

# NTIRE 2021 Challenge on High Dynamic Range Imaging: Dataset, Methods and Results

Eduardo Pérez-Pellitero <sup>†</sup>	Sibi Catley-Chandar <sup>†</sup>	Aleš Leonardis <sup>†</sup>	Radu Timofte <sup>†</sup>		
Xian Wang	Tao Wang	Fenglong Song	Zhen Liu	Wenjie Lin	
Xinpeng Li	Qing Rao	Ting Jiang	Mingyan Han	Haoqiang Fan	Jian Sun
Shuaicheng Liu	Xiangyu Chen	Yihao Liu	Zhengwen Zhang	Yu Qiao	
Chao Dong	Evelyn Yi Lyn Chee	Shanlan Shen	Yubo Duan	Guannan Chen	
Mengdi Sun	Yan Gao	Lijie Zhang	Akhil K A	Jiji C V	S M A Sharif
Rizwan Ali Naqvi	Mithun Biswas	Sungjun Kim	Chenjie Xia	Bowen Zhao	
Zhangyu Ye	Xiwen Lu	Yanpeng Cao	Jiangxin Yang	Yanlong Cao	
Green Rosh K S	Sachin Deepak Lomte	Nikhil Krishnan	B H Pawan Prasad		

## Abstract

*This paper reviews the first challenge on high-dynamic range (HDR) imaging that was part of the New Trends in Image Restoration and Enhancement (NTIRE) workshop, held in conjunction with CVPR 2021. This manuscript focuses on the newly introduced dataset, the proposed methods and their results. The challenge aims at estimating a HDR image from one or 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 by two tracks: In Track 1 only a single LDR image is provided as input, whereas in Track 2 three differently-exposed LDR images with inter-frame motion are available. In both tracks, the ultimate goal is to achieve the best objective HDR reconstruction in terms of PSNR with respect to a ground-truth image, evaluated both directly and with a canonical tonemapping operation.*

## 1. Introduction

Current consumer-grade cameras struggle to capture scenes with varying illumination with a single exposure shot due to the inherent limitations of the imaging sensor, which suffers from saturation in high-irradiance regions and from uncertainty in the readings for low-light regions.

In recent years, advances in computational photogra-

phy have enabled single-sensor cameras to acquire images with an extended dynamic range without the need of expensive, bulky and arguably more inconvenient multi-camera rigs, *e.g.* [11, 23, 35]. Generally, those algorithms exploit multiple LDR frames captured with different exposure values (EV) that are then fused into a single HDR image [8, 24], with some of those methods including frame alignment [15, 29, 38] or pixel rejection strategies [39].

Convolutional Neural Networks (CNNs) have greatly advanced the state-of-the-art for HDR reconstruction, especially for complex dynamic scenes [15, 27, 38, 39]. Additionally, CNNs have opened a new path into single-image HDR imaging thanks to their ability to learn complex and entangled vision tasks seamlessly, *e.g.* denoising, camera response function estimation, image in-painting, high-frequency and detail hallucination [20]. Despite the ill-posed nature of the single-image HDR reconstruction, most current methods obtain plausible results that, if not as accurate as those reconstructed from multiframe LDR images, can be a good alternative when multiple frames are not available or can not be captured due to time constraints.

The NTIRE 2021 HDR Challenge aims at stimulating research for computational HDR imaging, and better understand the state-of-the-art landscape for both single and multiple frame HDR processing. It is part of a wide spectrum of associated challenges with the NTIRE 2021 workshop: non-homogeneous dehazing [3], defocus deblurring using dual-pixel [2], depth guided image relighting [10], image deblurring [25], multi-modal aerial view imagery classification [19], learning the super-resolution space [22], quality enhancement of heavily compressed videos [40], video super-resolution [32], perceptual image quality assessment [12], burst super-resolution [4], high dynamic

<sup>†</sup> Eduardo Pérez-Pellitero ([e.perez.pellitero@huawei.com](mailto:e.perez.pellitero@huawei.com)), Sibi Catley-Chandar, Aleš Leonardis (Huawei Noah’s Ark Laboratory) and Radu Timofte (ETH Zurich) are the NTIRE 2021 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/ntire21/>

range [26].

## 2. Challenge

The NTIRE 2021 HDR Challenge is the first edition that addresses the HDR image enhancement task. This challenge aims to gauge and advance the state-of-the-art on HDR imaging. It is focused specially in challenging scenarios for HDR image reconstruction, *i.e.* wide range of scene illumination, accompanied by complex motions in terms of camera, scene and light sources. In this section we present details about the new dataset used for the challenge, as well as how the challenge tracks are designed.

### 2.1. Dataset

Both training and evaluation of HDR imaging algorithms require high quality annotated datasets. Specially for deep learning methods, the number of training examples and their diversity in terms of *e.g.* scene and camera motion, exposure values, textures, semantic content, is of crucial importance for the model performance and generalization capabilities. Creating a high quality HDR dataset with such features still poses several challenges. Current HDR datasets are generally captured using static image bracketing, with some efforts towards controlling the scene motion so that *stop-motion* dynamic scenes can be assembled: In the work of Kalantari *et al.* [15] a subject is asked to stay still in order to capture three bracketed exposure images on a tripod used to generate ground-truth, and afterwards two additional images are captured while the subject is asked to move, obtaining therefore a input LDR triplet with inter-frame motion and a reference HDR ground-truth image aligned to the central frame. Such capturing approaches are normally limited to small datasets, as this type of capturing is time consuming, and additionally it constrains the motions that can be captured while misalignment might still happen if the subject is not completely still.

For this challenge we introduce a newly curated HDR dataset. This dataset is 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 HDR videos using a professional two-camera rig with a semitransparent mirror for the purpose of HDR display evaluation. The contents of those videos include naturally challenging HDR scenes: *e.g.* moving light sources, brightness changes over time, high contrast skin tones, specular highlights and bright, saturated colors. As these images lack the necessary LDR input images, similarly to [16], we synthetically generate the respective LDR counterparts by fol-

lowing accurate image formation models that include several noise sources [13].

