Perceptual Quality Assessment of High-Dynamic-Range Image: A Benchmark Dataset and a No Reference Method

Zihao Zhou, Liquan Shen, Jun Lei, Zhaoyi Tian, Xiangyu Hu, Shiwei Wang, and Yang Chen

***Abstract*— High dynamic range (HDR) imaging technology has received increasing attention in recent years, and HDR image quality assessment (IQA) metrics are indispensable during the capturing, processing and displaying of HDR images. However, existing HDR-IQA datasets and methods neglect complex distortions during the HDR image processing schemes, leading to limited generalization performance on practical application. In this work, to facilitate the development of HDR-IQA dataset, we present HDRQAD, a large-scale HDR Quality Assessment Dataset, which possesses diversified distortions during HDR imaging technologies, abundant scenes and considerable quantity. Specifically, the HDRQAD dataset contains 1409 HDR images, which are derived from source scenes with six types of distortions during the HDR imaging schemes. In contrast to existing datasets that contain only compression artifacts, the HDRQAD includes Under-exposure, Over-exposure, Motion blur and Ghosting in HDR images achieved with multi-exposure fusion technology, conversion artifacts in HDR images achieved with single image reconstruction technology and compression artifacts during the transmission of HDR images. Furthermore, during the process of constructing the dataset, we identified three key challenges in HDR-IQA tasks: 1) dynamic range variations, 2) HDR visual artifacts with large overall gap, 3) inter-regional non-uniform image quality. Based on these observations, we propose a new end-to-end network for HDR-IQA tasks, which consists of a Distortion-aware Representation Learning (DRL) module and an Inter-Regional Quality Interaction (IRQI) module. The DRL learns the representations of dynamic range variations and HDR visual artifacts, enhancing the reliability of prior information extraction. The IRQI captures inter-regional quality dependencies with interacting and fusing intermediate distortion features for more accurately predicting image quality. Extensive experiments prove the superiority of proposed HDRQAD and demonstrate that the proposed network achieves state-of-the-art performance. The Dataset and Code will be made publicly available at HDR-IQA-Dataset.**

***Index Terms*— High dynamic range (HDR), deep learning, no reference, image quality assessment.**

I. Introduction

HIGH dynamic range (HDR) imaging technology is favored by the consumer market and professional color systems to render the brightness and color variations of real scene more accurately. During the past decade, advancements in HDR imaging technology and hardware devices have revolutionized the whole multimedia communications pipeline from acquisition to final display [1]. Recently, many HDR imaging methods were proposed, which can be classified into three types: Multi-exposure Fusion (MEF) frameworks [2], [3], Single Image Reconstruction (SIR) frameworks [4], [5] and HDR camera schemes [6], [7].

These imaging methods facilitate the acquisition of HDR content. However, as shown in Fig. 1, these methods may lead to different distortion patterns. In the context of MEF, ghosting artifacts that arise during the fusion process is a critical content of research [8]. For lack of shooting proficiency, issues such as motion blur, under-exposure, and over-exposure may occur during the fusion process. For SIR, the reconstructed HDR image suffers from saturation in highlights, noise in lowlights and severe color shift [9]. These distortions are referred to as single image construction artifacts. Although HDR cameras can directly capture HDR content, compression artifacts inevitably arise during storage and transmission.

Research in HDR IQA is significantly lagging. One main limitation is the absence of a uniform subjective quality assessment dataset with comprehensive distorted content. Pioneers have constructed some previous HDR-IQA datasets [10], [11], [12], [13], [14] that contain distorted images in various scenarios. They just utilize existing compression tools, such as JPEG, JPEG2K, JPEG-XT and HEVC to construct a dataset, resulting in limited distortion content. Besides, most of these datasets are limited in scale, featuring only several hundred images or even fewer than one hundred, and lack scene diversity, making it difficult to validate the effectiveness and robustness of the HDR-IQA algorithms.

