

Efficient Progressive High Dynamic Range Image Restoration via Attention and Alignment Network

Gaocheng Yu[†] Jin Zhang[†] Zhe Ma
 Hongbin Wang
 AntGroup

{yugaocheng.ygc, zj346862, mz281827, hongbin.whb}@antgroup.com

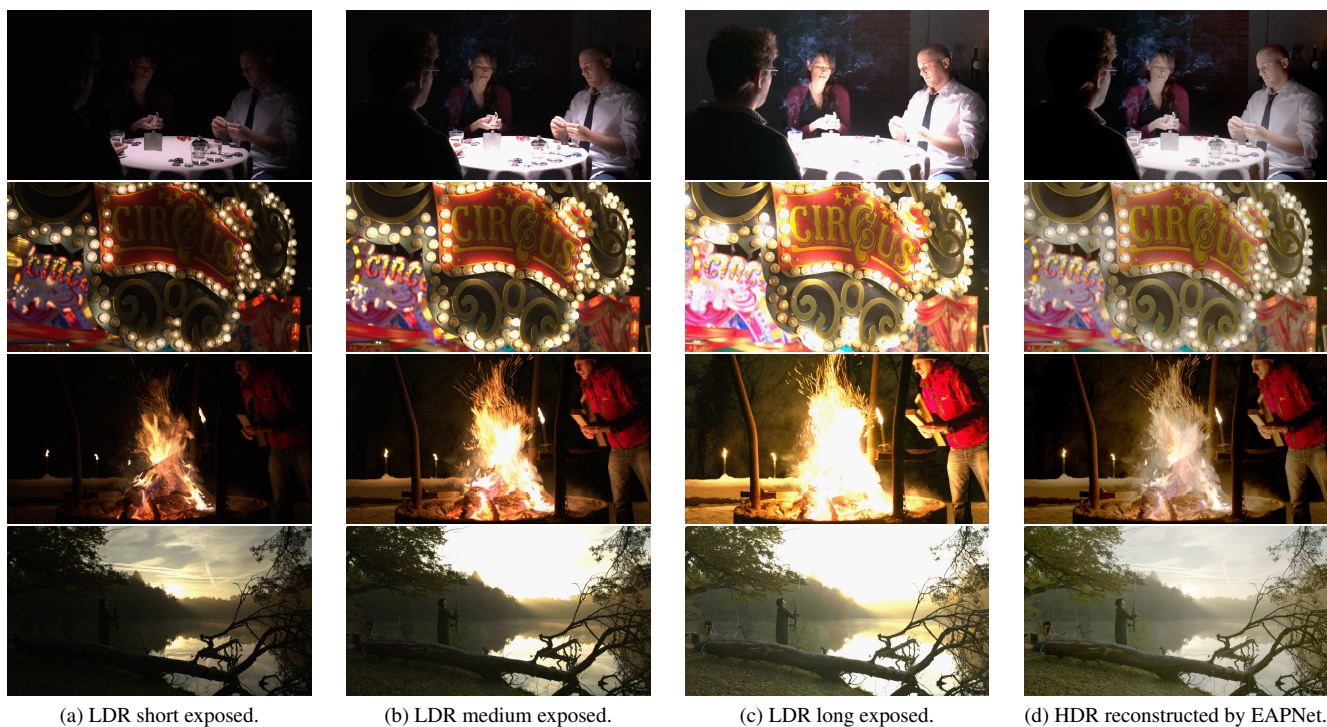


Figure 1. Examples of HDRs restored from LDRs by EAPNet. Fig. 1a, Fig. 1b and Fig. 1c are short, medium, and long exposed LDRs respectively, and the HDR images restored by our standard EAPNet are presented in Fig. 1d.