**Image Formation Model:** In order to model the HDR to LDR step, we use the following pixel measurement model as described in [13]:

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

where  $I_l$  is an LDR image,  $\Phi$  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. For our data processing, we assume  $\Phi$  to be well approximated by the ground-truth HDR images, and produce different LDR images by modifying the exposure time  $t$  parameter of any three consecutive frames.

**Noise Model:** In order to realistically reproduce the characteristic of common LDR images, we include a zero-mean signal whose variance comes from three independent sources, *i.e.* photon noise, read noise and analog-to-digital (ADC) gain and quantization noise (for 8-bit LDR images). For pixels under the saturation level, the variance of  $n$  reads:

$$\text{Var}(n) = \Phi t / g^2 + \sigma_{read}^2 / g^2 + \sigma_{ADC}^2 \quad (2)$$

Note that first photon-noise term is signal-dependent (normally represented by a Poisson distribution), while the read-noise term is gain-dependent.

We show in Figure 1 some examples of the HDR and the synthetically generated LDR images.

**Partitions:** We provide training, testing and validation data splits. With our synthetically processed set, we manually discard images to balance the number of frames per scene and to remove undesirable frames, mostly due to *e.g.* dominant presence of lights, lack of inter-frame motion, excessive presence of noise in the HDR image. This leads to roughly 1750 frames within 29 different scenes. The validation and testing splits are obtained randomly from 4 different scenes (*carousel fireworks 02*, *fireplace 02*, *fishing longshot*, *poker travelling slowmotion*) while the other 25 scenes are used for the training set, ensuring that there is no scene overlap between training and testing/validation. This results on a training set short of 1500 examples, and a validation and testing set of 60 and 201 examples respectively.

### 2.2. Challenge Design and Tracks

This challenge is organized into two different tracks, both of them sharing the same evaluation metrics and ground-truth data. The results from both tracks are thus directly comparable and can explain the performance differences between single and multi-frame HDR imaging.

#### 2.2.1 Track 1: Single Frame

This track evaluates the HDR reconstruction when only a single LDR frame is available. In contrast to the multi-frame approaches, single image HDR methods have only

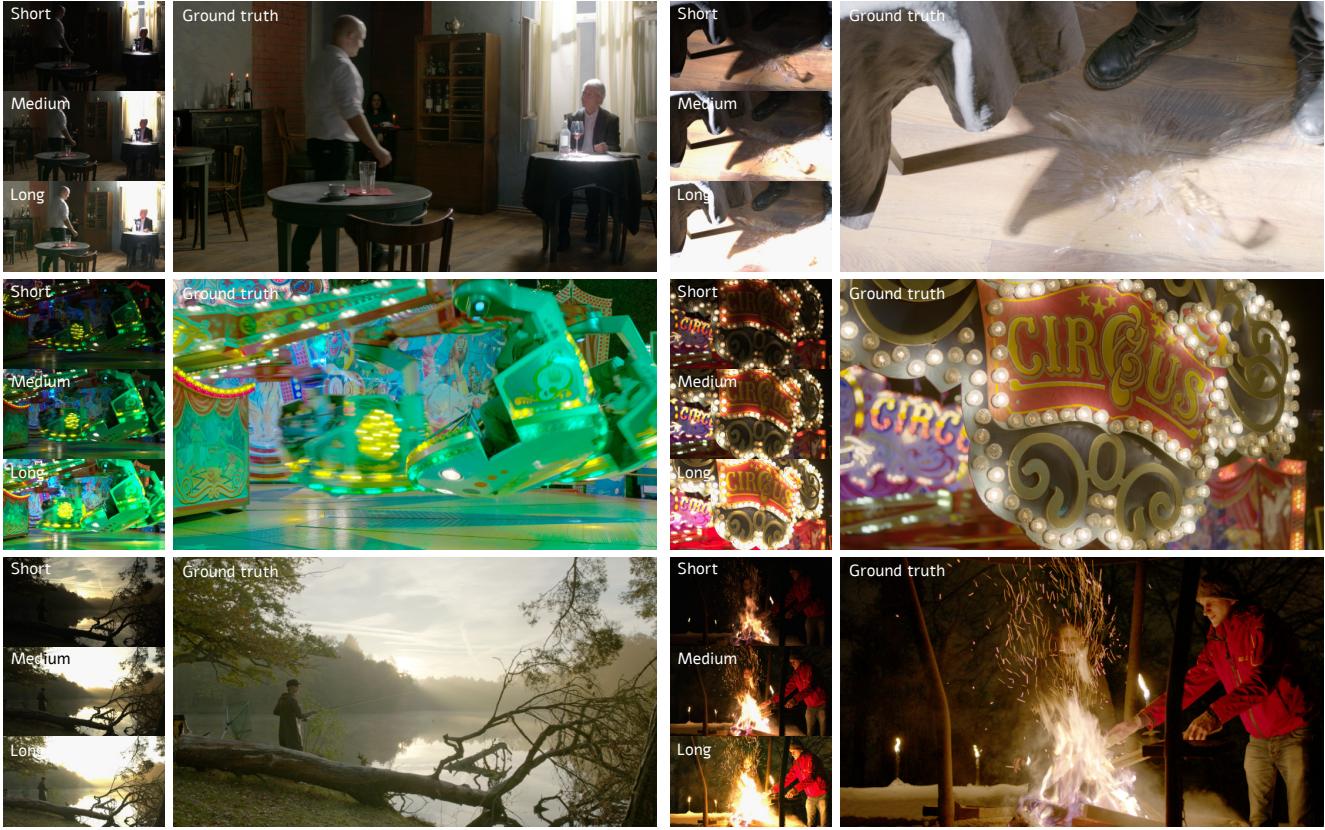


Figure 1. Visualizations of samples included in the newly curated training, validation and testing datasets. Each training example is composed by three input LDR images (short, medium and long) and a related ground-truth image aligned with the medium input LDR. Note that validation and testing ground-truth images were not made available to participants.

a single exposure (instead of a bracketed set) which is arguably more challenging when recovering under- and overexposed regions as no information from neighboring frames at different EVs can be leveraged. Similarly, single image denoising poses further challenges than its multiple frame counterpart as noise sources are of zero mean and less observations are available. On the other side, this track does not have to deal with motion related artifacts, *e.g.* ghosting, bleeding edges, which are common in the multiframe setup.

