending quality metrics to full luminance range images," in *Proc. SPIE Hum. Vis. Electron. Imaging XIII*, Feb. 2008, pp. 68060B:1–68060B:10.

[28] Z. Wang, E. Simoncelli, and A. Bovik, "Multiscale structural similarity for image quality assessment," in *Proc. Asilomar Conf. Signals, Syst. Comput.*, vol. 2, Nov. 2003, pp. 1398–1402.

[29] P. Hanhart, M. V. Bernardo, M. Pereira, A. M. G. Pinheiro, and T. Ebrahimi, "Benchmarking of objective quality metrics for HDR image quality assessment," *EURASIP J. Image Video Process.*, vol. 2015, no. 39, pp. 1–18, Dec. 2015.