Abstract

HDR is an important part of computational photography technology. In this paper, we propose a lightweight neural network called **Efficient Attention-and-alignment-guided Progressive Network (EAPNet)** for the challenge NTIRE 2022 HDR Track 1 and Track 2. We introduce a multi-scale lightweight encoding module to extract features. Besides, we propose **Progressive Dilated U-shape Block (PDUB)** which is a progressive plug-and-play module for dynam-

ically tuning MAccs and PSNR. Finally, we use fast and low-power feature-alignment module to deal with misalignment problem in place of the time-consuming Deformable Convolutional Network (DCN). The experiments show that our method achieves about $20\times$ compression on MAccs with better PSNR- μ and PSNR compared to the state-of-the-art method. We got the 2nd place of both two tracks during the testing phase. Fig. 1 shows the visualized result of NTIRE 2022 HDR challenge.

[†]These authors contributed equally to this work.

1. Introduction

HDR image restoration task aims to recover image contents, details, and color from one or several LDRs. Due to the limitations of camera sensor, the LDRs degraded model can be described as in [25]:

$$I_f = \min\{\Phi t/g + I_0 + n, I_{max}\} \quad (1)$$

where I_f denotes LDR image observations, Φ is the scene brightness, t is the exposure time, g is the sensor gain, I_0 is the constant offset current, n is the sensor noise, and I_{max} denotes the saturation point [25].

In the recent, learnable approaches attempt to recover HDR images via convolutional neural networks (CNNs) [3,5]. There are two ways of HDR image restoration: single frame restoration and multi frame restoration. The former focuses on learning a mapping from LDR to HDR [3, 8], and the latter processes Multi-Exposure Fusion (MEF), and works on the misalignment problem [2] additionally.

Generally, the ISP pipeline [11, 18, 28] shows that the LDRs suffer from problems such as motions, noises, truncations, and saturations in the image signal process. Taking the above issues into account, researchers work on optical flow [27] and denoise [34] technologies jointly with HDR. Meanwhile, the application of HDR restoration technology requires real time performance on mobile devices [6]. Due to the highly resolution of input LDRs, low-power and efficient technology will become more challenging than image detection [19] and image classification [4] tasks.

In this paper, we propose a lightweight network Efficient Attention-and-alignment-guided Progressive Network (EAPNet) for HDR restoration task. We consider the balance between performance and efficiency, and adopt depth-wise separable convolution [12, 13, 26]. To deal with misalignment problem, we introduce the feature-alignment network [32] to predict scale and bias to describe the local offset instead of DCN. And we propose a new module called Progressive Dilated U-shape Block (PDUB) with low computation cost and good performance. We participated in NTIRE 2022 High Dynamic Range Challenge (both two tracks), and won the 2nd place in the two tracks.

Our main contributions are summarized as follow:

- Compared with the SOTA method [30], we achieve a maximum 20x MAccs's compression (NTIRE 2022 High Dynamic Range Challenge, Track 2) under the same PSNR and PSNR- μ .
- We propose Progressive Dilated U-shape Block (PDUB), and the experiments show that we achieve better performance compared with DRDB [30].
- We introduce a lightweight Feature-Alignment module to deal with misalignment problem with low computation cost.

- Taking noise and MAccs into consideration, we propose an efficient Multi-Scale Encoder layer.

Figure 2 shows the overall structure of our method.

2. Related Work

2.1. Single Frame HDR

Single Frame HDR has been extensively studied. Typically, without the negative impact of ghost artefact, single frame HDR researches on the LDR-to-HDR mapping problem. [5, 31] propose a CNN based method to predict the HDR images directly. [3, 36] use a network to predict bilateral grid of coefficients in low-resolution, and lead to image enhancement. [23] devises a multi-branch network which extracts the local and global features respectively, and fuses them via concatenation operator, leading to a good performance.

Researches on efficient and low-power single frame HDR solutions are wide. [1, 3] propose a dequantize network to recover missing details from the low-bit LDR image. And [8, 16] formulate the HDR task as a curve estimation problem via neural network, and adjust the LDRs pixel-wise. All of the above methods learn a compressing space to describe the mapping between LDR and HDR, discarding the details of under or over exposed area.