### 2.2.2 Track 2: Multiple Frame

This track evaluates the HDR reconstruction for three differently exposed LDR images (*i.e.* short, medium, long) with diverse motion between the respective frames, including camera motion, non-rigid scene motion with an emphasis on complex moving and changing light sources. The bracketed input frames were separated by steps of 2 or 3 EV between them, similarly to other existing datasets [15]. In order to enable direct comparison between both tracks, the medium LDR frame in Track 2 corresponds to the single-frame LDR input on Track 1 and thus both tracks share the

same ground-truth data.

### 2.3. Evaluation

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  $\mu$ -law tonemapper, which is a simple and canonical operator used widely for benchmarking in the HDR literature [15, 27, 39]. From within these two metrics, we selected PSNR- $\mu$  as the main 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- $\mu$ , we apply the following tone-mapping operation  $T(H)$ :

$$T(H) = \frac{\log(1 + \mu H)}{\log(1 + \mu)} \quad (3)$$

where  $H$  is the HDR image, and  $\mu$  is a parameter that controls the compression, which we fix to  $\mu = 5000$  following

common HDR evaluation practises. In order to avoid excessive compression due to peak value normalization, for the  $\text{PSNR-}\mu$  computation we normalize using the 99 percentile of the ground-truth image followed by a tanh function to maintain the  $[0, 1]$  range.

### 3. Results

From 120 registered participants in Track 1, 16 teams participated during the development phase and finally 7 teams entered the final testing phase and submitted results and fact sheets. As for Track 2, from 126 registered participants, 28 teams participated during the development phase and finally 6 teams entered the final testing phase and submitted results and fact sheets. We report the final test phase results in Table 1 and 2 for track 1 and 2 respectively. A visualization of both metrics for each track separately can be found in Figure 2 and 3, and all the results from both tracks are aggregated in Figure 4. The methods and the teams that entered the final phase are described in Section 4, more detailed information about each team and their member’s affiliation can be found in Appendix A.

#### 3.1. Main ideas

In the single frame track, the majority of the proposed architectures consist of several sub-networks which aim to reverse single aspects of the HDR to LDR image pipeline, perhaps inspired by the success of Liu *et al.* [20]. Variants of the Residual Dense Block [43] are the most commonly used backbone although U-Net style architectures are used by a significant minority. In addition to the standard  $\ell_1$  loss, some methods also use perceptual colour losses.

In the multiple frames track, a major number of solutions are inspired by Yan *et al.* AHDRNet [39], with most submissions using their attention mechanism. Most methods also adopt the Dilated Residual Dense Block, although similarly to Track 1, U-Net style architectures with non-dense residual blocks are also present and achieve competitive performance. Ensemble approaches to improve performance via test time augmentations such as flips/transpose [34] are common among the participants, leading to increases of up to 0.5 dB. Some submissions aim to explicitly align input images instead of just rejecting unaligned regions with attention, including the first-ranked submission which aligns images using deformable convolutions [7, 21].

#### 3.2. Top results

**Track 1:** The top two methods (NOAHTCV and XPixel) obtain similar  $\text{PSNR-}\mu$  scores, only about 0.07 dB apart, while in terms of PSNR the difference is more noticeable, in the range of 0.6 dB. BOE-IOT-AIBD comes third in terms of  $\text{PSNR-}\mu$ , at around 0.4 dB gap to the first position, however they are ranked first in terms of PSNR by a notice-

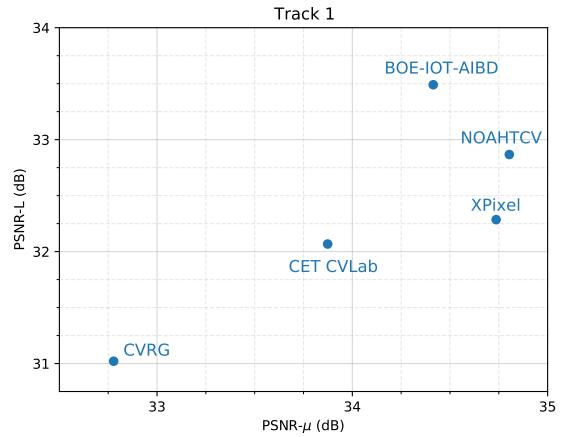


Figure 2. Combined  $\text{PSNR-}\mu$  and PSNR values of methods from the *Track 1 (single frame)*.

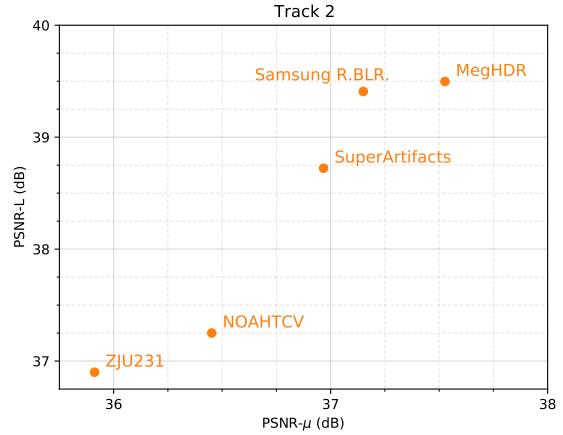


Figure 3. Combined  $\text{PSNR-}\mu$  and PSNR values of methods from the *Track 2 (multiple frames)*.

