https://doi.org/10.2352/EI.2023.35.14.COIMG-156

A Lightweight Exposure Bracketing Strategy for HDR Imaging without Access to Camera Raw

Jieyu Li*[1] *, Ruiwen Zhen*[2] *, Robert L. Stevenson*[1] *;
***University of Notre Dame*[1] *, Notre Dame, In 46556; SenseBrain Technology, San Jose, CA 95131.*[2]**

Abstract

A lightweight learning-based exposure bracketing strategy is proposed in this paper for high dynamic range (HDR) imaging with no access to camera RAW. Some low-cost, power-efficient cameras, such as webcams, video surveillance cameras, sport cameras, mid-tier cellphone cameras, and navigation cameras on robots, can only provide access to 8-bit low dynamic range (LDR) images. Exposure bracketing is a classical approach to capture HDR scenes by fusing images taken with different exposures into an 8-bit tone-mapped HDR image. A key question is what the optimal set of exposure settings should be to cover the scene dynamic range and achieve a desirable tone. The proposed lightweight neural network predicts the exposure values for a 3-shot exposure bracketing, given the input irradiance information from 1) the histograms of an auto-exposure LDR preview image, and 2) the maximum and minimum levels of the scene irradiance. Without the processing of the preview image streams, and the circuitous route of first estimating the scene HDR irradiance and then tone-mapping to 8-bit images, the proposed method gives a more practical HDR enhancement for real-time and on-device applications. Experiments on a number of challenging images reveal the advantages of our method in comparison with other state-of-the-art methods qualitatively and quantitatively.

Introduction

HDR imaging is an essential feature for capturing photos of the real world, given the typical dynamic range of a scene is larger than 120 dB while the dynamic range of a camera is around 60 dB. With multiple images taken with different exposure settings, it is relatively easy to compose HDR in camera RAW domain [1, 2], where images are describing the scene irradiance without gamma correction, demosaicing, white balance, etc., but there are many cameras that have no access to camera RAW, such as mid-tier cellphone cameras, webcams, sport cameras, surveillance cameras and navigation cameras. Although one can perform radiometric calibration to recover camera response curves [3, 4, 5], it introduces extra errors and processing time. Additionally, RAW file size is typically 2–6 times larger than JPEG file size, so applying HDR enhancement with RAW images consumes more power resources and DRAM transaction bandwidth. It is thus desirable to directly fuse multiple 8-bit LDR images into one 8-bit tone-mapped HDR image for real-time and on-device applications.

For exposure bracketing strategies, a key question is what the optimal set of exposure settings are. Often the optimal set is scene dependent, and it is a trade-off between dynamic range, SNR, tone and motion artifacts. In addition to the existing exposure bracketing strategies for generating the noise-optimal HDR irradiance map in camera RAW domain [3, 6, 7, 8, 9], there are a few methods that focus on tone-optimal exposure strategy for direct fusion [10, 11, 12, 13].

In [12, 13], the exposure strategy is generated on local statistics of a single auto-exposed (AE) image, where irradiance information beyond its limited dynamic range is not observed or estimated, and as a result fail to avoid saturation. In [10], a neural network (NN) is applied to the exposure bracketing strategy for the first time, leveraging not only the local statistics but also the semantic information to solve the saturation problem. However, this model contains more than 60 million parameters, and needs to be trained separately for day and night scenes. In this paper, we propose a lightweight exposure bracketing strategy network, which can produce images with great details and pleasing tone for different scenes. The comparison of processing time and performance among the state-of-the-art methods is shown in Figure 1. With the selected exposure bracketing, final 8-bit HDR images are fused by a widely used fusion method [14]. The performance is evaluated based on the peak signal-to-noise ratio (PSNR) against the ground truth images, which are generated by fusing 10 differently exposed images for each scene. Such large and tightly packed bracketing can effectively avoid contrast decrease due to saturation or color distortion and halo artifact caused by large exposure gaps.

