NTIRE 2022 Challenge on High Dynamic Range Imaging: Methods and Results

Eduardo Pérez-Pellitero[†]   Sibi Catley-Chandar[†]   Richard Shaw[†]   Aleš Leonardis[†]

Radu Timofte[†]   Zexin Zhang   Cen Liu   Yunbo Peng   Yue Lin   Gaocheng Yu
Jin Zhang   Zhe Ma   Hongbin Wang   Xiangyu Chen   Xintao Wang   Haiwei Wu
Lin Liu   Chao Dong   Jiantao Zhou   Qingsen Yan   Song Zhang   Weiye Chen
Yuhang Liu   Zhen Zhang   Yanning Zhang   Javen Qinfeng Shi   Dong Gong
Dan Zhu   Mengdi Sun   Guannan Chen   Yang Hu   Haowei Li   Baozhu Zou
Zhen Liu   Wenjie Lin   Ting Jiang   Chengzhi Jiang   Xinpeng Li   Mingyan Han
Haoqiang Fan   Jian Sun   Shuaicheng Liu   Juan Marín-Vega   Michael Sloth
Peter Schneider-Kamp   Richard Röttger   Chunyang Li   Long Bao   Gang He
Ziyao Xu   Li Xu   Gen Zhan   Ming Sun   Xing Wen   Junlin Li   Jinjing Li
Chenghua Li   Ruipeng Gang   Fangya Li   Chenming Liu   Shuang Feng   Fei Lei
Rui Liu   Junxiang Ruan   Tianhong Dai   Wei Li   Zhan Lu   Hengyan Liu
Peian Huang   Guangyu Ren   Yonglin Luo   Chang Liu   Qiang Tu   Fangya Li
Ruipeng Gang   Chenghua Li   Jinjing Li   Sai Ma   Chenming Liu   Yizhen Cao
Steven Tel   Barthelemy Heyrman   Dominique Ginhac   Chul Lee   Gahyeon Kim
Seonghyun Park   An Gia Vien   Truong Thanh Nhat Mai   Howoon Yoon   Tu Vo
Alexander Holston   Sheir Zaheer   Chan Y. Park

Abstract

This paper reviews the challenge on constrained high dynamic range (HDR) imaging that was part of the New Trends in Image Restoration and Enhancement (NTIRE) workshop, held in conjunction with CVPR 2022. This manuscript focuses on the competition set-up, datasets, the proposed methods and their results. The challenge aims at estimating an HDR image from multiple respective low dynamic range (LDR) observations, which might suffer from under- or over-exposed regions and different sources of noise. The challenge is composed of two tracks with an emphasis on fidelity and complexity constraints: In Track 1, participants are asked to optimize objective fidelity scores while imposing a low-complexity constraint (i.e. solutions can not exceed a given number of operations). In Track 2, participants are asked to minimize the complexity of their solutions while imposing a constraint on fidelity scores (i.e.

solutions are required to obtain a higher fidelity score than the prescribed baseline). Both tracks use the same data and metrics: Fidelity is measured by means of PSNR with respect to a ground-truth HDR image (computed both directly and with a canonical tonemapping operation), while complexity metrics include the number of Multiply-Accumulate (MAC) operations and runtime (in seconds).

1. Introduction

Advances in computational photography have enabled single-sensor cameras to acquire images with an extended dynamic range without the need for expensive, bulky and arguably more inconvenient multi-camera rigs, *e.g.* [11, 32, 43]. The principle behind such single-sensor camera designs relies upon multiple conventional frames captured with different exposure values (EV) that are then fused into a single HDR image [8, 33], with some of those methods including frame alignment [18, 42, 49] or pixel rejection strategies [50].

The current state-of-the-art in multi-frame HDR fusion is significantly dominated by Convolutional Neural Networks

---

† Eduardo Pérez-Pellitero (e.perez.pellitero@huawei.com), Sibi Catley-Chandar, Richard Shaw, Aleš Leonardis (Huawei Noah's Ark Laboratory) and Radu Timofte (ETH Zürich, University of Würzburg) are the NTIRE 2022 Challenge organizers, while the other authors participated in the challenge. Appendix A contains the authors' teams and affiliations. https://data.vision.ee.ethz.ch/cvl/ntire22/

(CNNs), especially for complex dynamic scenes [18, 23, 35, 37, 49, 50]. However, the computational requirements of most of the aforementioned state-of-the-art methods are far from real-time or even interactive processing on most hardware, especially when performing fine-grained motion estimation between frames. Despite several efforts in other computational photography applications [9, 30], HDR imaging has received limited attention in terms of computational efficiency.

The NTIRE 2022 HDR Challenge aims to stimulate research for efficient, low-complexity computational HDR imaging and better understand the state-of-the-art landscape of multiple frame HDR processing. It is part of a broad spectrum of associated challenges within the NTIRE 2022 workshop: spectral recovery [2], spectral demosaicing [1], perceptual image quality assessment [12], inpainting [38], night photography rendering [10], efficient super-resolution [20], learning the super-resolution space [29], super-resolution and quality enhancement of compressed video [52], high dynamic range [36], stereo super-resolution [45], and burst super-resolution [3].

2. Challenge

The NTIRE 2022 HDR Challenge addresses the efficient HDR image enhancement task. This challenge aims to gauge and advance the state-of-the-art on HDR imaging, with a particular emphasis on efficient and low-complexity solutions. It focuses on challenging scenarios for multiframe HDR image reconstruction, *i.e.* wide range of scene illumination, accompanied by complex motions in terms of camera, scene and light sources. The remainder of this section describes the challenge set-up, including the dataset, as well as how the challenge tracks are designed and evaluated.

2.1. Dataset

For the NTIRE 2022 HDR Challenges, we follow the previous edition [35] data set-up, consisting of a curated dataset composed of approximately 1500 training, 60 validation and 201 testing examples. Each example in the dataset is in turn composed of three input LDR images, *i.e.* short, medium and long exposures, and a related ground-truth HDR image aligned with the central medium frame. The images are collected from the work of Froelich *et al.* [11], where they capture an extensive set of diverse and challenging HDR videos using a professional two-camera rig with a semitransparent mirror. Input images are obtained using a pixel-measurement model [14], which includes several sources of noise. For more details about the imaging and noise models, as well as other details, we refer the reader to [35].

2.2. Challenge Design and Tracks

This challenge focuses on developing efficient multiframe HDR reconstruction methods. The challenge is organized in two different tracks that focus on different aspects of the fidelity-complexity trade-off.

2.2.1 Track 1: Fidelity (low-complexity constraint)

This track enforces a strict low-complexity constraint, requiring all submitted methods to be less than 200 GMACs for a testing input tensor of size $(3, 3, 1900, 1060)$, *i.e.* (sequence length, channels, width, height). All methods are required to perform any pre- or post-processing within the model definition, *e.g.* gamma correction, normalization, exposure alignment, downsampling. The submitted methods are then ranked based on fidelity metrics, with an focus on the tonemapped PSNR-μ. Participants are thus invited to optimize the fidelity scores while ensuring they meet the constraint on computational complexity.

2.2.2 Track 2: Low Complexity (fidelity constraint)

This track enforces a fidelity constraint, requiring all submitted methods to achieve greater PSNR and PSNR-μ values than the baseline model, AHDR [50]. The submitted methods are then ranked based on complexity metrics, with an emphasis on the number of GMACs. As in Track 1, GMACs are calculated with an input tensor of size $(3, 3, 1900, 1060)$. Participants are thus invited to minimize the computational complexity, while ensuring they meet the constraint on fidelity.

2.3. Challenge Phases

The challenge consists of two distinct phases, a development phase intended to allow participants to improve and validate their models and a testing phase designed to evaluate final submissions.

2.3.1 Development Phase

Participants are given access to the complete training set, including the input LDR sequences and corresponding HDR ground truths. In addition, participants have access to only the input LDR sequences of the validation set. Each track respective constraints are announced (complexity limit for Track 1 and baseline performance on the validation set for Track 2). The participants are able to compute fidelity metrics on the validation set by uploading their predictions to the Codalab challenge server, and are given access to scripts to compute complexity metrics which are in turn self-reported in the Codalab as well. The validation leaderboard is visible to all participants.