Besides datasets, another limitation is the lack of efficient IQA algorithms specifically designed for HDR content. The

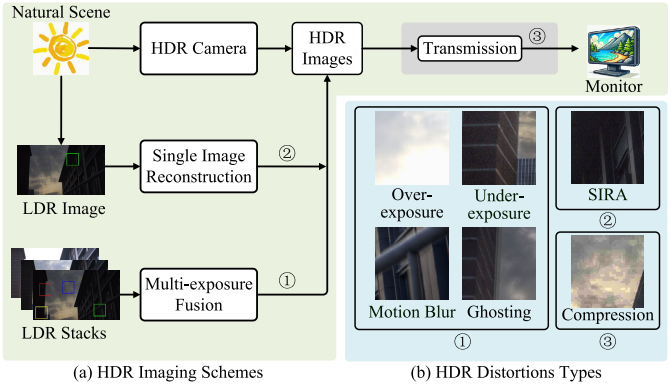
Received 17 October 2024; revised 28 December 2024, 15 January 2025, and 1 February 2025; accepted 8 February 2025. Date of publication 13 February 2025; date of current version 4 July 2025. This work was supported in part by the National Natural Science Foundation of China under Grant 61931022 and Grant 62271301, in part by the Shanghai Science and Technology Program under Grant 22511105200, in part by the Shanghai Excellent Academic Leaders Program under Grant 23XD1401400, and in part by the Natural Science Foundation of Shandong Province under Grant ZR2022ZD38. This article was recommended by Associate Editor W. Liu. *(Corresponding author: Liquan Shen.)*

Zihao Zhou, Zhaoyi Tian, Xiangyu Hu, Shiwei Wang, and Yang Chen are with the School of Communication and Information Engineering, Shanghai University, Shanghai 200444, China (e-mail: yi_yuan@shu.edu.cn; kinda@shu.edu.cn; arhu314@shu.edu.cn; ieemia@shu.edu.cn; rapidkoe@163.com).

Liquan Shen is with the Key Laboratory of Specialty Fiber Optics and Optical Access Networks and the Joint International Research Laboratory of Specialty Fiber Optics and Advanced Communication, Shanghai University, Shanghai 200444, China (e-mail: jsslq@163.com).

Jun Lei is with the School of Communication and Information Engineering, Shanghai University, Shanghai 200444, China, and also with China Mobile (Hangzhou) Information Technology Company Ltd., Hangzhou 310030, China (e-mail: leijun@cmhi.chinamobile.com).

Digital Object Identifier 10.1109/TCSVT.2025.3541772



(a) HDR Imaging Schemes (b) HDR Distortions Types

Fig. 1. Different HDR imaging schemes and various HDR distortions. The distortion types of proposed dataset consists of ①②③, while existing dataset only contains ③. SIRA is single image reconstruction artifacts.

IQA algorithms employed for HDR images can be divided into three categories: **1) general algorithms** that aim to extract comprehensive image distortion information through the use of either handcrafted [15] or learned features [16]; **2) extended LDR algorithms** [17], [18], [19], which transform linear content into a perceptually more uniform space for subsequent prediction; **3) HDR-IQA algorithms** [20], [21], [22], [23], [24], [25], which are designed to simulate the non-linear response characteristics of the human visual system. General algorithms and extended LDR metrics, which were initially intended for LDR image distortions, may not be as effective in evaluating HDR images, potentially leading to performance degradation.

In HDR-IQA algorithms, most traditional methods [20], [21], [22], [23] depend on simple mapping functions that convert image features into quality prediction. And most deep learning-based methods [24], [25] rely on general features, such as saliency and visual masking, which constrains their ability to accurately evaluate HDR-specific distortions, resulting in suboptimal performance. Although these methods achieved their intended goals, they were validated only on previous datasets only containing compression distortion. However, distortions from recent HDR imaging methods cannot be ignored, including Motion blur, Ghosting, Under-exposure and Over-exposure from MEF scheme, and conversion artifacts from SIR scheme. Existing HDR-IQA algorithms are not effective at assessing these distortions. Therefore, it is necessary to propose a new HDR-IQA algorithm to better predict the quality of HDR images with diverse HDR distortions.

HDR images display a diversity of distortions, each of which has unique distortion characteristics. These factors motivate us to consider various distortions from alternative perspectives in the design of an HDR-IQA algorithm. **How to represent the HDR distortions?** There are two key aspects: dynamic range distortions and HDR visual artifacts. Dynamic range variations etc. can lead to deviation of human cognition, while HDR visual artifacts resulting from different distortions have large overall gap, such as ghosting and over-exposure, induce discomfort during observation, resulting in serious degradaion of subjective quality. **How to address the problem of inter-regional non-uniform quality in single HDR image?** The HDR distortions introduce non-uniform quality degradation in single HDR image, complicating quality assessment. As shown in Fig. 2, image with over-exposure exhibits severe detail loss in highlight regions while low-light areas maintain a considerable degree of integrity, which demonstrates that both low-quality and high-quality regions coexist in an HDR image. This necessitates a holistic representation of image quality, encompassing quality variations across different regions and considering inter-regional quality dependencies.