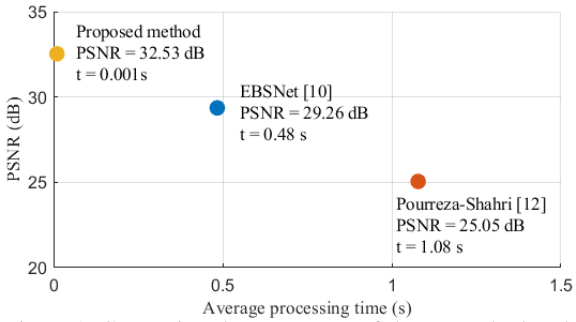

Figure 1: Comparison between state-of-the art methods. The proposed method achieves the highest PSNR against the ground truth and costs the minimum processing time (including the preprocessing of AE images and the prediction of exposure settings).

Instead of a heavy semantic branch [10], the proposed system takes the maximum and minimum scene irradiance along with the AE image as inputs to select the exposure settings covering the dynamic range. An image signal processor (ISP) normally changes the exposure time and measures the irradiance at some parts of the scene for auto-exposure (AE) and auto-focus (AF), so it could be capable to produce the minimum and maximum scene

irradiance without generating preview image streams or HDR histograms, which is time consuming but required in some other exposure strategy methods [6, 7, 8]. With an HDR RAW dataset simulating the scene irradiance, an image formation model and the exposure fusion algorithm [14] can be applied to generate the 8-bit HDR images. The model is trained to minimize the sum of square differences between the ground truth and final images. Weighed loss is introduced for imbalanced learning with emphasis on the tone of portraits and billboards. Experiments show that the proposed system can be generalized well to different scenes and takes about 1 ms for computation using a TESLA V100 GPU.

In summary, the main contributions are listed as follows:

- A lightweight network is developed for exposure strategy driven by the optimization of the tone of the fused results. The experiment validates that the system can generate fused images with desirable details and tone in HDR scenes.
- The model has only around 300K parameters and takes about 1 ms for computation using one TESLA V100 GPU, which is efficient for deployment in practical applications.

Related Work

Exposure Bracketing Strategy

The selection of exposure bracketing is made to cover and optimally sample the irradiance of a scene. Some existing methods are driven by the noise-optimal reconstruction of HDR irradiance map [3, 6, 15]. Because of the non-linearity of the cameras, brought by the optical characteristics of particular sensors and ISP, these methods depend on the access to RAW images or the estimation of camera response curve [3, 4, 5]. The full irradiance distribution of the scene, can be generated by combining a sequence of LDR data [2, 6, 8], and then the average SNR or the worst-case SNR can be minimized [1, 2, 6, 7, 8]. The estimated scene HDR irradiance then has to be tone mapped [16, 17, 18, 19] to 8-bit for display.

To avoid this circuitous route of first estimating the scene HDR irradiance and then tone-mapping the results, some other work focuses on selecting the exposure bracketing that has the optimal tone after direct fusion [10, 11, 12, 13]. These methods analyze scene information from an AE preview image, so that the processing of preview streams is not required. In [12], the AE image is segmented into dark, normal and bright regions, then correspondingly exposure settings for a 3-image stack are determined to match the mean luminance of each region to an optimal level (around $255/2$) according to some typical camera response curves. Without the estimation of the dynamic range, the method gives sub-optimal results with loss of details in saturation. Wang et al. [10] claim that the semantic information can be useful to infer the irradiance information beyond the preview image. For example, a billboard recognized in the images indicates larger maximum irradiance level and thus lower exposure is needed to capture its content. Based on this intuition, an exposure bracketing selection network (EBSNet) is developed with a histogram branch and a semantic branch, and allows better recovery of very dark/bright regions. However, this network contains more than 60M parameters and hard to be generalized to different scenarios, which might prevent the network from real-time applications.