Team	Username	PSNR	**PSNR-μ**	Runtime (s)	GMACs	Param. $\times 10^3$
ALONG	Good	39.417 $_{(1)}$	37.424 $_{(1)}$	0.324	198.47	489.01
antins_cv	lishen	38.607 $_{(3)}$	37.252 $_{(2)}$	0.185	198.38	576.23
XPixel-UM	Xy_Chen	38.015 $_{(6)}$	37.209 $_{(3)}$	0.276	199.88	1013.25
AdeTeam	yeeah	39.001 $_{(2)}$	37.163 $_{(4)}$	0.134	156.12	188.99
CZCV	pepper26545	37.388 $_{(10)}$	36.972 $_{(5)}$	0.431	193.93	633.69
BOE-IOT-AIBD	NBCS	38.131 $_{(5)}$	36.851 $_{(6)}$	0.169	199.58	526.25
VAlgo	buzzli	37.888 $_{(7)}$	36.723 $_{(7)}$	0.061	175.25	1229.61
MegHDR	liuzhen	37.611 $_{(9)}$	36.608 $_{(8)}$	0.249	199.16	103.15
WorkFromHome	gaogao	37.663 $_{(8)}$	36.451 $_{(9)}$	0.057	198.91	274.13
TeamLiangJian	dth914	38.159 $_{(4)}$	36.386 $_{(10)}$	0.213	192.88	1628.27
TVHDR	leochangliu	37.221 $_{(12)}$	36.089 $_{(11)}$	0.573	199.02	753.52
ForLight	FangyaLi	37.370 $_{(11)}$	35.988 $_{(12)}$	0.226	199.39	80.65
IMVIA	steventel	36.185 $_{(13)}$	34.588 $_{(13)}$	0.045	118.99	38.83
DGU-CILAB	cilab	35.405 $_{(14)}$	34.525 $_{(14)}$	0.441	199.11	1300.60
CVIP	hannah1258	37.028	36.050	0.8876	4818.98	2485.60
no processing	-	27.408	25.266	-	-	-

Table 1. Results and rankings of methods submitted to *Track 1: Fidelity (low-complexity constraint)*. Participants are asked to optimize fidelity scores (PSNR and PSNR-μ) while keeping complexity under 200 GMACs. We denote in bold the main metric of the track.

2.3.2 Testing Phase

Participants gain access to the input sequences of the testing set and are then asked to submit a subset of their HDR predictions (67 images) to the Codalab evaluation server and e-mail the complete set (201 images), code and factsheet to the organizers. Unlike the development phase, the results and related leaderboards are not visible to the participants during the testing phase. Once the testing phase is over (*i.e.* after 7 days), the organizers verify and run the provided code to obtain the final results.

2.4. Evaluation

2.4.1 Fidelity Metrics

The evaluation of the challenge submissions is based on the computation of objective metrics between the estimated and the ground-truth HDR images. We use the well-known standard peak signal-to-noise ratio (PSNR) both directly on the estimated HDR image and after applying the μ-law tonemapper, which is a simple and canonical operator used widely for benchmarking in the HDR literature [18, 37, 50]. Within these two metrics, we selected PSNR-μ as the pri-

mary metric to rank methods in the challenge.

For the PSNR directly computed on the estimated HDR images, we normalize the values to be in the range $[0, 1]$ using the peak value of the ground-truth HDR image.

For the PSNR-μ, we apply the following tone-mapping operation $\mathrm{T}(H)$:

$$\mathrm{T}(H) = \frac{\log(1 + \mu H)}{\log(1 + \mu)} \qquad (1)$$

where H is the HDR image, and μ is a parameter that controls the compression, which we fix to $\mu = 5000$ following common HDR evaluation practices. In order to avoid excessive compression due to peak value normalization, for the PSNR-μ computation, we normalize using the 99$^{\text{th}}$ percentile of the ground-truth image followed by a tanh function to maintain the $[0, 1]$ range.

2.4.2 Complexity Metrics

To assess the complexity of the submitted solutions, we measure the number of required operations in terms of Multiply-Accumulate operations (MAC) and runtime (in

Team	Username	PSNR	PSNR-μ	Runtime (s)	**GMACs**	Param. $\times 10^3$
ALONG	Good	38.843	37.033	0.183 $_{(3)}$	74.02 $_{(1)}$	177.41
antins_cv	QianPaiChiGua	38.766	37.100	0.155 $_{(2)}$	146.28 $_{(2)}$	393.73
AdeTeam	yeeah	39.001	37.163	0.136 $_{(1)}$	156.12 $_{(3)}$	188.99
BOE-IOT-AIBD	NBCS	38.150	37.225	0.473 $_{(6)}$	756.78 $_{(4)}$	1355.10
MegHDR	liuzhen	38.749	37.030	0.269 $_{(4)}$	790.48 $_{(5)}$	401.83
0noise	Juan	38.495	37.114	0.738 $_{(7)}$	1769.85 $_{(6)}$	1222.04
Winterfell	XuZiyao	38.205	37.037	0.437 $_{(5)}$	2073.42 $_{(7)}$	1466.21
ACALJJ32	ACALJJ32	39.042	37.273	1.825 $_{(8)}$	5704.17 $_{(8)}$	8218.37
Baseline (AHDR [50])	-	37.597	37.021	0.760	2916.92	1441.28
TeamLiangJian	dth914	38.587	36.858	0.820	2260.30	20476.96
KCML2	kcml2	38.969	36.302	1.110	6000.00	2499.00
no processing	-	27.408	25.266	-	-	-

Table 2. Results and rankings of methods submitted to the *Track 2: Low-Complexity (fidelity constraint)*. Participants are asked to minimize the complexity of their solutions (GMACs and runtime) while achieving at least the same fidelity scores (PSNR and PSNR-μ) as the baseline method AHDR [50]. We denote in bold the main metric of the track.

seconds). Both metrics are self-reported by participants during the development phase but measured by the organizers during the final testing phase using a unified hardware setup, namely an Intel Xeon Platinum 8168 CPU (2.70GHz) and a single NVIDIA Tesla V100 PCIe 32 GB. We choose GMACs as the primary metric relating to low complexity for the challenges as this is independent of the platform used and thus can provide meaningful ranking during the development phase. However, we incorporate both metrics in assessing the solutions complexity in the final testing phase. For completeness, we also measured the number of parameters this metric is however not directly used for the challenge.

3. Results

From 197 registered participants in Track 1, 15 teams entered the final testing phase and submitted results, source code and fact sheets. All submissions but one complied with the eligibility criteria of under 200 GMACs constraint. As for Track 2, from 168 registered participants, 10 teams entered the final testing phase and submitted results, source code and fact sheets. In this track, two teams were not able to reach the minimal PSNR scores as set by the AHDR baseline [50] and thus are not eligible. We report the final test phase results in Table 1 and 2 for Track 1 and 2 respectively. We show more visualizations of fidelity and complexity metrics for each track in Figure 1 and 2.

The methods and the teams that entered the final phase are described in Section 4, more detailed information about each team and their members' affiliation can be found in Appendix A.