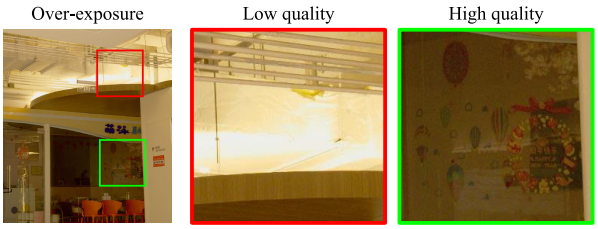


Fig. 2. Problem of non-uniform quality. The HDR image is degradate by over-exposure, where red is low quality region and green is high quality region.

To address the limitation of datasets, motivated by the inherent differences, we here propose a groundbreaking HDRQAD, High Dynamic Range image Quality Assessment Dataset, for HDR-IQA tasks. Compared with previous works [10], [11], [12], [13], [14] that only consider HDR carmera scheme, HDRQAD is the first large-scale IQA dataset that not only includes HDR camera scheme but also involves multi-exposure fusion and single image reconstruction schemes, addressing the research vacancy in the HDR subjective quality assessment. It encompasses a wide range of content and HDR distortions to cover the complete quality spectrum of HDR images created in practice. In particular, HDRQAD includes a total of 1,409 distorted HDR images with more than 20,000 human ratings. The dataset features a high diversity of real content (outdoor, indoor, daytime, nighttime, backlight, frontlight, etc.) and distortion types (over-exposure, under-exposure, single image reconstruction artifact, ghosting, motion blur, etc.), covering a wide spectrum of distortions encountered in most real-world applications.

To address the limitation of HDR-IQA algorithms, we propose a network to assess HDR image quality more accurately and reasonably. The network comprises a Distortion-aware Representation Learning module (DRL) and an Inter-Regional Quality Interaction module (IRQI). The DRL is designed to learn the representation of HDR distortions, which independently extracts prior information of distortions presenting in distinct regions of a HDR image, improving the robustness of subsequent efforts to capture inter-regional dependencies. The IRQI captures inter-regional quality dependencies with interacting and fusing intermediate distortion features to more accurately representing image quality. The contributions of our work can be summarized as follows:

- We embark on a groundbreaking endeavor to create the HDRQAD for developing a high-quality metrics for HDR IQA. A total of 1,409 distorted HDR images acquired through three imaging schemes, enabling HDRQAD

characterized by a high diversity of real distortion types (6 categories) and content (147 scenes). To our best knowledge, HDRQAD is the largest in scale, the most abundant in scenes, and contains comprehensive distortions.

- We design an end-to-end network for HDR IQA, which consists of a DRL module and an IRQI module. The DRL module independently considers distortions in different regions from both dynamic range and HDR visual artifacts perspectives, enhancing the robustness of quality prediction. The IRQI captures inter-regional quality dependencies to improve the accuracy of overall quality assessment.
- Extensive expreiments demonstrate the superiority of HDRQAD and proposed HDR-IQA algorithm. We re-evaluated the existing popular IQA algorithms on HDRQAD, providing a complete survey of performance of them for researchers to explore HDR IQA researches.

II. Related Works

A. Subjective HDR Quality Assessment

With the advancement of HDR research, many subjective HDR-IQA datasets have been proposed to explore the perceptual quality of HDR image. For instance, Narwaria et al. [10] studied the quality assessment in tone mapping-based HDR image compression, where they explored the optimal parameters of dynamic range reduction function for maximized visual quality and constructed a subjective dataset consisting of 140 compressed HDR images with 10 contents. The next year, Narwaria et al. [14] addressed the issue of how tone mapping affects the perceptual quality of the decompressed HDR signal produced by inverse tone mapping, and collected 216 decompressed HDR images with 6 contents. Valenzise et al. [11] proposed a subjective dataset containing 260 compressed images with 5 contents to validate the consistency of PSNR and SSIM with subjective perception in a perceptually uniform space. Korshunov et al. [12] proposed a dataset with JPEG-XT, which led to a total of 240 compressed images with 20 contents, including several images with different luminance, frames from HDR video, and CGI images. Rousselot et al. [26] selected 96 distorted images with 8 HDR images degraded by HEVC, gaussian noise and color gamut mismatch to study the impacts of viewing conditions on HDR-VDP2.

To facilitate development of HDR IQA, there are serveral works that have merged previous datasets. Zerman et al. [13] provided a complete and thorough survey of the performance of HDR full-reference image quality metrics. They collected several HDR image databases [10], [11], [12], [14] and created a new part of HDR images, which consists of 50 distorted images with 5 contents degraded by JPEG and JPEG2K, resulting in a total of 690 distorted images. Recently, Mikhailiuk et al. [27] constructed a consolidated dataset (UPIQ) consists of 3779 LDR images and 380 HDR images using psychometric scales, where the Mean Opinion Score (MOS) values of HDR images were obtained from two datasets [10], [12] by a specially designed algorithm.