Neural Network for HDR Imaging

Convolutional neural networks (CNN) are widely used for the generation and enhancement of HDR images. Some methods are developed to learn an HDR image from a single LDR image [20,21]. However, these algorithms can do well only when the captured LDR image contains sufficient information. So for image acquisition, learning based exposure strategies are designed. Onzon et al. [22] propose a network for auto-exposure control driven by object detection in HDR scenes. The network extracts the multi-scale histogram and semantic features from the previous frame, and is trained jointly, end-to-end with an object detector and an ISP pipeline. Wang et al. [10] apply CNNs to exposure bracketing selection (EBSNet). Similarly, the EBSNet has a histogram branch and a semantic branch based on AlexNet [23]. The EBSNet is trained by reinforcement learning to select the optimal exposure bracketing from 36 predefined candidates, which is limited and difficult for generalization. In the experiments, two models are trained with different candidate sets for night scenes and day scenes. We propose a lightweight exposure bracketing strategy network that contains only around 300K parameters and can predict exposure values for a 3-image stack under different scenarios.

CNNs are also introduced for multi-exposure fusion to improve the quality of fused images in challenging setups, such as extreme exposure image pairs [24] or images with ghosting artifacts [25,26]. These networks can be integrated into the proposed system, and trained jointly to adapt for different tone-mapped images as in the future work.

Proposed System

The system framework is illustrated in Figure 2. The network takes the AE preview LDR image, the minimum and maximum scene irradiance as inputs. A low-end ISP can easily produce the minimum and maximum scene irradiance during the AE/AF process, without generating preview image streams or HDR histograms. The network predicts the exposure values EV, as the product of the exposure time t and the camera gain g, for a 3-shot bracketing, that optimizes the tone of the fused image. For photo capturing, the exposure time is set to be the same as the AE settings, so the gain $g = EV/t$. In extreme cases, the computed gain is not practical and the exposure time will be adjusted.

Features

The exposure values are predicted based on the irradiance distribution learned from 1) the histogram statistics of an AE preview image and 2) the minimum and maximum levels of the scene irradiance. In order to use the irradiance information locally and globally, histograms at three different scales are computed. For the coarsest scale, a 128-bin histogram is computed on the whole image based on the luma channel. Similarly, at the intermediate scale and the finest scale, the image is divided respectively into 3 by 3 and 7 by 7 non-overlapped sub-images and the histograms are computed from every sub-image. The minimum and maximum irradiance levels are repeated and concatenated to each bins of the histogram stack, which gives an input to the neural network with shape [128, 61].

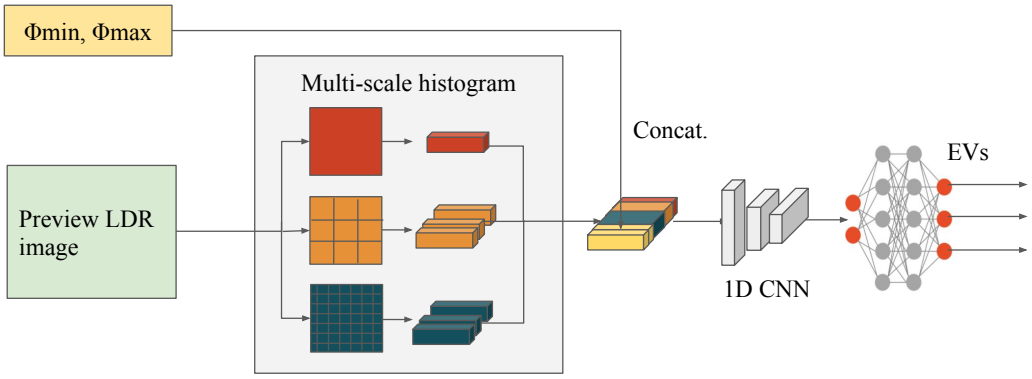

Figure 2: System overview.

Exposure Values Prediction

The feature is then sent to a neural network that consists three 1D convolutional layers with kernel size 4 and stride 4 and three fully connected layers, which predicts the exposure values, $\boldsymbol{EV} = \{EV_+, EV_n, EV_-\}$, relative to the AE settings for a 3-shot exposure bracketing. Similar to the work [22], the last layer is followed by a function that computes the final exposure value by

$$\boldsymbol{EV} = \exp\{2 \cdot [\text{sigmoid}(\boldsymbol{x}) - 0.5] \cdot \log(M_{\text{ev}})\} \tag{1}$$

where x is the output of last layer. The exposure change is bounded by $M_{\text{ev}} = 2^7$ as $EV \in [M_{\text{ev}}^{-1}, M_{\text{ev}}]$.