3.1. Main ideas

The emphasis on low complexity and efficiency forced participants to aggressively reduce GMACs compared to last year's challenges [35], while maintaining or even improving performance. One dominant trend is the use of depthwise and separable convolutions to reduce computational complexity. This strategy was employed by many of the top performers. The use of residual feature distillation blocks (RFDBs) is also seen as an efficient replacement for the dilated residual dense blocks (DRDB) backbone. Another trend that led to increased performance is the use of vision transformer architectures, either directly as part of the model backbone or as a teacher for knowledge distillation. A large majority of participants use a U-Net architecture due to its natural efficiency and performance advantages, and choose to process input images in feature space at a downsampled resolution or at multiple scales. Participants also use pixel shuffle layers to reduce computational complexity, and we note that participants typically trained models for more epochs when compared to the NTIRE 2021 HDR Challenge edition. Finally, the use of spatial attention, as introduced by AHDR [50], is practically ubiquitous
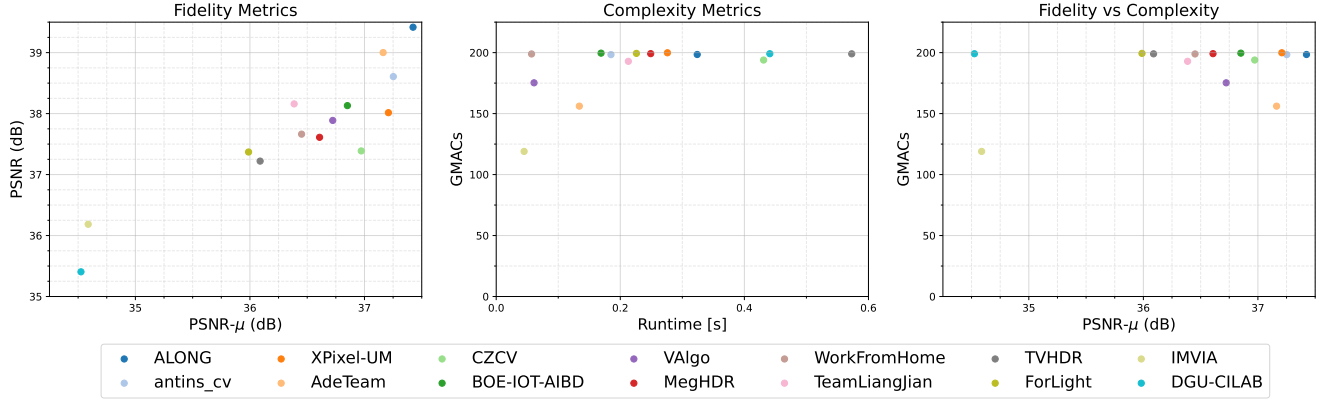
Figure 1. Several visualizations of the challenge *Track 1* methods, from left to right: combined fidelity metrics (PSNR-μ and PSNR), combined complexity metrics (GMACs and runtime) and combined fidelity and complexity metrics (GMACs and PSNR-μ).
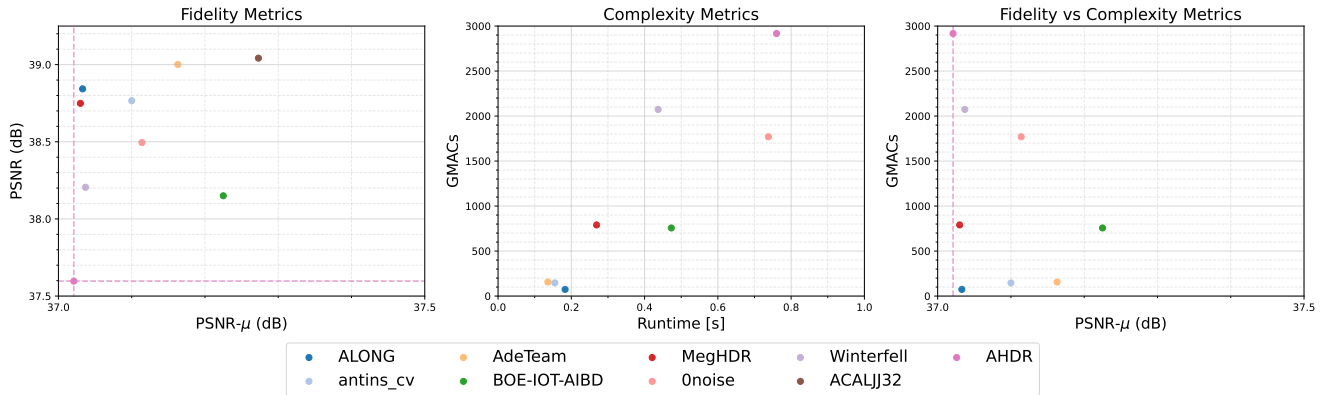

Figure 2. Several visualizations of the challenge *Track 2* methods, from left to right: combined fidelity metrics (PSNR-μ and PSNR), combined complexity metrics (GMACs and runtime) and combined fidelity and complexity metrics (GMACs and PSNR-μ).

across the different architectures.

3.2. Top results

Track 1: The top three methods are: ALONG, antins_cv and XPixel-UM. There is a moderate difference of 0.17 dB in PSNR-μ between Along and antins_cv, and a small difference of 0.04 dB between antins_cv and XPixel-UM. There is a much more significant difference in PSNR, with 0.81 dB between ALONG and antins_cv, and a further 0.59 dB between antins_cv and XPixel-UM. As expected, all methods are close to but just under 200 GMACs. Also of notable mention is the method submitted by AdeTeam which ranks fourth in PSNR-μ (by only 0.046 dB) but ranks second in PSNR and has a faster runtime and significantly fewer parameters than the top three solutions. Most of the remaining methods achieve scores of between 0.4-1.4 dB lower than the first-ranking solution.

Track 2: The top three methods are: ALONG, antins_cv and AdeTeam. ALONG achieves the required fidelity constraint using only 74.02 GMACs, while antins_cv and

AdeTeam require 146.28 and 156.12 GMACs respectively. There is a large difference between these and the remaining methods, which use between $\times 10$ and $\times 77$ as many GMACs as the first-ranked solution. We note that the first-ranked solution is able to achieve a $\times 39$ reduction in GMACs compared to the AHDR baseline while maintaining the same performance.

4. Team and Methods

4.1. ALONG

The team presents a solution that is composed of attention alignment, feature extraction and reconstruction stages, with an emphasis on efficient architecture backbones and teacher-student distillation training. Figure 3 shows an overview of the proposed approach. Firstly, 6 images (3 LDR images and 3 exposure aligned images) are fed to a pixel shuffle layer, reducing the feature spatial resolution. An efficient alignment module follows the attention module of AHDRNet [50]. The feature extraction uses an

RFDB [24] backbone with the ESA block in RFANet [25] rather than other common architectures *e.g.* DRDB [50]. Depth-wise and point-wise convolutions replace the conventional convolutions in both RFDB and the attention alignment module. Additionally, the ReLU activation is substituted by the Sigmoid Linear Unit (SiLU) [22]. Finally, the solution adopts a knowledge-distillation training strategy, where the teacher's output images and ground-truth images are used to guide the student network.
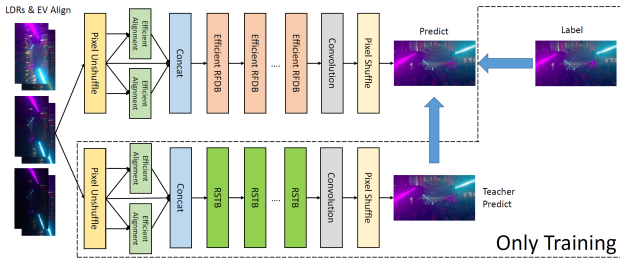

Figure 3. Architecture of EfficientHDR: Residual Feature Network and Transformer Teacher for Efficient High Dynamic Range, proposed by the ALONG team.

4.2. Antins_cv

The team proposes an efficient progressive HDR network [54] based on AHDRNet [50] and ADNet [26]. Figure 4 shows the overall architecture. Firstly, considering noise, a multi-scale encoder layer is proposed to extract high and low-frequency features from the input signals. Secondly, to deal with misalignment problems, following [53], a feature-alignment module is introduced in place of the deformable convolutional network [7], leading lower computational cost. Finally, a Progressive Dilated U-shape Block (PDUB) is proposed, composed of multiple tiny u-form structures to progressively restore features, achieving a good balance between PSNR-μ and GMACs in both Track 1 and Track 2. Compared to DRDB [50], the proposed PDUB does not work at full resolution and supports a progressive plug-and-play mode to support models with different GMACs requirements. Separable convolution [16, 17, 41] is applied throughout the main body of the network yielding improved low-complexity performance.

4.3. XPixel-UM

The team introduces a Transformer-based method for HDR imaging based on the Swin Transformer [27]. As shown in Figure 5, the overall architecture of this approach mainly consists of a large patch embedding module, a series of Swin Transformer blocks and a pixel shuffle layer. Specifically, a convolutional layer with kernel size and stride of 4 is utilized for the large patch embedding to map the concatenated input LDRs to a high-dimensional rep-
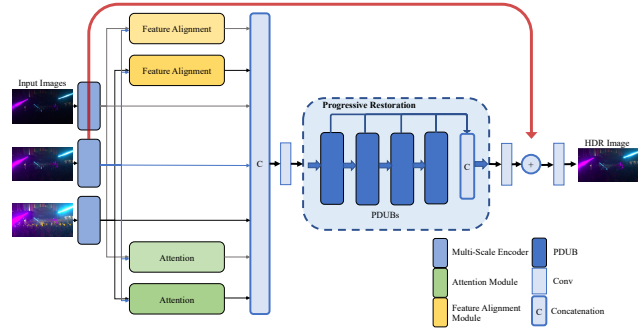

Figure 4. Architecture of Efficient Progressive High Dynamic Range Image Restoration via Attention and Alignment Network, proposed by the antins_cv team.

resentation. Then, four Swin Transformer blocks similar to [21] with a window size of 16 are adopted to extract the deep features. A long skip connection is added for better optimization, and a pixel shuffle layer is used to reconstruct the high-resolution HDR result.
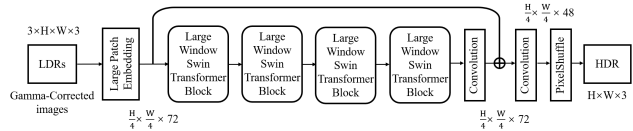

Figure 5. Architecture of SwinHDR: Efficient Swin Transformer for High Dynamic Range Imaging, proposed by the XPixel-UM team.