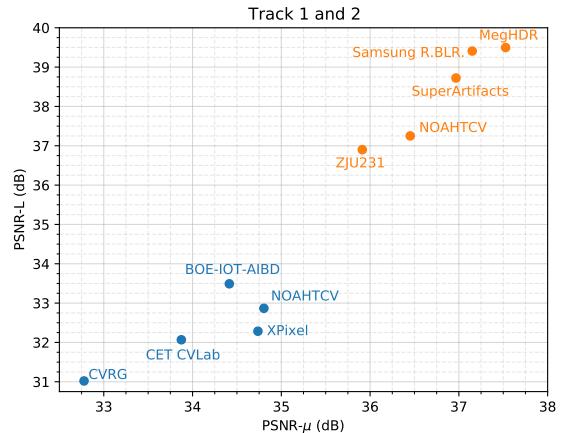


Figure 4. Combined  $\text{PSNR-}\mu$  and PSNR values of methods from the *Track 1 (in blue)* and *Track 2 (in orange)*.

able margin (0.6 dB) to the second best-performer in that metric (NOAHTCV). The rest of competing teams obtain scores within 2 and 1 dB gaps when compared to the best-performer in terms of PSNR and PSNR- $\mu$  respectively.

**Track 2:** In this track, both metrics behave similarly and exactly the same ranking is obtained with either of them. The MegHDR team obtains the first position, with a lead of 0.56 dB in terms of PSNR- $\mu$  and a broader difference of 0.78 in terms of PSNR when compared against the runner-up team in the leaderboard (SuperArtifacts). NOAHTCV follows at roughly 0.5 dB and 0.8 dB performance gap with respect to MegHDR in terms of PSNR- $\mu$  and PSNR respectively. The rest of competing teams obtain scores within 1.6 and 2.6 dB gaps when compared to the best-performer in terms of PSNR and PSNR- $\mu$  respectively.

## 4. Team and Methods

### 4.1. NOAHTCV

NOAHTCV have proposed two methods, one for single frame and one for multi-frame. Both methods are discussed here.

**Single Image HDR Reconstruction in a Multi-stage Manner** The team propose a multi-stage method which decomposes the problem into two sub-tasks; denoising and hallucination. The input image,  $I$  is first passed through a denoising network to get the denoised image  $D$ . Both  $I$  and  $D$  are processed by the hallucination network to obtain  $H$ . Finally  $I$ ,  $D$  and  $H$  are fused by a refinement network. The general architecture can be seen in Figure 5. MIRNet [42] is employed as the denoising network, while the hallucination network uses masked features as in [28] to reconstruct details in the over-exposed regions. The refinement network is a U-Net equipped with coordinate attention [14].

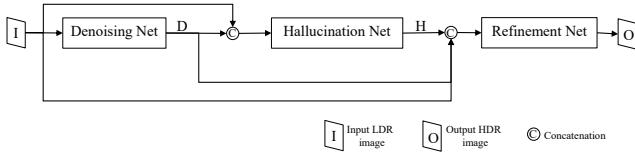


Figure 5. Architecture of Single Image HDR Reconstruction in a Multi-stage Manner, proposed by the NOAHTCV team.

**Alignment Augmentation and Multi-Attention Guided HDR Fusion** The team propose a three stage method consisting of an Alignment and Augmentation module, an Attention Based Information Extraction module, and an Enhancement and Fusion module. The architecture can be seen in Figure 6. The Alignment and Augmentation module uses a pretrained PWC-Net [33] to warp the short and long input images with a predicted optical flow. Both the original images and warped images are fed into the network. The Attention Based Information Extraction module employs the occluded attention

mechanism from AHDRNet [39] to reduce misalignment distortion. Channel attention is also used on shallow features extracted by a shared convolutional layer to re-weight features generated by different frames. The Enhancement and Fusion module employs the network architecture from AHDRNet [39] with the final sigmoid layer removed.

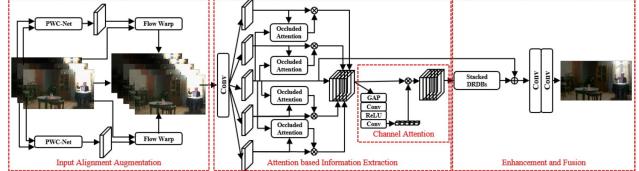


Figure 6. Architecture of Alignment Augmentation and Multi-Attention Guided HDR Fusion, proposed by the NOAHTCV team.

### 4.2. MegHDR

**ADNet: Attention-guided Deformable Convolutional Networks for High Dynamic Range Imaging** The team propose ADNet [21], a novel multi-frame imaging pipeline where the LDR images and their corresponding gamma-corrected images are processed separately, instead of being concatenated together. This is motivated by the intuition that images in the LDR domain are helpful for detecting noisy or saturated regions, while images in the HDR domain help to detect misalignment. The PCD align module aligns the gamma corrected images using pyramid, cascading and deformable convolutions based on EDVR [36]. The spatial attention module suppresses undesired saturation and noisy regions in the LDR images while highlighting the regions useful for fusion. The resulting features are concatenated and processed by dilated residual dense blocks (DRDBs) as in AHDRNet [39]. The architecture can be seen in Figure 7.

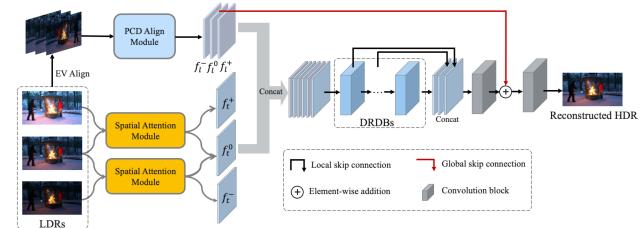


Figure 7. Architecture of ADNet, proposed by the MegHDR team.