In addition, subjective experiments have been conducted using several HDR datasets. Fang et al. [28] constructed a deghosting quality assessment dataset to investigate the effectiveness of different deghosting algorithms. Hanji et al. [9] proposed a HDR single reconstruction dataset to rank the results of single image reconstruction methods. It is regretful that these datasets lack MOS values, resulting in their unavailability for HDR IQA.

Most of the above datasets contain only compression distortion and are limited in scale. These datasets focus on compression distortion, reflecting a relatively mature study of its subjective quality, while other common HDR distortions, such as ghosting and single-image reconstruction artifacts, remain insufficiently explored. Although there are datasets [9], [28] containing the specific distorted content of HDR imaging, they lack the MOS necessary for HDR-IQA tasks. Therefore, a large-scale dataset with a wide range of distortions and diverse content is highly desired.

B. Objective Quality Assessment

1) General IQA Algorithms: Objective IQA algorithms are designed to predict image quality consistent with human perception. According to the types of reference image, IQA algorithms can be divided into three types: Full Reference (FR), Reduced Reference (RR) and No Reference (NR). FR algorithms are typically achieved by comparing the pristine reference image against a distorted version to quantify the visual quality degradation [29], [30]. RR IQA models utilize partial reference information to assess visual quality [31], [32]. NR methods evaluate visual quality without a reference image and instead employ information extracted from distorted images [33], [34], making them the most practical for real-world applications. Before the advent of deep neural network methods, traditional IQA approaches were predominantly full-reference and relied on handcrafted features to represent image quality. The popular traditional IQA models include PSNR, SSIM [35], VIF [36], FSIM [37], GFM [38], NIQE [39], PIQE [40] and Brisque [15], among others. Different from traditional handcrafted-based methods, deep learning-based IQA approaches are mostly no-reference and can establish an end-to-end mapping between the image and its quality. The popular deep learning-based IQA approaches contain HyperNet [16], VCRNet [41], TReS [42], TempQT [43], among others.

2) Extended LDR Metrics: Extended LDR metrics aim to better inherit the performance of LDR quality metrics with a special coding method. Considering the large gap between HDR linear content and human perception, Aydın et al. [17] designed a Perceptual Uniform (PU) code, which maps HDR linear content to perceptually uniformed range, to employ PSNR and SSIM for HDR image prediction. Azimi et al. [19] improved the PU function relied on the latest CSF [45] and extended it to cover more LDR metrics. Shang et al. [46] proposed a new framework with a nonlinear transform to enhance distortions occurring in higher and lower light portions of the HDR image, improving the performance of LDR metrics in HDR-IQA tasks. Recently, Cao et al. [47]

decomposed HDR images into LDR image stacks with different exposures, and then evaluated these decomposed images with the well-established LDR quality metrics to predict HDR image quality. Besides, although the Gamma function and the PQ [18] transform are designed for display model, they have the ability to extend LDR metrics.

3) Dedicated HDR IQA Algorithms: HDR IQA remains a nascent field. Most existing HDR-IQA metrics are traditional methods, even learning-based methods are underdeveloped. For traditional methods, HDR visual difference predictor (HDR-VDP) [20] is one of the most classical HDR FR-IQA methods. The series of HDR-VDP predicts the visible differences by modeling the optical and retinal pathways in the human visual system (HVS) [20], [21], [48]. HDR-VQM is another equally important metric, which extracts frequency features with a Gabor filter in a perceptual domain for predicting the video quality [49]. Some authors have utilized structural information between the reference image and a distorted version, Liu et al. [23] employed Gabor and Butterworth filter to model response of HVS in frequency, and Zhang et al. [22] utilized gradient similarity with the designed stabilization parameters. Moreover, other researchers have addressed HDR image quality from the assessment chromatic aspect, employing different HDR Uniform Color Spaces [50] or utilizing color difference models [51].

For deep learning-based models, most of them were designed for compression distortion and have not undergone subsequent development. To the best of our knowledge, Jia et al. [24] first proposed a deep learning-based model with saliency map and verified the feasibility of converting LDR image features to HDR images. The NR method proposed by Kottayil et al. [25] predicts pixel-level error and perceptual resistance to image error with two deep network units. Mikhailiuk et al. [27] trained the PU-PieAPP model based on the construction of a uniform photometric image quality (UPIQ) dataset. Besides, some authors attempted other methods, such as modeling HDR-VDP2 prediction [52] and using newly-conceived differential natural scene statistics [53].