4.4. AdeTeam

The team presents a CNN with both high accuracy and high efficiency for reconstructing ghosting-free HDR images [51]. As shown in Figure 6, a hybrid framework is proposed to fuse high-resolution features and multi-scale features with a lightweight block. Unlike previous CNN-based methods, the high-resolution and encoder-decoder structure are integrated into a model. Following AHDRNet [50], spatial attention is used to remove motion and refine the features of non-reference images during the feature extraction stage. Additionally, a residual operation, which highlights useful information based on the dual attention module, forces the fusion stage to learn more details about degenerated regions. In the fusion stage, inspired by [13], a lightweight (LW) module fuses features with fewer parameters. For the high-resolution branch, depthwise separable convolution [6] is used to maintain high-resolution features. On the other hand, since the encoder-decoder network tends to rapidly capture a larger receptive field for HDR deghosting, the LW block is inserted into the encoder-decoder network to learn different scale features. Finally, the estimated

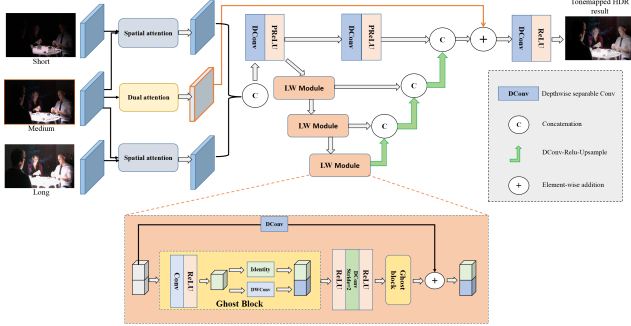image is generated with a depthwise separable convolution.

Figure 6. Architecture of A Lightweight HDR Deghosting Network with LW Module, proposed by AdeTeam.

4.5. BOE-IOT-AIBD

The team proposes a multi-level attention U-Net architecture. The overall architecture used for Track 1 is shown in Figure 7. The team adopts spatial attention modules [25, 26] to select the most appropriate regions of LDR images and corresponding exposure aligned images for feature fusion. A multi-branch U-Net [40] architecture is used to process and merge features at three different resolutions. At the low resolution, image features are concatenated and processed by feature distillation modules [24] in which feature channels are halved gradually. A 1×1 convolutional layer is used after each feature concatenation, and downsampling is implemented by the a combination of pixel shuffle and two convolutional layers of kernel size 1 and 3 respectively.

In Track 2, the framework is the same as that submitted to Track 1. To further increase the network performance, the team utilizes spatial attention and the feature fusion group (FFG) [34] that consists of several multi-attention blocks (MAB) to replace the feature distillation. The MAB makes the attention mechanisms more efficient by introducing dilated convolutions with different filter sizes.
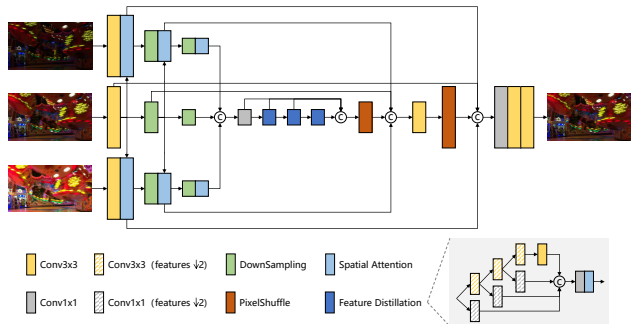

Figure 7. Architecture of Multi-Level Attention U-Net for HDR Reconstruction, proposed by the BOE-IOT-AIBD team.

4.6. CZCV

The team exploits the transformer architecture for HDR reconstruction. The proposed solution is presented in Figure 8. It consists of a convolutional feature extractor for feature extraction and a transformer for HDR reconstruction. The team builds the convolutional feature extractor using part of ADNet [26], including exposure alignment, the pyramid, cascading and deformable (PCD) alignment module [46], and the spatial attention module. The transformer part employs the recently proposed Restormer architecture [55], due to its reported low complexity and high performance. To further reduce the complexity of the model, the team replaces some regular convolution layers with depth-wise convolution followed by 1×1 convolution.
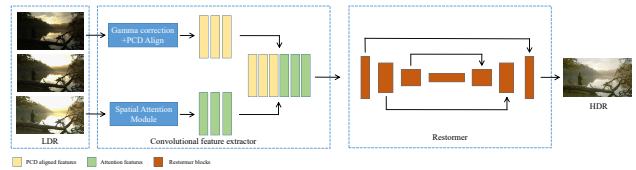

Figure 8. Architecture of Multi-frame High Dynamic Range Image Reconstruction using Transformer, proposed by the CZCV team.

4.7. MegHDR

For efficient HDR imaging, the team proposes a multi-stage feature distillation network (MFDN). The architecture of the proposed network is illustrated in Figure 9. The proposed MFDN mainly consists of two parts: 1) the feature extraction network and 2) the HDR reconstruction network. Given the input LDR images, spatial features are extracted through a feature attention module. The feature extraction network follows previous work [50], but in Track 1, the attention module is implemented using coordinate attention [15] to reduce the complexity of the model further. The extracted features are then fed into the HDR reconstruction network, reconstructing the corresponding HDR image. The reconstruction network consists of N successive residual feature distillation blocks (RFDBs). The RFDBs are adopted to distil the coarse features (with noisy, under- and over-exposed regions), and the retained features in each stage are used to reconstruct noise-free high-quality HDR results. In Track 1, N is set to 3 and the channel number of each RFDB is set to 16, whereas in Track 2, N is set to 2 and the RFDB channels is set to 32 (refer to Figure 9 for related parameter details).

4.8. 0noise

The team presents Dual Branch Residual Network for HDR Imaging (DRHDR) [31]. The overall architecture is shown in Figure 10 and comprises two routes, one
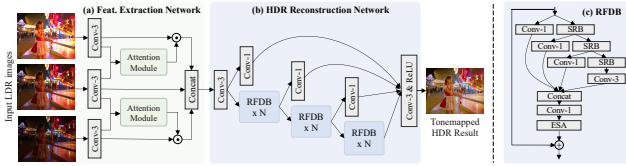
Figure 9. Architecture of Multi-stage Feature Distillation Network for Efficient High Dynamic Range Imaging, proposed by the MegHDR team.

branch that operates at full resolution, and a second branch that operates at a fourth of the original resolution. The full resolution branch adopts a deformable convolutional block [26] while the low resolution branch adopts spatial attention [50]. Both branches utilize DRDBs. Features from both branches are fused through a Dual Branch Fusion. The output layer incorporates a dense global skip connection.

Figure 10. Architecture of Dual Residual HDR (DRHDR) Network, proposed by the 0noise team.

4.9. VAlgo

The team proposes a U-Net-style network with a spatial attention mechanism (AUNet). The overall architecture is shown in Figure 11. In order to fuse information from short and long exposure inputs, the team refers to AHDRNet [50] and uses a spatial attention module to evaluate the importance of short and long images. To lower the computational complexity, inputs are downscaled to half of the original resolution and residual maps are saved to recover the final results. In the final stage, features and residual maps are concatenated to generate the final HDR image.
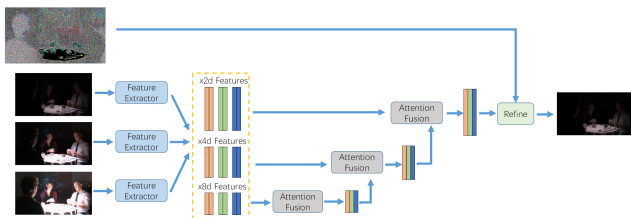

Figure 11. Architecture of AUNet for HDR, proposed by the VAlgo team.