Loss Function

With HDR RAW images providing scene irradiance, the network is trained end-to-end with a differentiable image formation model simulating single-shot LDR image capturing without noise and an exposure fusion module generating final fused images. The pipeline for loss calculation is shown in Figure 3.

Under certain scene irradiance $\boldsymbol{\phi}$, $\boldsymbol{\phi}t$ electrons are collected over an exposure time of t seconds, and converted to a voltage. The voltage is amplified by camera gain g, and converted to a digital number as a pixel value. For a camera recording 12-bit RAW images, the pixels with intensity larger than $I_{\text{max}} = 2^{12} - 1$ are clipped due to saturation. The process can be described by the following equation.

$$\boldsymbol{I}_{\text{RAW}} = \min\{\lfloor g\boldsymbol{\phi}t \rfloor, I_{\text{max}}\}, \tag{2}$$

where $\lfloor \cdot \rfloor$ represents quantization.

The LDR RAW images are then fed into the software ISP, processed by demosaicing algorithm and gamma correction, to generate the 8-bit LDR images. The exposure value, as the output of the network, is the product of exposure time t and camera gain g. In practice, t is set to be the same as the AE settings, so the gain $g = EV/t$.

According to the process described above, different exposure settings result in different saturated regions and brightness in the LDR images. The 3-shot exposure bracketing is then fused directly in an 8-bit image by [14], where images are weighted according to their saturation, contrast, well-exposedness, and blended using pyramidal image decomposition. The loss is the sum of weighted squared error between the fused image and the ground truth, where the ground truth is generated by fusing a large LDR stack (e.g. 10 images) with geometric exposure variation.

$$\mathscr{L} = \sum_{p \in P} w_p (I_p - \hat{I}_p(\boldsymbol{g}, \boldsymbol{t}))^2. \tag{3}$$

For each pixel p, its weight w_p is given for the learning of imbalanced data where some small portions of the images are more important for perception, such as faces and billboards. Minimizing the loss ensures that the details are captured without saturation and with good quantization intervals.

Experiments

Dataset

We conduct experiments on two datasets: MIT-Adobe FiveK [27] and our night sight dataset. The MIT-Adobe FiveK dataset contains 5000 images in RAW format, but mostly for day scenes. We introduce a night sight dataset, including 300 HDR RAW images. The dataset contains a set of 24-bit Linear HDR images. Each of the HDR image is generated by fusing 8 LDR RAW images, which are processed by denoising and deghosting algorithms to remove potential noise and ghost artifacts. The two datasets cover different scenarios, including but not limited to scenes with and without people as well as indoor and outdoor.

From the HDR data, synthetic LDR images can be generated with different exposure settings according to Equation 4. The ground truth are created by fusing 10 differently exposed images [14]. With the exposure value of the AE preview image denoted as EV_a, the ten images are exposed by $[2^3, 2^2, 2, 1, 2^{-1}, 2^{-2}, 2^{-3}, 2^{-4}, 2^{-5}, 2^{-6}] \cdot EV_a$.

In order to get the data distributed evenly for night and day scenes, 400 images of day scenes and 100 images of night scenes are selected randomly from MIT-Adobe FiveK dataset [27], and combined with the 300 night sight data to form the dataset for the experiments. The dataset is then split into 600 images for training, 100 images for evaluation and 100 images for testing.

Experiment Setups

For training, the HDR RAW images in the dataset are used to simulate the real-world irradiance. As the network inputs, the resolution of the AE preview images is 384×510, and the minimum and maximum scene irradiance are estimated respectively

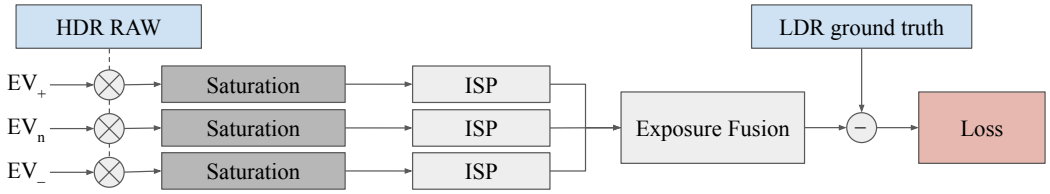

Figure 3: Diagram for loss calculation in training: HDR RAW image in dataset provides the simulated scene irradiance. The 3-shot bracketing can be generated by the image formation model, and fused by Exposure fusion.