2.2. Multi Frame HDR

Specifically, with the bracketed exposure strategy, [11, 24] use image pyramid, and calculate brightness weight, contrast weight, and exposure weight to fuse LDRs. Furthermore, [17, 21] propose a Structural Patch Decomposition (SPD) approach, decompose the signal into intensity and structure components, restore them individually, and fuse the restored components. The traditional approaches generally recover HDR image by means of decomposition and reconstruction.

Many learnable methods have been developed [2, 3, 20, 29]. [29] proposes a merge and fine-tune architecture to solve large-scale motions. [2, 3, 20] adopt an attention-guided mechanism to accomplish ghost-free merging. [20] additionally uses Deformable Convolution Network (DCN) to predict optical flow and align the other frames to the reference frame, achieving a good performance on detail restoration.

2.3. Efficient Convolutional Neural Network

The MobileNets [12, 13, 26] decompose the standard convolution operator into two parts: depthwise convolution and pointwise convolution. [10] uses group convolution to reuse redundant features. [22, 35] propose a channel-shuffle operator, receiving a high precision on image classification and detection tasks. [9] introduces the attention mechanism

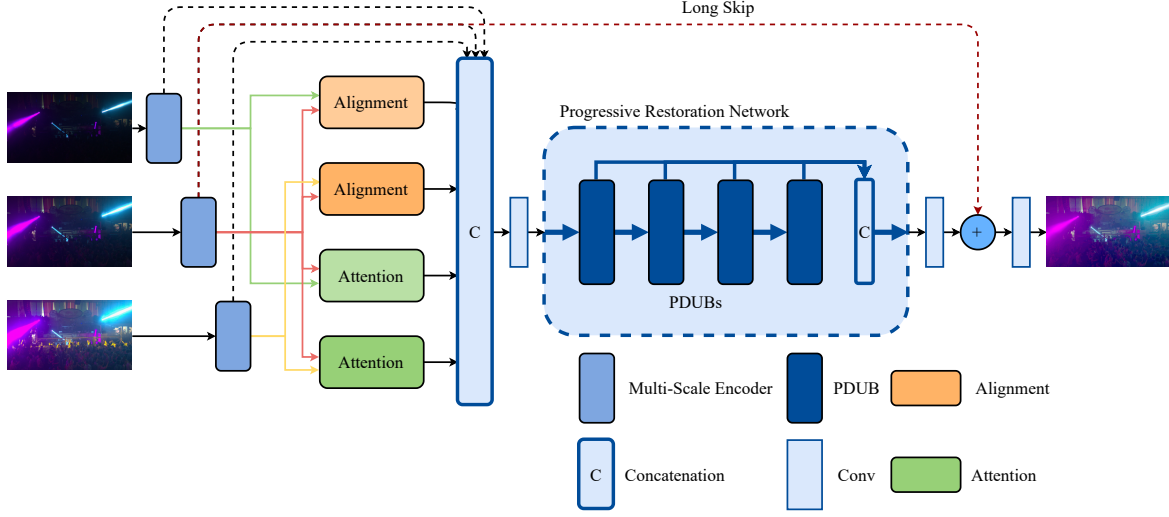


Figure 2. Overall structure of EAPNet.

on the basis of [12, 13, 26], and proposes a Large Kernel Attention (LKA) module to enable self-attention and self-adaptation according to the high-wide receptive field.

3. Proposed Method

We propose an end-to-end network EAPNet consisting of a Feature Extraction Network and a Progressive Restoration Network. The overall structure of EAPNet is presented in Fig. 2.

The Feature Extraction Network first extracts features with a Multi-Scale Encoder Module, and then attends features with an Attention Module and a Feature-Alignment Module. The Progressive Restoration Network first fuses features by guidance from attention and alignment module, and then progressively restores features with proposed PDUBs. The restored features are then upsampled to the original resolution to fuse with the long skip feature, and then recovered with several standard convolutional layers.

We consider three LDR images, i.e., $I_i, i = 1, 2, 3$ as input and let the second LDR image I_2 be the reference frame. With the Exposure Value (EV), we conduct EV alignment, and directly concatenate them with the original input, and get 6-channel input:

$$I_i^{input} = \text{concat}([I_i, f(I_i)]), \quad i = 1, 2, 3 \quad (2)$$

where I_i^{input} is the input of each frame, I_i is the original LDR image, and f is the EV alignment function.