4.10. Winterfell

The team propose a lightweight condition attention-guided reconstruction network (CARN) with multi-exposure fusion for HDR. The architecture of the proposed network is illustrated in Figure 12. First, an exposure condition module generates the condition attention maps for the entire network. The structure of the exposure condition module is derived from [4], and the input of this module involves the short, medium and long exposure alignment images. Meanwhile, the long and medium features are fused. Then, inspired by [47], condition attention (CA) is introduced to adjust these fused features with exposure condition maps. The same procedure processes the short exposed image. Hence, long-medium fusion-modulation maps and short-medium fusion-modulation maps are obtained. The medium exposed features are also sent to a CA to generate self-fusion-modulation maps. Finally, the concatenated fusion-modulation maps are fed to a reconstruction module to obtain the 16-bit HDR image. Specifically, the backbone of the reconstruction module is a variant of U-Net [39]. Dilated residual dense blocks (DRDBs) are adopted to distil the high-frequency information at a multi-scale level, and the condition attention maps are utilized at a lower scale to optimize the reconstruction maps.
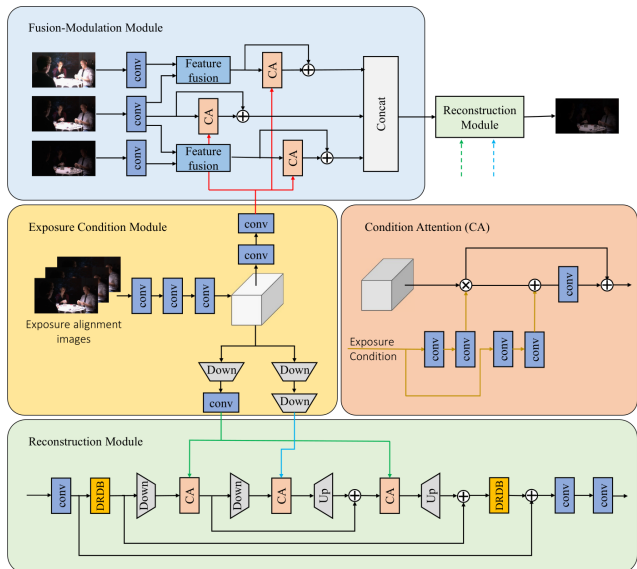

Figure 12. Architecture of Condition Attention-guided Reconstruction Network with Multi-exposure Fusion for High Dynamic Range Images, proposed by the Winterfell team.

4.11. ACALJJ32

The team presents a solution called MA-UNet to handle artifacts and colour aliasing in HDR image reconstruction. The network structure of MA-UNet is presented in

Figure 13. First, the gamma-corrected LDR images are aligned by a Pyramid, Cascading and Deformable (PCD) alignment module [46], and then the LDR images are fused by spatial attention modules [26]. Second, a U-Net structure is adopted to process the features into different scales. Instead of fusing the multi-scale features in a gradually increasing way, the proposed EAFF module [5] connects features in parallel, which can remove the colour aliasing and hallucinate reasonable details. Third, to allow useful information to propagate further, the SFT module is proposed to process the outputs from the EAFF module. To reduce computational complexity, simple concatenation is used in EAFF and SFT is not adopted during the testing phase.
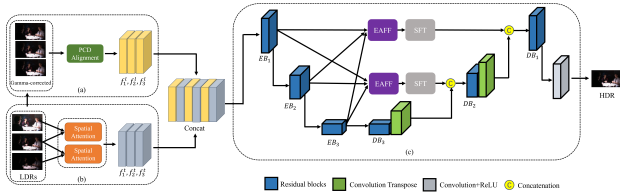

Figure 13. Architecture of Multi-scale Asymmetric Learning for High Dynamic Range Imaging, proposed by the ACALJJ32 team.

4.12. WorkFromHome

The team proposes Deep Residual U-Net for High Dynamic Range Imaging. The model architecture, shown in Figure 14, takes three LDR images as input and directly outputs the synthesized HDR image. First, three convolutional kernels extract the visual feature maps of the three different LDR inputs and then concatenate them along the channel dimension. Then these features are fed into a U-Net architecture with 2 downsampling and 2 upsampling operations. Note, the dashed shortcuts in Figure 14 are adding operations rather than concatenation operations used in regular U-Net. During this process, the image features of different exposure LDR inputs are fused thoroughly and finally transformed into the final HDR image with a regular convolutional kernel.
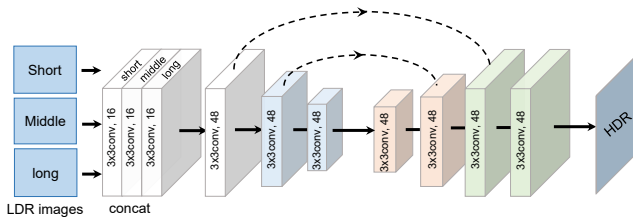

Figure 14. Architecture of Deep Residual U-Net for High Dynamic Range Imaging, proposed by the WorkFromHome team.

4.13. TeamLiangJian

The team presents a frequency-guided network (FHDR-Net) to explicitly deal with signals of different frequency sub-bands for HDR imaging (see Figure 15), using the discrete wavelet transform (DWT) to decompose the input into different frequency sub-bands. Specifically, ghosting artifacts are removed in the low-frequency sub-band, and high-frequency sub-bands are used to preserve details. The proposed network contains three parts: an encoder, a merger, and a decoder. In the encoder, the inputs are sent into three independent sub-networks. In each sub-network, the DWT is used for decomposing the feature maps into different frequency sub-bands, among which only the low-frequency sub-band is used for the next stage (scale) processing. All frequency sub-bands are also sent to the corresponding frequency-guided upsampling modules through skip connections. The merger fuses the three inputs (in the low-frequency sub-band) into a ghost-free one, which is then sent to the decoder. The merging module takes only low-frequency components from the previous stage as input and generates a merged result, focusing on structural information. In the decoder, the frequency-guided upsampling module is used to process features in the low-frequency and high-frequency sub-bands independently and then reconstruct the feature maps to a finer scale using the IDWT.
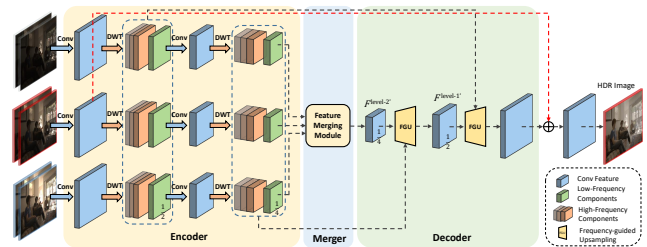

Figure 15. Architecture of Wavelet-based Neural Network for Efficient High Dynamic Range Imaging, proposed by the Team-LiangJian team.

4.14. TVHDR

The team proposes an encoder-decoder structure similar to U-Net [40]. Multi-scale features are extracted from multiple LDR images and aligned to reconstruct an HDR image. The architecture of the proposed network is illustrated in Figure 16. Firstly LDR images are aligned in terms of their relative exposure value (EV), and later fed into the encoder to extract multi-scale features. To align the multi-scale features, the team proposes a modified PCD module [46], aligning the features from low resolution to high resolution. The decoder utilizes all the aligned features to obtain the HDR image. To reduce the network complexity in the modified PCD module, conventional convolutions are replaced by depth-wise and point-wise convolutions, and the conven-

tional convolutions in the encoder and decoder are replaced with ConvNeXt blocks [28].
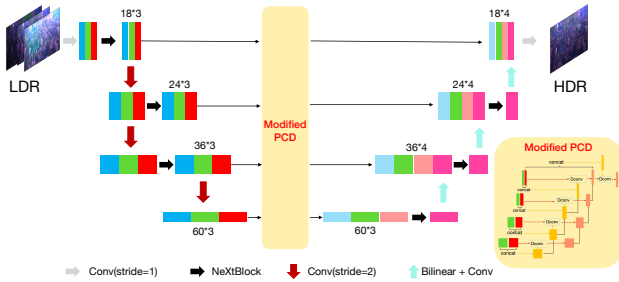

Figure 16. Architecture of HDR Reconstruction with a Lightweight Multi-Level Attention Network, proposed by the TVHDR team.