by 1% and 99% percentile of the sorted HDR pixels. Exposure values $\boldsymbol{EV}$ for a 3-shot exposure bracketing are predicted, and the ISO gains are set as $\boldsymbol{g} = \boldsymbol{EV}/\boldsymbol{t}_{AE}$. With the HDR images, losses in Section are calculated and used to train the network parameters with learning rate as 10^{-5} and batch size as 8 by Adam [28]. Note that, the HDR images, simulating the real-world irradiance, are only used in training and testing, but not needed in inference stage.

To better simulate the real-world scenario, noise is added in testing. With the noise model, RAW image formulation in Equation 2 is modified as

$$\boldsymbol{I}_{\text{RAW}} = \min\{\lfloor g(\boldsymbol{\phi}\boldsymbol{t} + \boldsymbol{n}_{\text{pre}}) + \boldsymbol{n}_{\text{post}} \rfloor, I_{\max}\}, \tag{4}$$

where $\boldsymbol{n}_{\text{pre}}$ is the noise introduced before amplification, including the noise from photon arrivals and from sensor readout. $\boldsymbol{n}_{\text{post}}$ is the noise after amplification, mainly from analog-to-digital conversion (ADC). The captured exposure bracketing is then fused by the Exposure Fusion algorithm [14]. The performance is evaluated by PSNR between the fused images and the ground truth.

Comparison with State-of-the-art Methods

To validate the effectiveness of the proposed method, comparison is conducted against the state-of-the-art approaches proposed by Pourreza-Shahri et al [12], and Wang et al [10]. Similar to the proposed method, these approaches determine the exposure bracketing based on an AE preview image and without the dependency of RAW images or accurate camera response functions.

For fair comparison, all the selected bracketing from the above methods is composed of three images and fused directly by the same fusion algorithm [14]. In [10], a fusion module is needed to provide reward for the training of the exposure bracketing selection network (EBSNet). The fusion module can be either the proposed multi-exposure fusion network (MEFNet) or the Exposure Fusion algorithm [14] as shown in their experiments. We train two EBSNets individually with these different fusion modules on our dataset, and the results show that [14] can do better than the MEFNet. Although this is not consistent with the results in the paper [10], but comparable performance is achieved by the retrained model.

The quantitative results can be found in Table 1. The whole test dataset includes 50 day scenes and 50 night scenes. The exposure bracketing selected by the proposed method results in the most similar fused images as the ground truth. Some qualitative examples can be found in Figure 4. As shown by the AE images in Column (2), with one LDR image, the contrast decrease due to saturation is inevitable in high dynamic range scenes, as shown by the poor tone in tunnel (a) and the detail loss in billboards (d). [12]

	Method	PSNR
All	Pourreza-Shahri [12]	25.05
	EBSNet [10]	29.36
	Proposed	**32.53**
Day	Pourreza-Shahri [12]	22.53
	EBSNet [10]	29.04
	Proposed	**33.41**
Night	Pourreza-Shahri [12]	27.56
	EBSNet [10]	29.67
	Proposed	**31.65**

Table 1: Comparison of the state-of-the-art exposure strategy methods.

predicts the exposure values based only on the local statistics of the AE image, and thus cannot capture details in saturated regions as shown in Column (3). The large variance of PSNR across the dataset also indicates that it has difficulty to generalize to scenes with different irradiance ranges. The EBSNet [10] achieves less information loss in very bright/dark regions, which benefits from the semantic features extracted by the heavy network. However, the fused images suffer from low contrast as shown in the Column (4). The proposed network gives the best results. For example in Figure 4(d), the proposed method gives a better recovery in the billboard areas and a more visually-pleasing tone of the backlit portrait.