Depthwise Separable Convolution. Depthwise separable convolution, proposed by [13], factorizes a standard convolution into a depthwise convolution and a 1×1 convolution called pointwise convolution. The depthwise separable convolution can drastically reduce computation cost and

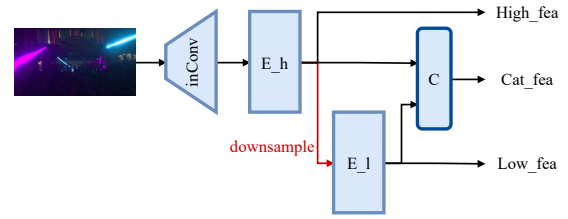


Figure 3. Structure of Multi-Scale Encoder.

model size. We replace the standard convolution in AHDR with depthwise separable convolution, except for the first one in the encoder and the last one in the restoration network.

Multi-Scale Encoder Module. Given the LDRs, we propose the Multi-scale Encoder module (MSEnc) to encode different scales of information with different frequencies. As shown in Fig. 3, the MSEnc first encodes the input with a standard convolution layer, and then extracts high resolution features which usually contain high frequency features with a high extractor. The high features are then downsampled and processed by a low extractor to produce low resolution features. The high and low features are further processed by attention and alignment module separately, and concatenated features are passed to the restoration network. The reference frame features are also passed to the restoration network as the long skip feature.

$$f_{h,i} = E_{h,i}(I_i^{input}), \quad i = 1, 2, 3 \quad (3)$$

$$f_{l,i} = E_{l,i}(f_{h,i}), \quad i = 1, 2, 3 \quad (4)$$

Exp.	MS	PDUB	share	align	conv-up	PSNR(db)	PSNR- μ (db)	MAccs(G)
1			Y			38.23	36.48	152.11
2	✓		Y			38.54	36.70	137.66
3	✓	✓	Y			38.67	36.87	121.03
4	✓	✓	Y	full		38.24	36.79	146.28
5	✓	✓	N			38.60	36.79	121.03
6*	✓	✓	N	full		38.34	36.92	146.28
7	✓	✓	N	half		38.54	36.89	126.43
8*	✓	✓	N	half	✓	38.74	37.02	198.38

Table 1. Ablation studies of effectiveness of network modules based on μ -law L_1 loss. Exp. 6 is our lightweight model, and this is the model version in the listed ranking results for Track 2 in Tab. 5. In exp. 7, we half both the feature map resolution of attention and alignment module. Exp.8 is our standard model, and this is the model version in the listed ranking results for Track 1 in Tab. 4.



Figure 7. Qualitative comparison of our EAPNet with SOTA method AHDR. Fig. 7a, Fig. 7b and Fig. 7c are short, medium, and long exposed LDRs respectively. The HDR images restored by AHDR are shown in Fig. 7d, and the HDR images restored by our standard EAPNet are shown in Fig. 7e, and the HDR ground-truth images are shown in Fig. 7f.

#PDUBs	PSNR(db)	PSNR- μ (db)	MAccs(G)
2	38.45	36.77	128.25
3	38.34	36.92	146.28
4	38.74	36.95	164.31
5	38.77	36.97	182.34

Table 2. Comparison on number of PDUBs. Experiments are conducted based on lightweight model with μ -law L_1 loss.

4.2. Implementation Details

We set the number of PDUBs as 4, downsample the alignment feature map resolution, and apply transposed convolution in PDUB for our standard EAPNet model. While for the lightweight model, we cut one PDUB, keep full resolution for alignment module, and use bilinear up-sampling to save computation cost and speed up.