4.15. ForLight

The team proposed GSANet (Gamma-enhanced Spatial Attention Network) for HDR reconstruction [19]. An overview of the two-stage framework architecture is shown in Figure 17. The first stage performs feature extraction. The second stage aims for feature fusion and HDR reconstruction. To remove noise and ghosting effects, the gamma transformed input is processed by a small U-Net, resulting in *denoised* features. The features then undergo a spatial attention module (as proposed by [50]) and are finally fused via a channel attention [48] and a dilated residual dense block.
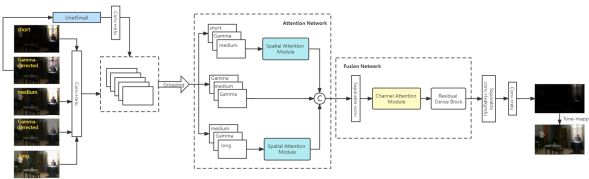

Figure 17. Architecture of HDR Reconstruction with Gamma-enhanced Spatial Attention Network, proposed by the ForLight team.

4.16. IMVIA

The team presents HDRES, a network with low computational cost using a novel merger architecture based on a cascading channel-spatial attention operation enabling a low number of parameters (38K). The architecture of the proposed network is illustrated in Figure 18. The encoder network is limited to only 16 feature maps to keep a low computational cost. Following previous work [50] attention maps are generated for each non-reference input. However,

unlike [50] the attention blocks share the same weights to limit the number of weights in the network. The merger network firstly merges the non-reference tensors and then merges this non-reference tensor with the reference tensor. The merger architecture consists of cascading operations of sum and concatenation, offering a solution of spatial and channel attention with low computational cost. Residual connections are added to the reference tensor to facilitate the network's training. A simple decoder network, composed of 3 sequential convolution operations, produces the final HDR image from the merger's 16-channel output.
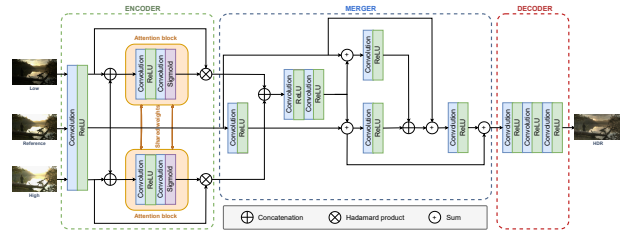

Figure 18. Architecture of HDRES: Lightweight Network for Ghost-free High Dynamic Range Imaging on Embedded System, proposed by the IMVIA team.

4.17. DGU-CILAB

The team propose a learning-based HDR imaging algorithm based on bidirectional motion estimation [44]. The architecture of the proposed network is illustrated in Figure 19. First, the motion estimation network (MENet) with cyclic cost volume and spatial attention maps estimates accurate optical flows between input LDR images. Then, the dynamic local fusion network (DLFNet) combines the warped and reference inputs to generate a synthesized output by exploiting local information. Finally, a global refinement network (GRNet) generates a residual image by using global information to improve the synthesis performance further.
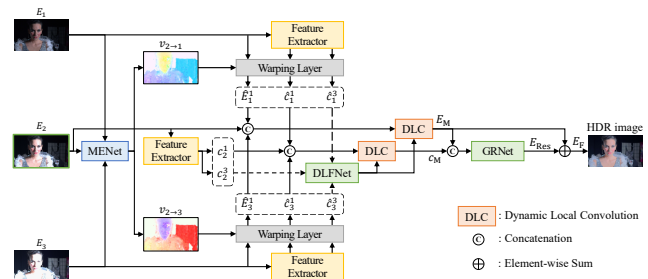

Figure 19. Architecture of Bidirectional Motion Estimation with Cyclic Cost Volume for High Dynamic Range Imaging, proposed by the DGU-CILAB team.

4.18. CVIP

The team proposes Multi-Level Attention U-Net (MLAUNet) inspired by AHDRNet [50]. The architecture of the proposed network is illustrated in Figure 20. The overall structure of MLAUNet primarily consists of two components: 1) an LDR encoder using a spatial attention module that identifies beneficial features from the long and short exposure frames, and 2) a decoder for final HDR reconstruction. First, the encoder successively downsamples the input frames into two levels, and the spatial attention module extracts attention feature maps at each level. Then, a decoder upsamples the feature maps to full resolution using skip-connections between each level of concatenated attention feature maps. The features from each level are merged using a convTranspose block and are passed through a series of residual dense blocks (DRDBs) to generate the final HDR image with a global residual learning strategy.
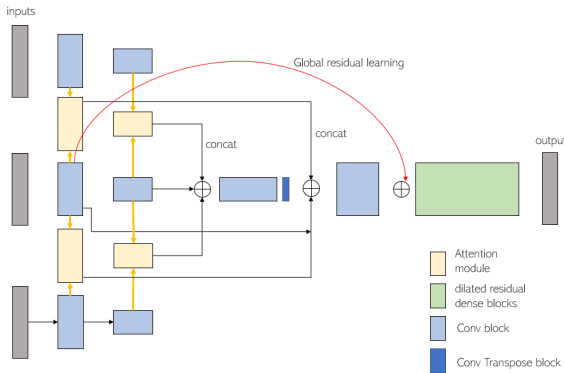

Figure 20. Architecture of Multi Level Attention U-Net, proposed by the CVIP team.

4.19. KCML2

The team presents an attention-guided non-local network (ATNLC) to deal with possible motion between LDR frames. The attention mechanism weights the contribution of each input LDR, while the non-local mechanism works for motion removal. The architecture of the proposed network is illustrated in Figure 21. The proposed ATNLC comprises four modules: attention encoder, local, global, and decoder. The attention encoder, applied to each input LDR at every stage, includes a Dual-Attention (DA) module, which in turn consists of a Spatial Attention (SA) and Channel Attention (CA) modules, evaluating the importance of each pixel and channel, respectively. The local module extracts local features at different scales with varying kernel sizes, while the global module spans a large receptive field. Finally, the HDR image is obtained by employing several decoder modules; the output of each decoder is concate-

nated with the masked-encoded features from the attention encoder before being fed to the next decoder.
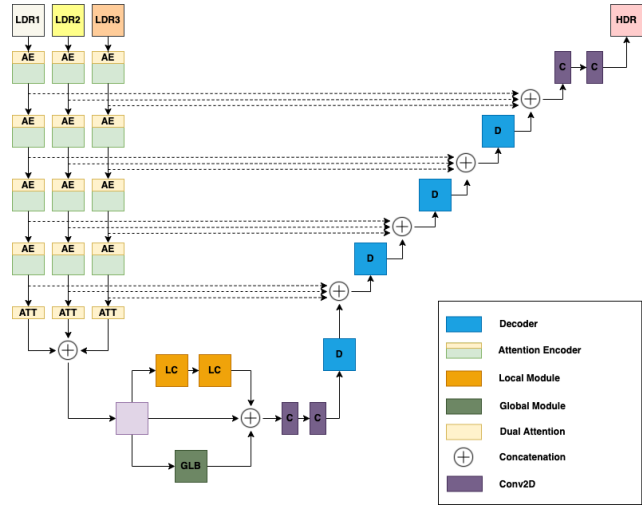

Figure 21. Architecture of Attention-guided Non-Local Network (ATNLC) for High Dynamic Range Image Synthesis, proposed by the KCML2 team.