### 4.3. XPixel

**HDRUNet: Single Image HDR Reconstruction with Denoising and Dequantization** The team propose HDRUNet [6], which consists of three sub-networks: the base network, the condition network and the weighting network. The architecture can be seen in Figure 8. The base network is a U-Net style encoder-decoder model. The condition network and spatial feature transform (SFT) layers [37] are introduced to achieve adaptive modulation

Team	Username	PSNR- $\mu$	PSNR	Runtime (s)	GPU	Ensemble
NOAHTCV	noahtcv	34.804 (1)	32.867 (2)	61.52 (5)	Tesla P100	flips, transpose
XPixel	Xy_Chen	34.736 (2)	32.285 (3)	0.53 (2)	RTX 2080 Ti	-
BOE-IOT-AIBD	chenguannan1981	34.414 (3)	33.490 (1)	5.00 (4)	Tesla V100	-
CET CVLab	akhilkashok	33.874 (4)	32.068 (4)	0.20 (1)	Tesla P100	flips, rotation
CVRG	sharif_apu	32.778 (5)	31.021 (5)	1.10 (3)	GTX 1060	-
<i>no processing</i>	-	25.266 (6)	27.408 (6)	-	-	-

Table 1. Results and rankings of methods submitted to the *Track 1: Single frame HDR*. Please note that running times are self-reported.

Team	Username	PSNR- $\mu$	PSNR	Runtime (s)	GPU	Ensemble
MegHDR	liuzhen	37.527 (1)	39.497 (1)	1.35 (2)	RTX 2080 Ti	flips, transpose
SuperArtifacts	evelynchee	36.968 (2)	38.723 (2)	3.80 (4)	RTX 2080 Ti	-
NOAHTCV	noahtcv	36.452 (3)	37.250 (3)	1.26 (1)	Tesla V100	-
ZJU231	ZJU231	35.912 (4)	36.900 (4)	2.96 (3)	RTX 2080 Ti	flips, rotation, $\times 4$ models
Samsung Research Bangalore*	AnointedKnight	37.151	39.408	15.77	Tesla P40	flips, transpose
<i>no processing</i>	-	25.266 (5)	27.408 (5)	-	-	-

Table 2. Results and rankings of methods submitted to the *Track 2: Multiple frames HDR*. Please note that running times are self-reported.

based on the features being processed. Besides, inspired by [9], a mask is calculated for the global residual, as adding it directly is sub-optimal. Finally, a tanh  $\ell_1$  loss function is adopted to balance the impact of over-exposed values and well-exposed values on the network learning.

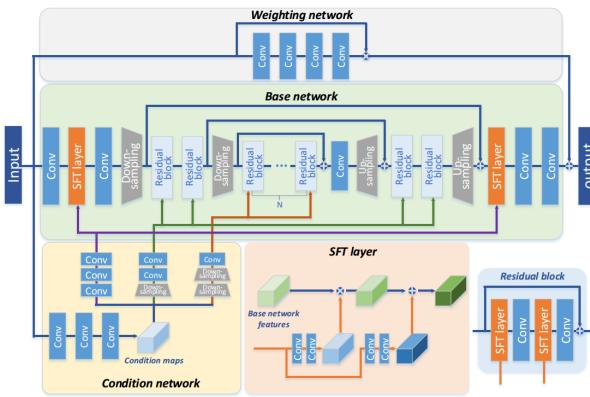


Figure 8. Architecture of HDRUNet: Single Image HDR Reconstruction with Denoising and Dequantization, proposed by the XPixel team.

\* Incomplete submission (no reproducibility) thus not in the challenge ranking.

#### 4.4. BOE-IOT-AIBD

**Task-specific Network based on Channel Adaptive RDN** The team propose a method [5] which consists of three sub-networks which each perform a different task: Image Reconstruction (IR), Detail Restoration (DR) and Local Contrast Enhancement (LCE) [17]. The IR network reconstructs the coarse HDR image from the input LDR image. The DR network can further refine the image details by adding its output to the coarse HDR output of IR. Finally the LCE network predicts a luminance equalization mask which is multiplied by the refined HDR image for contrast adjustment. The total architecture can be seen in Figure 9. All three sub-networks use the same backbone, named the Channel Adaptive RDN. This consists of the standard Residual Dense Network [43] with the Gate Channel Transformation layer [41] added to each RDB block.

#### 4.5. SuperArtifacts

**Multi-Level Attention on Multi-Exposure Frames for HDR Reconstruction** The team propose a multi-level architecture which processes and merges features at three different resolutions. On top of the architecture of AHDR-Net [39], the model encodes the frames into three levels, with each feature being half the resolution of the previous level. This increases the receptive field and helps to bet-

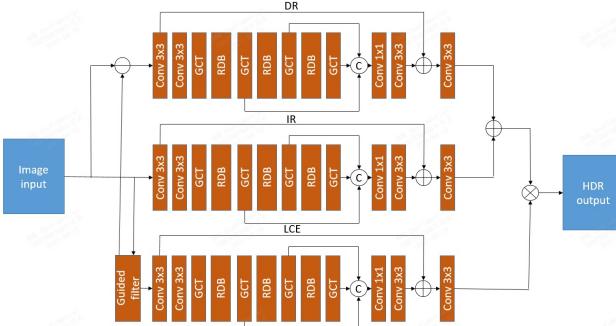


Figure 9. Architecture of Task-specific Network based on Channel Adaptive RDN, proposed by the BOE-IOT-AIBD team.

ter handle large foreground motion. At each level, the attention mechanism is used to identify which regions to use from the long and short exposure frames. The features at each level are merged independently first before being up-sampled back to the original resolution. The features from all three levels are then merged together using some fusion blocks to generate the final HDR image. The architecture can be seen in Figure 10.

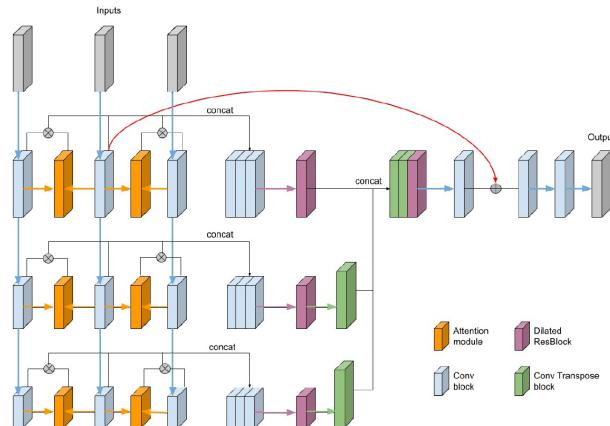


Figure 10. Architecture of Multi-Level Attention on Multi-Exposure Frames for HDR Reconstruction, proposed by the SuperArtifacts team.