Loss	PSNR(db)	PSNR- μ (db)
L_1	38.30	36.40
Tan_L_1	38.75	36.91
μ -law L_1	38.34	36.92
$Tan_L_1 + \mu$ -law L_1	38.66	36.92
μ -law $L_1 + half^*$	38.54	36.89
$Tan_L_1 + \mu$ -law $L_1 + half^*$	39.00	36.96

Table 3. Comparison on losses. Experiments are conducted based on lightweight model. With experiments noted by *, we downsample the feature map resolution of attention module and alignment module.

During the training stage, we crop the input LDRs and ground-truth to 256x256-sized patches with an overlap of 128. The network is trained from scratch, and optimized by an Adam optimizer [14]. We train the model for 10 million iterations with a batch size of 32. The initial learning rate is

Team	PSNR(db)	PSNR- μ (db)	Runtime(s)	MAccs(G)	Params.(k)
ALONG	39.417	37.424	0.324	198.47	489.01
Ours	38.607	37.252	0.185	198.38	576.23
XPixel-UM	38.015	37.209	0.276	199.88	1013.25
AdeTeam	39.001	37.163	0.134	156.12	188.99
CZCV	37.388	36.972	0.431	193.93	633.69

Table 4. Results of NTIRE 2022 HDR Challenge Track 1 top 5 methods on online testset [25]. Standard EAPNet is used in Track 1.

set as $8e-4$, and decayed every 100k iterations by a factor of 0.75. We use gradual warmup strategy [7] for the first 1000 steps to overcome early optimization difficulties. All models are built on the PyTorch framework, and trained with 4 NVIDIA V100 GPUs, costing about 7 days.

Temporal ensemble [15] is applied to generate an exponential moving average (EMA) model of the trained model. We update the EMA model every 5k steps with a decay rate of 0.5. The final EMA model is used to generate predictions for testing.

During the validation and testing stage, we pad the input images from 1900×1060 to 1920×1080 , and crop to the original size after model inference. Inferencing is conducted on NVIDIA V100 GPU. It takes 0.24 second per image with 198.38G MAccs for our standard model, and 0.15 second per image with 146.28G MAccs for our lightweight model specifically.

4.3. Results

We participated in the NTIRE 2022 HDR Challenge and won the second place in both track 1 and track 2. The comparisons of our method with other top methods and the challenge baseline are listed in Tab. 4 and Tab. 5.

We also compare our method with some state-of-the-art methods, including AHDR [30] and ADNet [20]. The quantitative results are provided in Tab. 6. In terms of fidelity, our standard EAPNet exceeds AHDR provided by NTIRE 2022 HDR organizers by 0.16db in PSNR- μ and 0.4db in PSNR, with about $15 \times$ compression in MAccs. And in terms of complexity, our lightweight EAPNet achieves about $20 \times$ compression in MAccs, with PSNR maintained the same and a slight improvement in PSNR- μ .

We present some reconstructed HDR examples in Fig. 7 produced from LDRs by EAPNet. Details show that long-exposed and short-exposed areas are recovered from LDRs, and fine details are restored in the reconstructed HDRs.

4.4. Ablation Studies

In this part, we present ablation studies to analysis the effectiveness of each part of our model. Experiment results are listed in Tab. 1.

We replace the standard convolution in AHDR [30] with depthwise separable convolution [13], and then reduce the

Team	PSNR(db)	PSNR- μ (db)	Runtime(s)	MAccs(G)	Params.(k)
ALONG	38.843	37.033	0.183	74.02	177.41
Ours	38.766	37.100	0.155	146.28	393.73
AdeTeam	39.001	37.163	0.136	156.12	188.99
BOE-IOT-AIBD	38.150	37.225	0.473	756.78	1355.10
MegHDR	38.749	37.030	0.269	790.48	401.83
Baseline(AHDR)	37.597	37.021	0.760	2916.92	1141.28

Table 5. Results of NTIRE 2022 HDR Challenge Track 2 top 5 methods on online testset [25]. The result is produced by lightweight EAPNet.