Acknowledgments

A. Teams and Affiliations

NTIRE 2022 organizers

Title: NTIRE 2022 Challenge on High Dynamic Range Imaging
Members: Eduardo Pérez-Pellitero[1]
(e.perez.pellitero@huawei.com), Sibi Catley-Chandar[1], Richard Shaw[1], Aleš Leonardis[1], Radu Timofte[2]
Affiliations: [1] Huawei Noah's Ark Laboratory
[2] University of Würzburg and ETH Zürich

ALONG

Title: EfficientHDR: Residual Feature Network and Transformer Teacher for Efficient High Dynamic Range
Members: Zexin Zhang (zhangzexin@corp.netease.com), Cen Liu, Yunbo Peng, Yue Lin
Affiliations: Netease Games AI Lab, China

Antins_cv

Title: Efficient Progressive High Dynamic Range Image Restoration via Attention and Alignment Network
Members: Gaocheng Yu (yu-gaocheng.ygc@antgroup.com), Jin Zhang, Zhe Ma,

Hongbin Wang
Affiliations: AntGroup

XPixel-UM

Title: SwinHDR: Efficient Swin Transformer for High Dynamic Range Imaging
Members: Xiangyu Chen[1,2] (chxy95@gmail.com), Xintao Wang[3], Haiwei Wu[1], Lin Liu[4], Chao Dong[2], Jiantao Zhou[1]
Affiliations: [1] University of Macau, [2] Shenzhen Institutes of Advanced Technology, CAS, [3] Tencent PCG, [4] University of Science and Technology of China

AdeTeam

Title: A Lightweight Network for High Dynamic Range Imaging
Members: Qingsen Yan [1] (qingsenyan@gmail.com), Song Zhang[2], Weiye Chen[2], Yuhang Liu[1], Zhen Zhang[1], Yanning Zhang[3], Javen Qinfeng Shi[1], Dong Gong[4]
Affiliations: [1] The University of Adelaide, [2] Xidian University, [3] Northwestern Polytechnical University, [4] The University of New South Wales

BOE-IOT-AIBD

Title: Multi-Level Attention U-Net for HDR Reconstruction
Members: Dan Zhu (zhudan@boe.com.cn), Mengdi Sun, Guannan Chen
Affiliations: BOE Technology Group Co., Ltd., China

CZCV

Title: Multi-frame High Dynamic Range Image Reconstruction using Transformer
Members: Yang Hu (huyang@czur.com), Haowei Li, Baozhu Zou
Affiliations: CZUR Technology Group Co., Ltd., China

MegHDR

Title: Multi-stage Feature Distillation Network for Efficient High Dynamic Range Imaging
Members: Zhen Liu (liuzhen03@megvii.com), Wenjie Lin, Ting Jiang, Chengzhi Jiang, Xinpeng Li, Mingyan Han, Haoqiang Fan, Jian Sun and Shuaicheng Liu
Affiliations: Megvii Technology

0noise

Title: DRHDR: Dual Branch Residual Network for Multi-Bracket High Dynamic Range Imaging
Members: Juan Marín-Vega[1,2] (marin@imada.sdu.dk), Michael Sloth[2], Peter Schneider-Kamp[1], Richard Röttger[1]
Affiliations: [1] Department of Mathematics and Computer Science (IMADA), University of Southern Denmark, [2] Esoft Systems

VAlgo

Title: A UNet-style Network with Spatial Attention for High Dynamic Range Imaging
Members: Chunyang Li (lichunyang6@xiaomi.com), Long Bao
Affiliations: Xiaomi

Winterfell

Title: Condition Attention-guided Reconstruction Network with Multi-exposure Fusion for High Dynamic Range Images
Members: Gang He[1] (ghe@xidian.edu.cn), Ziyao Xu[1], Li Xu[1], Gen Zhan[2], Ming Sun[3], Xing Wen[3], Junlin Li[2]
Affiliations: [1] Xidian University, [2] ByteDance, [3] Kuaishou Technology

ACALJJ32

Title: Multi-scale Asymmetric Learning for High Dynamic Range Imaging
Members: Jinjing Li[1] (1792418414@qq.com), Chenghua Li[2], Ruipeng Gang[3], Fangya Li[1], Chenming Liu[3], Shuang Feng[1]
Affiliations: [1] Communication University of China, [2] Institute of Automation, Chinese Academy of Sciences, [3] Academy of Broadcasting Science, NRTA

WorkFromHome

Title: Deep Residual U-Net for High Dynamic Range Imaging
Members: Fei Lei (leifei217@gmail.com), Rui Liu, Junxiang Ruan
Affiliations: Tetras AI Technology

TeamLiangJian

Title: Wavelet-based Neural Network for Efficient High Dynamic Range Imaging
Members: Tianhong Dai[1] (tianhong.dai15@imperial.ac.uk), Wei Li[2], Zhan Lu[2], Hengyan Liu[1], Peian Huang[3], Guangyu Ren[1]
Affiliations: [1] Imperial College London, [2] Tsinghua University, [3] The University of Edinburgh

TVHDR

Title: HDR Reconstruction with a Lightweight Multi-Level Attention Network
Members: Yonglin Luo[1] (luoylin2007@126.com), Chang Liu[2], Qiang Tu[3]
Affiliations: [1] Sun Yat-sen University, [2] Shanghai Jiao Tong University, [3] Beijing Institute of Technology

ForLight

Title: Gamma-enhanced Spatial Attention Network for Efficient High Dynamic Range Imaging

Members: Fangya Li[1] (fly72109@163.com), Ruipeng Gang[2], Chenghua Li[3], Jinjing Li[1], Sai Ma[2], Chenming Liu[2], Yizhen Cao[1]
Affiliations: [1] Communication University of China, [2] Academy of Broadcasting Sciencience, NRTA, [3] Institute of Automation, Chinese Academy of Science

IMVIA

Title: HDRES: Lightweight Network for Ghost-free High Dynamic Range Imaging on Embedded System
Members: Steven Tel (steven.tel@u-bourgogne.fr), Barthelemy Heyrman, Dominique Ginhac
Affiliations: ImViA laboratory, University of Burgundy

DGU-CILAB

Title: Bidirectional Motion Estimation with Cyclic Cost Volume for High Dynamic Range Imaging
Members: Chul Lee (chullee@dongguk.edu), Gahyeon Kim, Seonghyun Park, An Gia Vien, Truong Thanh Nhat Mai
Affiliations: Department of Multimedia Engineering, Dongguk University

CVIP

Title: Multi Level Attention U-Net
Members: Howoon Yoon (hannah1258@naver.com)
Affiliations: Gachon university

KCML2

Title: An Attention Non Local Network for High Dynamic Range Image Synthesis
Members: Tu Vo (tuvv@kc-ml2.com), Alexander Holston, Sheir Zaheer, Chan Y. Park
Affiliations: KC Machine Learning Lab

References

[1] Boaz Arad, Radu Timofte, Rony Yahel, Nimrod Morag, Amir Bernat, et al. NTIRE 2022 spectral demosaicing challenge and dataset. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, 2022. 2

[2] Boaz Arad, Radu Timofte, Rony Yahel, Nimrod Morag, Amir Bernat, et al. NTIRE 2022 spectral recovery challenge and dataset. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, 2022. 2

[3] Goutam Bhat, Martin Danelljan, Radu Timofte, et al. NTIRE 2022 burst super-resolution challenge. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, 2022. 2

[4] 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 Workshops*, 2021. 8

[5] Sung-Jin Cho, Seo-Won Ji, Jun-Pyo Hong, Seung-Won Jung, and Sung-Jea Ko. Rethinking coarse-to-fine approach in single image deblurring. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 4641–4650, 2021. 9

[6] François Chollet. Xception: Deep learning with depthwise separable convolutions. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1251–1258, 2017. 6

[7] Jifeng Dai, Haozhi Qi, Yuwen Xiong, Yi Li, Guodong Zhang, et al. Deformable convolutional networks. In *Proceedings of the IEEE international conference on computer vision*, pages 764–773, 2017. 6

[8] Paul E. Debevec and Jitendra Malik. Recovering high dynamic range radiance maps from photographs. In *ACM SIGGRAPH 2008 Classes*, 2008. 1

[9] Jana Ehmann, Lun-Cheng Chu, Sung-Fang Tsai, , and Chia-Kai Liang. Real-time video denoising on mobile phones. In *IEEE International Conference on Image Processing (ICIP)*, 2018. 2

[10] Egor Ershov, Alex Savchik, Denis Shepelev, Nikola Banic, Michael S. Brown, Radu Timofte, et al. NTIRE 2022 challenge on night photography rendering. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, 2022. 2

[11] Jan Froehlich, Stefan Grandinetti, Bernd Eberhardt, Simon Walter, Andreas Schilling, et al. Creating cinematic wide gamut HDR-video for the evaluation of tone mapping operators and HDR-displays. In *Proc. of SPIE Electronic Imaging*, 2014. 1, 2

[12] Jinjin Gu, Haoming Cai, Chao Dong, Jimmy Ren, Radu Timofte, et al. NTIRE 2022 challenge on perceptual image quality assessment. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, 2022. 2

[13] Kai Han, Yunhe Wang, Qi Tian, Jianyuan Guo, Chunjing Xu, et al. GhostNet: More features from cheap operations. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 1580–1589, 2020. 6

[14] Samuel W. Hasinoff, Frédo Durand, and William T. Freeman. Noise-optimal capture for high dynamic range photography. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2010. 2