#### 4.6. CET-CVLAB

# Single Image HDR Synthesis with Densely Connected

**Dilated ConvNet** The team propose an architecture which consists of a densely connected stack of dilated residual dense blocks (DRDBs) [1]. The dilation rate of convolutional layers used within the proposed DRDB progressively grows from 1 to 3 and then progressively decreases from 3 to 1. The DRDBs themselves are also connected as shown in Figure 11 to improve the representation capability of the network.

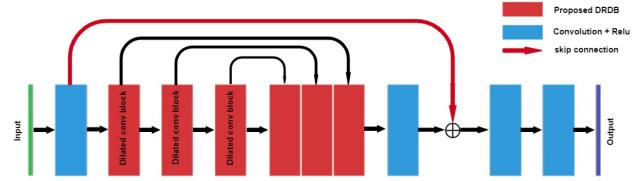


Figure 11. Architecture of Single Image HDR Synthesis with Densely Connected Dilated ConvNet, proposed by the CET-CVLAB team.

## 4.7. CVRG

**Deep Single-Shot LDR to HDR** The team propose a two stage method [31]: Stage I (inspired by [30]) performs denoising and recovers the 8-bit HDR image from the single LDR input; Stage II tonemaps the image into the linear domain and recovers the 16-bit HDR image. The architecture can be seen in Figure 12. The team proposes the Residual Dense Attention Block (RDAB) as the building block of the model. The RDAB, which combines the residual dense block and the spatial attention module, can be seen in Figure 13.

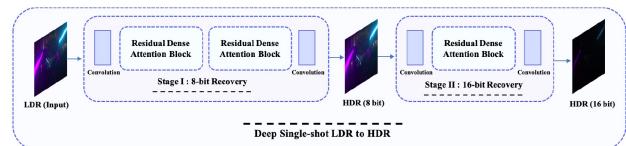


Figure 12. Architecture of Deep Single-Shot LDR to HDR, proposed by the CVRG team.

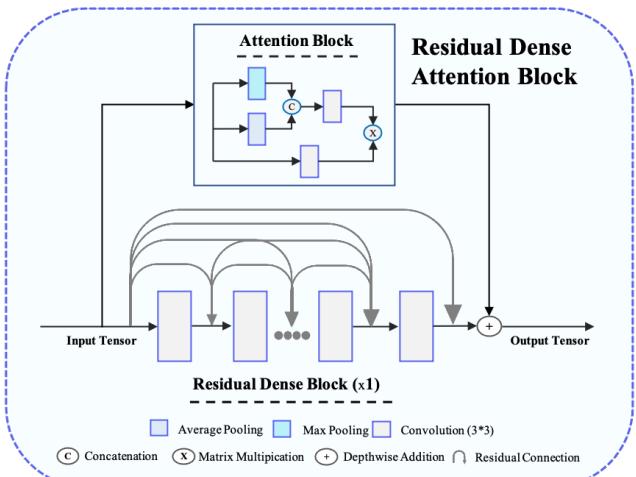


Figure 13. Residual Dense Attention Block, proposed by the CVRG team.

## 4.8. ZJU231

**Reference-Guided Multi-Exposure Fusion Network for HDR Imaging** The team propose a two-stage architecture which consists of the ghost reduction sub-network and the multi-exposure information fusion sub-network. Inspired by AHDRNet [39], the ghost reduction sub-network uses the reference image to generate an attention map for the short and long exposure images. The extracted features are guided via element-wise multiplication with the attention maps. The guided features are concatenated and merged by the fusion sub-network, which consists of five DRDBs followed by three convolutions, as shown in Figure 14.

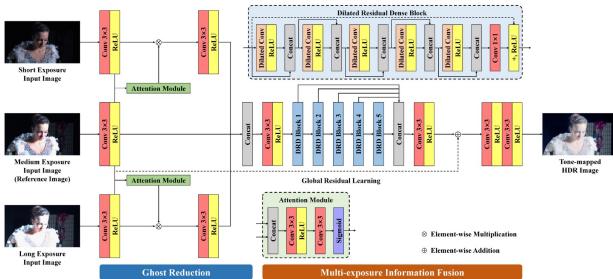


Figure 14. Architecture of Reference-Guided Multi-Exposure Fusion Network for HDR Imaging, proposed by the ZJU231 team.

## 4.9. Samsung Research Bangalore

**HDR Merging using Multi Branch Residual Networks** The team propose a multi-branch U-Net architecture inspired by [38] and [18] which consists of an encoder, a residual body and a decoder as seen in Figure 15. The building blocks of the network are Double Convolutional Residual Blocks (DCRB). This consists of two convolutions with prelu activations and the input is skipped to the output using a 1x1 convolution.

Each input image is processed with separate branches. The encoder consists of three blocks which successively downsample the input image. The features are then concatenated and processed using six residual blocks, followed by three decoder blocks which upsample the image back to full resolution. There are skip connections between all three of the encoder and decoder blocks. Self-ensembling strategy by averaging 8 ensembles created using flip and transpose operations are used to further improve the results.

## Acknowledgements

We thank the NTIRE 2021 sponsors: Huawei, Facebook Reality Labs, Wright Brothers Institute, MediaTek, OPPO and ETH Zurich (Computer Vision Lab).