	PSNR(db)	PSNR- μ (db)	MAccs(G)
AHDR [†]	38.34	36.86	2916.92
AHDR*	38.98	36.82	2710.12
ADNet	39.34	37.21	6249.43
Lightweight EAPNet	38.34	36.92	146.28
Standard EAPNet	38.74	37.02	198.38

Table 6. Comparison with state-of-the-art methods. AHDR* is reproduced by authors, and AHDR[†] is provided by NTIRE 2022 HDR organizers [25].

feature map resolution by half in the restoration network to reduce model complexity. We take this model as our base model, with 152.11G MAccs, corresponding to exp. 1 in Tab. 1.

Multi-Scale Encoder. We replace the naive encoder block in the base model with our proposed Multi-Scale Encoder in exp.2 in Tab. 1. Specifically, we use a 2-scale encoder, and obtain 0.22db gain on PSNR- μ performance with about 15G compression in MAccs.

DRDB V.S. PDUB. We conduct experiments to explore the performance of DRDB and PDUB in exp. 2 and 3 in Tab. 1. Results show that PDUB gets 0.17db higher on PSNR- μ and 0.07db higher on PSNR, with 16.6G MAccs lower than DRDB.

Alignment module and weight sharing. Encoder and attention module weights are shared by default in AHDR. In our method, these two modules are set apart with no weight sharing, since short-exposed and long-exposed frames may have different types of features. An additional alignment module helps to improve model performance by attending feature merging. Exp. 4, 5 and 6 in Tab. 1 show that alignment module and no weight sharing strategy together help to improve model performance, and this is our lightweight model.

Standard model. We further shrink the model by half the feature map resolution of alignment module, and then we replace bilinear upsampling with transposed convolutional upsampling in exp.8, and obtain the final standard model.

Number of PDUBs. As listed in Tab. 2, experiments on the number of PDUBs show that the model achieves better performance as the number of PDUBs increases. Consider-

ing model efficiency, we set the number of PDUBs to 4 for our standard model and 3 for our lightweight model.

Loss. As listed in Tab. 3, μ -law L_1 yields slightly better PSNR- μ , and with Tan_L_1 we get higher PSNR. The proposed loss achieves a balance between PSNR and PSNR- μ .

5. Conclusion

In this paper, we describe the solution EAPNet for the NTIRE 2022 HDR Track 1 and Track 2. We use depth-wise separable convolution and downsample feature map resolution to improve efficiency. The multi-scale encoder module concatenates shallow and deep features which are strongly required by HDR restoration. We propose PDUB that can progressively recover HDR images with low computation cost. To deal with misalignment problems, we use a lightweight and highly efficient feature-alignment module rather than DCN. To summarize, our method achieves good performance and outperforms the state-of-the-art method proven by experiments.