The processing time, on a TESLA V100 GPU, is reported in Table 2. [12] optimizes the exposure values by an iterative algorithm, which consumes longer time. For the EBSNet [10] and the proposed method, the processing time includes the extraction of multi-scale histogram from a low-resolution image and the running time of the CNN networks. Since the EBSNet contains more than 60 million parameters, while the proposed lightweight network has around 300K parameters, the proposed system can be a more efficient and practical solution on low-cost devices and real-time application.

	time (s)
EBSNet [10]	0.484
Pourreza-Shahri [12]	1.078
Proposed	**0.001**

Table 2: Comparison of the state-of-the-art exposure strategy methods in terms of processing time.

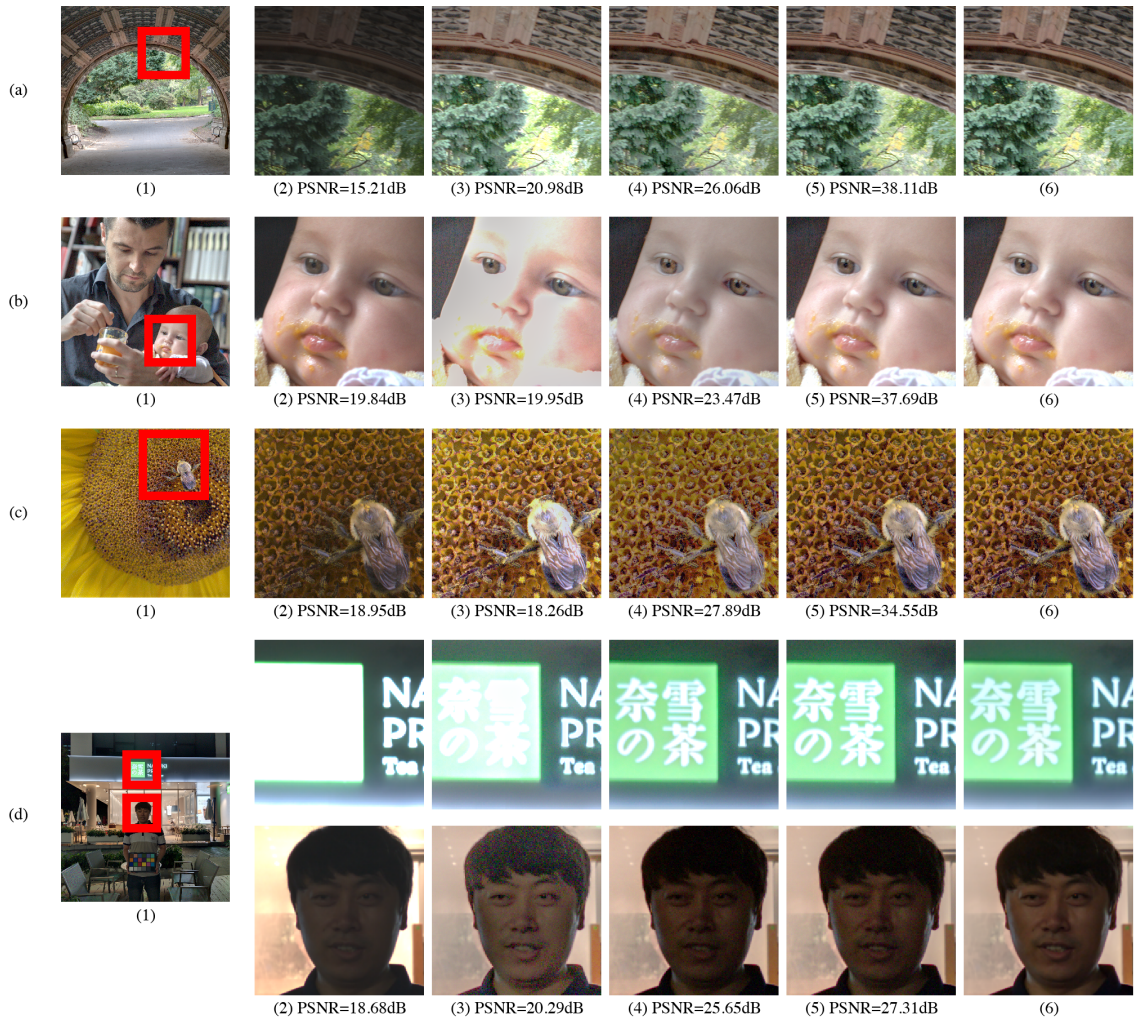

Figure 4: Resulting images of the state-of-the-art exposure bracketing strategy methods, with PSNR against the ground truth in captions. Column (1): ground truth with zoomed-in regions (red boxes); Column (2): AE images; Column (3): results of Pourreza-Shahri [12]; Column (4): results of EBSNet [10]; Column (5): results of the proposed method; Column (6): zoomed-in ground truth.

Ablation Study

Day and Night Scenes

The scene dynamic range varies from day to night, which challenges the generalization of exposure bracketing strategies. As for the system in [10], two models need to be trained separately for day and night scenes with different sets of exposure bracketing candidates. To validate the generalization of the proposed system, three models are trained and evaluated on different dataset, i.e. day scenes, night scenes and the combination. Model NN_{day} and NN_{night} are trained respectively on dataset of 200 day images and 200 night images, while model NN_{all} are trained on the joint dataset. The quantitative result is shown in Table 3. The performance of three networks, tested in their corresponding dataset, is similar, which indicates that NN_{all} can be generalized well to scenes with various irradiance distribution.

	NN_{day}	NN_{night}	NN_{all}
Day	34.13	26.08	33.41
Night	16.29	31.94	31.65

Table 3: Performance of models trained with day and night dataset, separately and jointly, in terms of PSNR against the ground truth images.

Conclusion