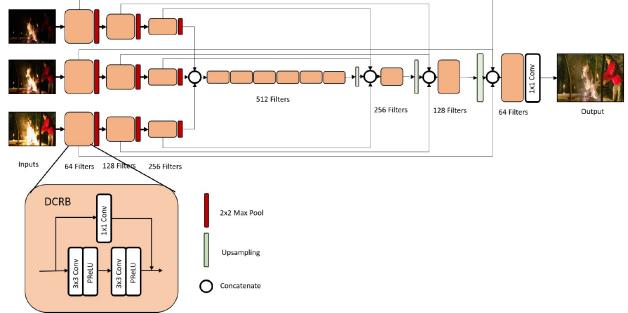


Figure 15. Architecture of HDR Merging using Multi Branch Residual Networks, proposed by the Samsung Research Bangalore team

## A. Teams and Affiliations

### NOAHTCV

**Title:** Single Image HDR Reconstruction in a Multi-stage Manner / Alignment Augmentation and Multi-Attention Guided HDR Fusion

**Members:** Xian Wang<sup>1</sup> ([wangxian10@huawei.com](mailto:wangxian10@huawei.com)), Yong Li<sup>1</sup>, Tao Wang<sup>1</sup> and Fenglong Song<sup>1</sup>

**Affiliations:** <sup>1</sup>Huawei Noah's Ark Lab

### MegHDR

**Title:** ADNet: Attention-guided Deformable Convolutional Networks for High Dynamic Range Imaging

**Members:** Zhen Liu<sup>1</sup> ([liuzhen03@megvii.com](mailto:liuzhen03@megvii.com)) Wenjie Lin<sup>1</sup>, Xinpeng Li<sup>1</sup>, Qing Rao<sup>1</sup>, Ting Jiang<sup>1</sup>, Mingyan Han<sup>1</sup>, Haoqiang Fan<sup>1</sup>, Jian Sun<sup>1</sup> and Shuaicheng Liu<sup>1</sup>

**Affiliations:** <sup>1</sup>Megvii Technology

### XPixel

**Title:** HDRUNet: Single Image HDR Reconstruction with Denoising and Dequantization

**Members:** Xiangyu Chen<sup>1</sup> ([chxy95@gmail.com](mailto:chxy95@gmail.com)), Yihao Liu<sup>1</sup>, Zhengwen Zhang<sup>1</sup>, Yu Qiao<sup>1</sup>, Chao Dong<sup>1</sup>

**Affiliations:** <sup>1</sup>Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences

### SuperArtifacts

**Title:** Multi-Level Attention on Multi-Exposure Frames for HDR Reconstruction

**Members:** Evelyn Yi Lyn Chee<sup>1</sup> ([evelyn.chee@bst.ai](mailto:evelyn.chee@bst.ai)), Shanlan Shen<sup>1</sup>, Yubo Duan<sup>1</sup>

**Affiliations:** <sup>1</sup>Black Sesame Technologies (Singapore)

### BOE-IOT-AIBD

**Title:** Task-specific Network based on Channel Adaptive RDN

**Members:** Guannan Chen<sup>1</sup> ([chenguannan@boe.com.cn](mailto:chenguannan@boe.com.cn)), Mengdi Sun<sup>1</sup>, Yan Gao<sup>1</sup>, Lijie Zhang<sup>1</sup>

**Affiliations:** <sup>1</sup>BOE Technology Co., Ltd.

## CET-CVLAB

**Title:** Single Image HDR Synthesis with Densely Connected Dilated ConvNet

**Members:** Akhil K A<sup>1</sup> ([akhil.93.ka@gmail.com](mailto:akhil.93.ka@gmail.com)), Jiji C V<sup>1</sup>

**Affiliations:** <sup>1</sup>College of Engineering Trivandrum, India

## ZJU231

**Title:** Reference-Guided Multi-Exposure Fusion Network for HDR Imaging

**Members:** Chenjie Xia<sup>1,2</sup> ([chenjiexia@zju.edu.cn](mailto:chenjiexia@zju.edu.cn)), Bowen Zhao<sup>1,2</sup> ([bowenzhao@zju.edu.cn](mailto:bowenzhao@zju.edu.cn)), Zhangyu Ye ([qiushizai@zju.edu.cn](mailto:qiushizai@zju.edu.cn)), Xiwen Lu, Yanpeng Cao<sup>1,2</sup>, Jiangxin Yang<sup>1,2</sup>, Yanlong Cao<sup>1,2</sup>

**Affiliations:** <sup>1</sup>State Key Laboratory of Fluid Power and Mechatronic Systems, Zhejiang University, <sup>2</sup>Key Laboratory of Advanced Manufacturing Technology, Zhejiang University

## CVRG

**Title:** Deep Single-Shot LDR to HDR

**Members:** S M A Sharif<sup>2</sup> ([sma.sharif.cse@ulab.edu.bd](mailto:sma.sharif.cse@ulab.edu.bd)), Rizwan Ali Naqvi<sup>1</sup>, Mithun Biswas, and Sungjun Kim

**Affiliations:** <sup>1</sup>Sejong University, South Korea, <sup>2</sup>Rigel-IT, Bangladesh

## Samsung Research Bangalore

**Title:** HDR Merging using Multi Branch Residual Networks

**Members:** Green Rosh K S<sup>1</sup> ([green-rosh.ks@samsung.com](mailto:green-rosh.ks@samsung.com)), Sachin Deepak Lomte<sup>1</sup>, Nikhil Krishnan<sup>1</sup>, B H Pawan Prasad<sup>1</sup>