References

- [1] SM A Sharif, Rizwan Ali Naqvi, Mithun Biswas, and Sungjun Kim. A two-stage deep network for high dynamic range image reconstruction. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 550–559, 2021. 2
- [2] Sibi Catley-Chandar, Thomas Tanay, Lucas Vandroux, Aleš Leonardis, Gregory Slabaugh, and Eduardo Pérez-Pellitero. Flexhdr: Modelling alignment and exposure uncertainties for flexible hdr imaging, 2022. 2
- [3] Xiangyu Chen, Yihao Liu, Zhengwen Zhang, Yu Qiao, and Chao Dong. Hdrunet: Single image hdr reconstruction with denoising and dequantization. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 354–363, 2021. 2, 4
- [4] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. Imagenet: A large-scale hierarchical image database. In *2009 IEEE conference on computer vision and pattern recognition*, pages 248–255. Ieee, 2009. 2
- [5] Gabriel Eilertsen, Joel Kronander, Gyorgy Denes, Rafal K Mantiuk, and Jonas Unger. Hdr image reconstruction from a single exposure using deep cnns. *ACM transactions on graphics (TOG)*, 36(6):1–15, 2017. 2
- [6] Michaël Gharbi, Jiawen Chen, Jonathan T Barron, Samuel W Hasinoff, and Frédo Durand. Deep bilateral learning for real-time image enhancement. *ACM Transactions on Graphics (TOG)*, 36(4):118, 2017. 2
- [7] Priya Goyal, Piotr Dollár, Ross Girshick, Pieter Noordhuis, Lukasz Wesolowski, Aapo Kyrola, Andrew Tulloch, Yangqing Jia, and Kaiming He. Accurate, large mini-batch sgd: Training imagenet in 1 hour. *arXiv preprint arXiv:1706.02677*, 2017. 6
- [8] Chunle Guo Guo, Chongyi Li, Jichang Guo, Chen Change Loy, Junhui Hou, Sam Kwong, and Runmin Cong. Zero-reference deep curve estimation for low-light image enhancement. In *Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR)*, pages 1780–1789, June 2020. 2
- [9] Meng-Hao Guo, Cheng-Ze Lu, Zheng-Ning Liu, Ming-Ming Cheng, and Shi-Min Hu. Visual attention network. *arXiv preprint arXiv:2202.09741*, 2022. 2
- [10] Kai Han, Yunhe Wang, Qi Tian, Jianyuan Guo, Chunjing Xu, and Chang Xu. Ghostnet: More features from cheap operations. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 1580–1589, 2020. 2
- [11] Samuel W Hasinoff, Dillon Sharlet, Ryan Geiss, Andrew Adams, Jonathan T Barron, Florian Kainz, Jiawen Chen, and Marc Levoy. Burst photography for high dynamic range and low-light imaging on mobile cameras. *ACM Transactions on Graphics (ToG)*, 35(6):1–12, 2016. 2
- [12] Andrew Howard, Mark Sandler, Grace Chu, Liang-Chieh Chen, Bo Chen, Mingxing Tan, Weijun Wang, Yukun Zhu, Ruoming Pang, Vijay Vasudevan, et al. Searching for mobilenetv3. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 1314–1324, 2019. 2, 3
- [13] Andrew G Howard, Menglong Zhu, Bo Chen, Dmitry Kalenichenko, Weijun Wang, Tobias Weyand, Marco Andreetto, and Hartwig Adam. Mobilenets: Efficient convolutional neural networks for mobile vision applications. *arXiv preprint arXiv:1704.04861*, 2017. 2, 3, 6
- [14] Diederik P Kingma and Jimmy Ba Adam. A method for stochastic optimization. arxiv preprint: 1412.6980. *Google Scholar*, 2014. 5
- [15] Samuli Laine and Timo Aila. Temporal ensembling for semi-supervised learning. *arXiv preprint arXiv:1610.02242*, 2016. 6
- [16] Chongyi Li, Chunle Guo, and Chen Change Loy. Learning to enhance low-light image via zero-reference deep curve estimation. *arXiv preprint arXiv:2103.00860*, 2021. 2
- [17] Hui Li, Kede Ma, Hongwei Yong, and Lei Zhang. Fast multi-scale structural patch decomposition for multi-exposure image fusion. *IEEE Transactions on Image Processing*, 29:5805–5816, 2020. 2
- [18] Orly Liba, Kiran Murthy, Yun-Ta Tsai, Tim Brooks, Tianfan Xue, Nikhil Karnad, Qirui He, Jonathan T Barron, Dillon Sharlet, Ryan Geiss, et al. Handheld mobile photography in very low light. *ACM Transactions on Graphics (TOG)*, 38(6):1–16, 2019. 2
- [19] Tsung-Yi Lin, Michael Maire, Serge Belongie, James Hays, Pietro Perona, Deva Ramanan, Piotr Dollár, and C Lawrence Zitnick. Microsoft coco: Common objects in context. In *European conference on computer vision*, pages 740–755. Springer, 2014. 2
- [20] Zhen Liu, Wenjie Lin, Xinpeng Li, Qing Rao, Ting Jiang, Mingyan Han, Haoqiang Fan, Jian Sun, and Shuaicheng Liu. Adnet: Attention-guided deformable convolutional network for high dynamic range imaging. In *CVPRW*, pages 463–470, 2021. 2, 4, 6
- [21] Kede Ma, Hui Li, Hongwei Yong, Zhou Wang, Deyu Meng, and Lei Zhang. Robust multi-exposure image fusion: a structural patch decomposition approach. *IEEE Transactions on Image Processing*, 26(5):2519–2532, 2017. 2