[15] Qibin Hou, Daquan Zhou, and Jiashi Feng. Coordinate attention for efficient mobile network design. In *CVPR*, 2021. 7

[16] Andrew G. Howard, Mark Sandler, Grace Chu, Liang-Chieh Chen, Bo Chen, et al. Searching for MobileNetV3. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 1314–1324, 2019. 6

[17] Andrew G. Howard, Menglong Zhu, Bo Chen, Dmitry Kalenichenko, Weijun Wang, et al. MobileNets: Efficient convolutional neural networks for mobile vision applications. *arXiv preprint arXiv:1704.04861*, 2017. 6

[18] Nima Khademi Kalantari and Ravi Ramamoorthi. Deep high dynamic range imaging of dynamic scenes. *ACM Transactions on Graphics (Proceedings of SIGGRAPH 2017)*, 36(4), 2017. 1, 2, 3

[19] Fangya Li, Ruipeng Gang, Chenghua Li, Jinjing Li, Sai Ma, et al. Gamma-enhanced spatial attention network for efficient high dynamic range imaging. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, 2022. 10

[20] Yawei Li, Kai Zhang, Radu Timofte, Luc Van Gool, et al. NTIRE 2022 challenge on efficient super-resolution: Methods and results. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, 2022. 2

[21] Jingyun Liang, Jiezhang Cao, Guolei Sun, Kai Zhang, Luc Van Gool, and Radu Timofte. SwinIR: Image restoration using swin transformer. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 1833–1844, 2021. 6

[22] Zudi Lin, Prateek Garg, Atmadeep Banerjee, Salma Abdel Magid, Deqing Sun, et al. Revisiting RCAN: Improved training for image super-resolution. *arXiv preprint arXiv:2201.11279*, 2022. 6

[23] Jerrick Liu, Oliver Nina, Radu Timofte, et al. NTIRE 2021 multi-modal aerial view object classification challenge. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, 2021. 2

[24] Jie Liu, Jie Tang, and Gangshan Wu. Residual feature distillation network for lightweight image super-resolution. In *European Conference on Computer Vision*, pages 41–55. Springer, 2020. 6, 7

[25] Jie Liu, Wenjie Zhang, Yuting Tang, Jie Tang, and Gangshan Wu. Residual feature aggregation network for image super-resolution. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 2359–2368, 2020. 6, 7

[26] Zhen Liu, Wenjie Lin, Xinpeng Li, Qing Rao, Ting Jiang, et al. ADNet: Attention-guided deformable convolutional network for high dynamic range imaging. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, 2021. 6, 7, 8, 9

[27] Ze Liu, Yutong Lin, Yue Cao, Han Hu, Yixuan Wei, et al. Swin transformer: Hierarchical vision transformer using shifted windows. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 10012–10022, 2021. 6

[28] Zhuang Liu, Hanzi Mao, Chao-Yuan Wu, Christoph Feichtenhofer, Trevor Darrell, et al. A ConvNet for the 2020s. *arXiv preprint arXiv:2201.03545*, 2022. 10

[29] Andreas Lugmayr, Martin Danelljan, Radu Timofte, et al. NTIRE 2022 challenge on learning the super-resolution space. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, 2022. 2

[30] Matteo Maggioni, Yibin Huang, Cheng Li, Shuai Xiao, Zhongqian Fu, and Fenglong Song. Efficient multi-stage video denoising with recurrent spatio-temporal fusion. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2021. 2

[31] Juan Marín-Vega, Michael M. S. L. Sloth, Peter Schneider-Kamp, and Richard Röttger. DRHDR: A dual branch residual network for multi-bracket high dynamic range imaging. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, 2022. 7

[32] Morgan McGuire, Wojciech Matusik, Hanspeter Pfister, Billy Chen, John F. Hughes, et al. Optical splitting trees for high-precision monocular imaging. *IEEE Computer Graphics and Applications*, 27(2), 2007. 1

[33] Tom Mertens, Jan Kautz, and Frank Van Reeth. Exposure fusion. In *Pacific Conference on Computer Graphics and Applications*, 2007. 1

[34] Abdul Muqeet, Jiwon Hwang, Subin Yang, Jung Heum Kang, Yongwoo Kim, et al. Multi-attention based ultra lightweight image super-resolution. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, 2020. 7

[35] Eduardo Pérez-Pellitero, Sibi Catley-Chandar, Ales Leonardis, Radu Timofte, et al. NTIRE 2021 challenge on high dynamic range imaging: Dataset, methods and results. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, 2021. 2, 4

[36] 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 *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, 2022. 2

[37] K. Ram Prabhakar, Susmit Agrawal, Durgesh Kumar Singh, Balraj Ashwath, and R. Venkatesh Babu. Towards practical and efficient high-resolution HDR deghosting with CNN. In *European Conference Computer Vision*, volume 12366 of *Lecture Notes in Computer Science*, pages 497–513. Springer, 2020. 2, 3

[38] Andres Romero, Angela Castillo, Jose M Abril-Nova, Radu Timofte, et al. NTIRE 2022 image inpainting challenge: Report. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, 2022. 2

[39] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-Net: Convolutional networks for biomedical image segmentation. In *MICCAI*, 2015. 8

[40] Green Rosh K. S., Sachin Deepak Lomte, Nikhil Krishnan Li, Jian Sun, and B. H. Pawan Prasad. HDR merging using multi branch residual networks. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, 2021. 7, 9

[41] Mark Sandler, Andrew G. 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. 6

[42] Pradeep Sen, Nima Khademi Kalantari, Maziar Yaesoubi, Soheil Darabi, Dan B. Goldman, et al. Robust patch-based HDR reconstruction of dynamic scenes. *ACM Trans. Graph.*, 31(6):203–1, 2012. 1

[43] Michael D. Tocci, Chris Kiser, Nora Tocci, and Pradeep Sen. A versatile HDR video production system. In *SIGGRAPH*, 2011. 1

[44] An Gia Vien, Seonghyun Park, Truong T. N. Mai, Gahyeon Kim, and Chul Lee. Bidirectional motion estimation with cyclic cost volume for high dynamic range imaging. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, 2022. 10

[45] Longguang Wang, Yulan Guo, Yingqian Wang, Juncheng Li, Shuhang Gu, Radu Timofte, et al. NTIRE 2022 challenge on stereo image super-resolution: Methods and results. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, 2022. 2

[46] Xintao Wang, Kelvin C. K. Chan, Ke Yu, Chao Dong, and Chen Change Loy. EDVR: Video restoration with enhanced deformable convolutional networks. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, June 2019. 7, 9

[47] Xintao Wang, Ke Yu, Chao Dong, and Chen Change Loy. Recovering realistic texture in image super-resolution by deep spatial feature transform. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2018. 8

[48] Sanghyun Woo, Jongchan Park, Joon-Young Lee, and In So Kweon. CBAM: Convolutional block attention module. In *Proceedings of the European conference on computer vision (ECCV)*, pages 3–19, 2018. 10

[49] Shangzhe Wu, Jiarui Xu, Yu-Wing Tai, and Chi-Keung Tang. End-to-end deep HDR imaging with large foreground motions. In *European Conference on Computer Vision*, 2018. 1, 2

[50] Qingsen Yan, Dong Gong, Qinfeng Shi, Anton van den Hengel, Chunhua Shen, et al. Attention-guided network for ghost-free high dynamic range imaging. In *Computer Vision and Pattern Recognition*, 2019. 1, 2, 3, 4, 5, 6, 7, 8, 10, 11

[51] Qingsen Yan, Song Zhang, Weiye Chen, Yuhang Liu, Zhen Zhang, et al. A lightweight network for high dynamic range imaging. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, 2022. 6

[52] Ren Yang, Radu Timofte, et al. NTIRE 2022 challenge on super-resolution and quality enhancement of compressed video: Dataset, methods and results. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, 2022. 2

[53] Lucas D. Young, Fitsum A. Reda, Rakesh Ranjan, Jon Morton, Jun Hu, et al. 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. 6

[54] Gaocheng Yu, Jin Zhang, Zhe Ma, and Hongbin Wang. Efficient progressive high dynamic range image restoration via attention and alignment network. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, 2022. 6

[55] Syed Waqas Zamir, Aditya Arora, Salman Khan, Munawar Hayat, Fahad Shahbaz Khan, et al. Restormer: Efficient transformer for high-resolution image restoration. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2022. 7