In this paper, a lightweight network for exposure bracketing strategy is proposed for HDR imaging with no access to camera RAW. A neural network is developed to learn irradiance information and predict the exposure values of a 3-shot bracketing driven by the optimization of the fused tone. The network takes multi-scale histograms of an LDR preview image and the maximum and minimum scene irradiance as inputs, which can be easily accessed by camera APIs. The result shows that the proposed system outperforms state-of-the-art methods in terms of fused tone as well as time efficiency. Without depending on the RAW images, the proposed system can provide a promising solution for low-cost, power-efficient cameras and real-time application. The proposed method assumes that the multiple LDR images are aligned, and noise is not considered. It can be a future work to include mo-

tion, noise and tone into one framework for exposure bracketing strategy design.

Acknowledgement

This work was supported by SenseBrain Technology. Thanks to the colleagues for the dataset and Dr. Jinwei Gu for advice.

References

[1] S. W. Hasinoff, F. Durand, and W. T. Freeman, "Noise-optimal capture for high dynamic range photography," in *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. IEEE, 2010, pp. 553–560.

[2] K. Seshadrinathan, S. H. Park, and O. Nestares, "Noise and dynamic range optimal computational imaging," in *2012 19th IEEE International Conference on Image Processing*. IEEE, 2012, pp. 2785–2788.

[3] M. A. Robertson, S. Borman, and R. L. Stevenson, "Estimation-theoretic approach to dynamic range enhancement using multiple exposures," *Journal of Electronic Imaging*, vol. 12, no. 2, pp. 219–228, 2003.

[4] M. D. Grossberg and S. K. Nayar, "What is the space of camera response functions?" in *2003 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2003. Proceedings.*, vol. 2. IEEE, 2003, pp. II–602.

[5] P. E. Debevec and J. Malik, "Recovering high dynamic range radiance maps from photographs," in *ACM SIGGRAPH 2008 classes*, 2008, pp. 1–10.

[6] M. Granados, B. Ajdin, M. Wand, C. Theobalt, H.-P. Seidel, and H. P. Lensch, "Optimal hdr reconstruction with linear digital cameras," in *2010 IEEE computer society conference on computer vision and pattern recognition*. IEEE, 2010, pp. 215–222.

[7] T. Chen and A. El Gamal, "Optimal scheduling of capture times in a multiple-capture imaging system," in *Sensors and Camera Systems for Scientific, Industrial, and Digital Photography Applications III*, vol. 4669. International Society for Optics and Photonics, 2002, pp. 288–296.