- [22] Ningning Ma, Xiangyu Zhang, Hai-Tao Zheng, and Jian Sun. Shufflenet v2: Practical guidelines for efficient cnn architecture design. In *Proceedings of the European conference on computer vision (ECCV)*, pages 116–131, 2018. 2
- [23] Demetris Marnerides, Thomas Bashford-Rogers, Jonathan Hatchett, and Kurt Debattista. Expandnet: A deep convolutional neural network for high dynamic range expansion from low dynamic range content. In *Computer Graphics Forum*, volume 37, pages 37–49. Wiley Online Library, 2018. 2
- [24] Tom Mertens, Jan Kautz, and Frank Van Reeth. Exposure fusion. In *15th Pacific Conference on Computer Graphics and Applications (PG'07)*, pages 382–390. IEEE, 2007. 2
- [25] Eduardo Pérez-Pellitero, Sibi Catley-Chandar, Richard Shaw, Ales Leonardis, Radu Timofte, et al. NTIRE 2022 challenge on high dynamic range imaging: Methods and results. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, 2022. 2, 4, 6
- [26] Mark Sandler, Andrew Howard, Menglong Zhu, Andrey Zhmoginov, and Liang-Chieh Chen. Mobilenetv2: Inverted residuals and linear bottlenecks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 4510–4520, 2018. 2, 3
- [27] Zachary Teed and Jia Deng. Raft: Recurrent all-pairs field transforms for optical flow. In *European conference on computer vision*, pages 402–419. Springer, 2020. 2
- [28] Bartłomiej Wronski, Ignacio Garcia-Dorado, Manfred Ernst, Damien Kelly, Michael Krainin, Chia-Kai Liang, Marc Levoy, and Peyman Milanfar. Handheld multi-frame super-resolution. *ACM Transactions on Graphics (TOG)*, 38(4):1–18, 2019. 2
- [29] Shangzhe Wu, Jiarui Xu, Yu-Wing Tai, and Chi-Keung Tang. Deep high dynamic range imaging with large foreground motions. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pages 117–132, 2018. 2, 4
- [30] Qingsen Yan, Dong Gong, Qinfeng Shi, Anton van den Hengel, Chunhua Shen, Ian Reid, and Yanning Zhang. Attention-guided network for ghost-free high dynamic range imaging. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1751–1760, 2019. 2, 4, 6
- [31] Xin Yang, Ke Xu, Yibing Song, Qiang Zhang, Xiaopeng Wei, and Rynson WH Lau. Image correction via deep reciprocating hdr transformation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 1798–1807, 2018. 2
- [32] Lucas D Young, Fitsum A Reda, Rakesh Ranjan, Jon Morton, Jun Hu, Yazhu Ling, Xiaoyu Xiang, David Liu, and Vikas Chandra. Feature-align network with knowledge distillation for efficient denoising. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pages 709–718, 2022. 2
- [33] Lucas D Young, Fitsum A Reda, Rakesh Ranjan, Jon Morton, Jun Hu, Yazhu Ling, Xiaoyu Xiang, David Liu, and Vikas Chandra. Feature-align network with knowledge distillation for efficient denoising. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pages 709–718, 2022. 4
- [34] Kai Zhang, Wangmeng Zuo, Yunjin Chen, Deyu Meng, and Lei Zhang. Beyond a gaussian denoiser: Residual learning of deep cnn for image denoising. *IEEE transactions on image processing*, 26(7):3142–3155, 2017. 2
- [35] Xiangyu Zhang, Xinyu Zhou, Mengxiao Lin, and Jian Sun. Shufflenet: An extremely efficient convolutional neural network for mobile devices. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 6848–6856, 2018. 2
- [36] Zhuoran Zheng, Wenqi Ren, Xiaochun Cao, Tao Wang, and Xiuyi Jia. Ultra-high-definition image hdr reconstruction via collaborative bilateral learning. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 4449–4458, 2021. 2