**Affiliations:** <sup>1</sup>Samsung R&D Institute India Bangalore (SRI-B)

## References

- [1] Akhil K A and Jiji C V. Single image hdr synthesis using a densely connected dilated convnet. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, 2021.
- [2] Abdullah Abuolaim, Radu Timofte, Michael S Brown, et al. NTIRE 2021 challenge for defocus deblurring using dual-pixel images: Methods and results. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, 2021.
- [3] Codruta O Ancuti, Cosmin Ancuti, Florin-Alexandru Vasluiianu, Radu Timofte, et al. NTIRE 2021 nonhomogeneous dehazing challenge report. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, 2021.
- [4] Goutam Bhat, Martin Danelljan, Radu Timofte, et al. NTIRE 2021 challenge on burst super-resolution: Methods and results. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, 2021.
- [5] Guannan Chen, Lijie Zhang, Mengdi Sun, Yan Gao, Pablo Navarrete Michelini, and Yanhong Wu. Single-image hdr reconstruction with task-specific network based on channel adaptive RDN. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, 2021.
- [6] 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.
- [7] J. Dai, H. Qi, Y. Xiong, Y. Li, G. Zhang, H. Hu, and Y. Wei. Deformable convolutional networks. In *2017 IEEE International Conference on Computer Vision*, 2017.
- [8] P. E.Debevec and Jitendra M. Recovering high dynamic range radiance maps from photographs. In *ACM SIGGRAPH 2008 Classes*, 2008.
- [9] Gabriel Eilertsen, Joel Kronander, Gyorgy Denes, Rafał 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.
- [10] Majed El Helou, Ruofan Zhou, Sabine Süsstrunk, Radu Timofte, et al. NTIRE 2021 depth guided image relighting challenge. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, 2021.
- [11] J. Froehlich, S. Grandinetti, B. Eberhardt, S. Walter, A. Schilling, and H. Brendel. Creating cinematic wide gamut hdr-video for the evaluation of tone mapping operators and hdr-displays. In *Proc. of SPIE Electronic Imaging*, 2014.
- [12] Jinjin Gu, Haoming Cai, Chao Dong, Jimmy S. Ren, Yu Qiao, Shuhang Gu, Radu Timofte, et al. NTIRE 2021 challenge on perceptual image quality assessment. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, 2021.
- [13] S. W. Hasinoff, F. Durand, and W. T. Freeman. Noise-optimal capture for high dynamic range photography. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2010.
- [14] Qibin Hou, Daquan Zhou, and Jiashi Feng. Coordinate attention for efficient mobile network design. In *CVPR*, 2021.
- [15] 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.
- [16] Nima Khademi Kalantari and Ravi Ramamoorthi. Deep HDR Video from Sequences with Alternating Exposures. *Computer Graphics Forum*, 2019.
- [17] Soo Kim, Jihyong Oh, and Munchurl Kim. Jsi-gan: Gan-based joint super-resolution and inverse tone-mapping with pixel-wise task-specific filters for uhd hdr video. *Proceedings of the AAAI Conference on Artificial Intelligence*, 34:11287–11295, 04 2020.
- [18] G. R. K.S., A. Biswas, M. S. Patel, and B. H. P. Prasad. Deep multi-stage learning for hdr with large object motions. In *2019 IEEE International Conference on Image Processing (ICIP)*, pages 4714–4718, 2019.
- [19] 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.
- [20] Yu-Lun Liu, Wei-Sheng Lai, Yu-Sheng Chen, Yi-Lung Kao, Ming-Hsuan Yang, Yung-Yu Chuang, and Jia-Bin Huang. Single-image HDR reconstruction by learning to reverse the camera pipeline, 2020.
- [21] 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 *IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, 2021.
- [22] Andreas Lugmayr, Martin Danelljan, Radu Timofte, et al. NTIRE 2021 learning the super-resolution space challenge. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, 2021.
- [23] M. McGuire, W. Matusik, H. Pfister, B. Chen, J. F. Hughes, and S. K. Nayar. Optical splitting trees for high-precision monocular imaging. *IEEE Computer Graphics and Applications*, 27(2), 2007.
- [24] T. Mertens, J. Kautz, and F. Van Reeth. Exposure fusion. In *Pacific Conference on Computer Graphics and Applications*, 2007.
- [25] Seungjun Nah, Sanghyun Son, Suyoung Lee, Radu Timofte, Kyoung Mu Lee, et al. NTIRE 2021 challenge on image deblurring. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, 2021.
- [26] Eduardo Pérez-Pellitero, Sibi Catley-Chandar, Aleš 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.
- [27] 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.
- [28] Marcel Santana Santos, Tsang Ing Ren, and Nima Khademi Kalantari. Single image HDR reconstruction using a cnn with masked features and perceptual loss. *ACM Transactions on Graphics*, 39(4), Jul 2020.
- [29] Pradeep Sen, Nima Khademi Kalantari, Maziar Yasoubi, Soheil Darabi, Dan B Goldman, and Eli Shechtman. Robust patch-based hdr reconstruction of dynamic scenes. *ACM Trans. Graph.*, 31(6):203–1, 2012.
- [30] S M A Sharif, Rizwan Ali Naqvi, and Mithun Biswas. Learning medical image denoising with deep dynamic residual attention network. *Mathematics*, 8(12), 2020.
- [31] S M A Sharif, Rizwan Ali Naqvi, Mithun Biswas, and Sungjun Kim. A two-stage deep network for high dynamic range image reconstruction. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, 2021.
- [32] Sanghyun Son, Suyoung Lee, Seungjun Nah, Radu Timofte, Kyoung Mu Lee, et al. NTIRE 2021 challenge on video super-resolution. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, 2021.
- [33] D. Sun, X. Yang, M. Liu, and J. Kautz. Pwc-net: Cnns for optical flow using pyramid, warping, and cost volume. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2018.
- [34] Radu Timofte, Rasmus Rothe, and Luc Van Gool. Seven ways to improve example-based single image super resolution. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 1865–1873, 2016.
- [35] M. D. Tocci, C. Kiser, N. Tocci, and P. Sen. A versatile hdr video production system. In *SIGGRAPH*, 2011.
- [36] 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.
- [37] 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.
- [38] 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.
- [39] Qingsen Yan, Dong Gong, Qinfeng Shi, Anton van den Hengel, Chunhua Shen, Ian D. Reid, and Yanning Zhang. Attention-guided network for ghost-free high dynamic range imaging. In *Computer Vision and Pattern Recognition*, 2019.
- [40] Ren Yang, Radu Timofte, et al. NTIRE 2021 challenge on quality enhancement of compressed video: Methods and results. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, 2021.
- [41] Zongxin Yang, Linchao Zhu, Yu Wu, and Yi Yang. Gated channel transformation for visual recognition. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2020.
- [42] Syed Waqas Zamir, Aditya Arora, Salman Khan, Munawar Hayat, Fahad Shahbaz Khan, Ming-Hsuan Yang, and Ling Shao. Learning enriched features for real image restoration and enhancement. In *ECCV*, 2020.
- [43] Yulun Zhang, Yapeng Tian, Yu Kong, Bineng Zhong, and Yun Fu. Residual dense network for image super-resolution. In *Computer Vision and Pattern Recognition*, 2018.