[8] O. Gallo, M. Tico, R. Manduchi, N. Gelfand, and K. Pulli, "Metering for exposure stacks," in *Computer Graphics Forum*, vol. 31, no. 2pt2. Wiley Online Library, 2012, pp. 479–488.

[9] P. van Beek, "Improved image selection for stack-based hdr imaging," *arXiv preprint arXiv:1806.07420*, 2018.

[10] Z. Wang, J. Zhang, M. Lin, J. Wang, P. Luo, and J. Ren, "Learning a reinforced agent for flexible exposure bracketing selection," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020, pp. 1820–1828.

[11] K.-F. Huang and J.-C. Chiang, "Intelligent exposure determination for high quality hdr image generation," in *2013 IEEE International Conference on Image Processing*. IEEE, 2013, pp. 3201–3205.

[12] R. Pourreza-Shahri and N. Kehtarnavaz, "Exposure bracketing via automatic exposure selection," in *2015 IEEE International Conference on Image Processing (ICIP)*. IEEE, 2015, pp. 320–323.

[13] ——, "Automatic exposure selection for high dynamic range photography," in *2015 IEEE International Conference on Consumer Electronics (ICCE)*. IEEE, 2015, pp. 471–472.

[14] T. Mertens, J. Kautz, and F. Van Reeth, "Exposure fusion," in *15th Pacific Conference on Computer Graphics and Applications (PG'07)*. IEEE, 2007, pp. 382–390.

[15] K. Kirk and H. J. Andersen, "Noise characterization of weighting schemes for combination of multiple exposures." in *BMVC*, vol. 3. Citeseer, 2006, pp. 1129–1138.

[16] F. Drago, K. Myszkowski, T. Annen, and N. Chiba, "Adaptive logarithmic mapping for displaying high contrast scenes," in *Computer graphics forum*, vol. 22, no. 3. Wiley Online Library, 2003, pp. 419–426.

[17] J. Duan, M. Bressan, C. Dance, and G. Qiu, "Tone-mapping high dynamic range images by novel histogram adjustment," *Pattern Recognition*, vol. 43, no. 5, pp. 1847–1862, 2010.

[18] F. Durand and J. Dorsey, "Fast bilateral filtering for the display of high-dynamic-range images," in *Proceedings of the 29th annual conference on Computer graphics and interactive techniques*, 2002, pp. 257–266.

[19] P. Debevec and S. Gibson, "A tone mapping algorithm for high contrast images," in *13th eurographics workshop on rendering: Pisa, Italy. Citeseer*. Citeseer, 2002.

[20] J. H. Kim, S. Lee, and S.-J. Kang, "End-to-end differentiable learning to hdr image synthesis for multi-exposure images," *arXiv preprint arXiv:2006.15833*, 2020.

[21] 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," in *Computer Graphics Forum*, vol. 37, no. 2. Wiley Online Library, 2018, pp. 37–49.

[22] E. Onzon, F. Mannan, and F. Heide, "Neural auto-exposure for high-dynamic range object detection," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2021, pp. 7710–7720.

[23] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," *Advances in neural information processing systems*, vol. 25, 2012.

[24] K. Ram Prabhakar, V. Sai Srikar, and R. Venkatesh Babu, "Deepfuse: A deep unsupervised approach for exposure fusion with extreme exposure image pairs," in *Proceedings of the IEEE international conference on computer vision*, 2017, pp. 4714–4722.

[25] K. R. Prabhakar, R. Arora, A. Swaminathan, K. P. Singh, and R. V. Babu, "A fast, scalable, and reliable deghosting method for extreme exposure fusion," in *2019 IEEE International Conference on Computational Photography (ICCP)*. IEEE, 2019, pp. 1–8.

[26] S.-Y. Chen and Y.-Y. Chuang, "Deep exposure fusion with deghosting via homography estimation and attention learning," in *ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2020, pp. 1464–1468.

[27] V. Bychkovsky, S. Paris, E. Chan, and F. Durand, "Learning photographic global tonal adjustment with a database of input / output image pairs," in *The Twenty-Fourth IEEE Conference on Computer Vision and Pattern Recognition*, 2011.

[28] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2014.