III. Proposed Dataset

To advance the development of HDR IQA, the largest High Dynamic Range image Quality Assessment Dataset (HDRQAD) is proposed. Most of the existing HDR-IQA datasets [10], [11], [12], [13], [14], [26], [27] focus on at exploring the effect of transmission on image quality. However, the diversity of HDR imaging schemes, which mainly is divided into multi-exposure fusion (MEF) [2], [3], single image reconstruction (SIR) [4], [5] and HDR Camera scheme [6], [7], introduces HDR distortions not limited to transmission. In this section, we focus on distortions that may occur in HDR imaging schemes, and provides essential guidance for constructing our dataset.

A. HDR Image Distortion Types

Three imaging schemes mentioned in Sec. I exhibit different distortion patterns. In the context of multi-exposure fusion (MEF), ghosting artifacts have been a significant focus of research [8]. Due to a lack of shooting proficiency, issues such as motion blur, under-exposure, and over-exposure may occur during the capture process. SIR scheme often produces HDR images exhibiting saturation in highlights, noise in lowlights, and significant color shifts [9]. These distortions are referred to as single image construction artifacts. Although HDR cameras can directly capture HDR content, compression artifacts are inevitably introduced during transmission. Moreover, the HDR image distortions include *ghosting*, *motion blur (MB)*, *under-exposure (UE)* and *over-exposure (OE)* introduced by MEF scheme, *single image reconstruction artifact (SIRA)* occurring in SIR scheme, and *compression artifacts* result from HDR Camera scheme.

B. Building Dataset

A multi-exposure fusion scheme, single image reconstruction scheme, and HDR camera scheme are involved in the proposed dataset, with details as follows:

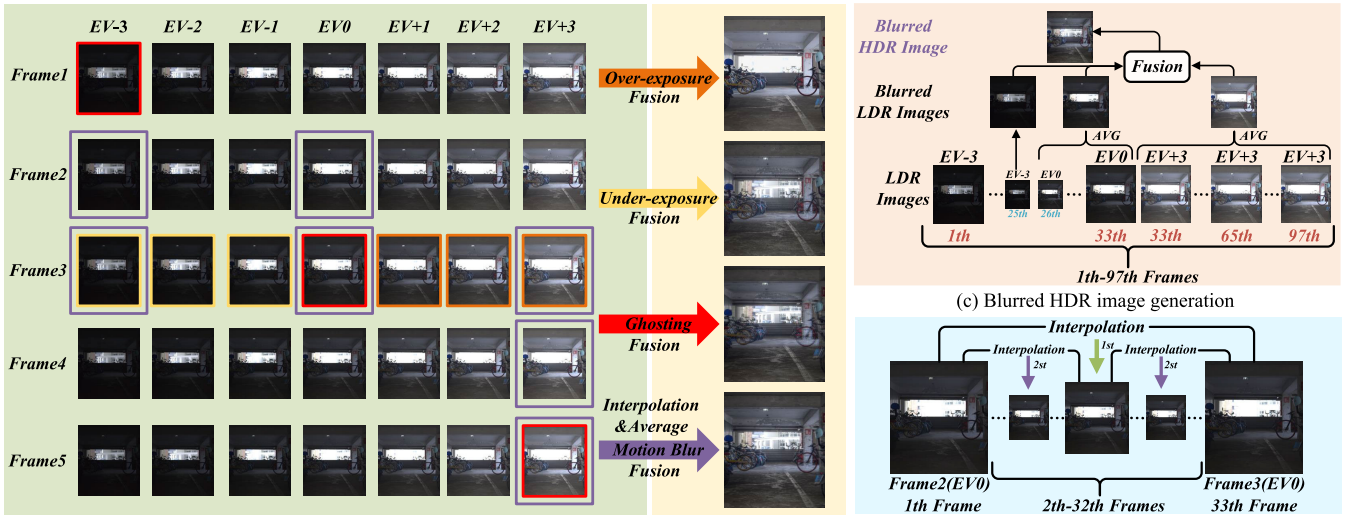
1) MEF Scheme: The LDR image stacks are captured by high-speed mirrorless camera, the Canon R6 Mark 2, which enables us to acquire high-quality multi-exposure LDR image stacks. As shown in Fig. 3, five multi-exposure LDR image stacks are captured in a manually motion-continuous way. Three images with exposure level *{EV-3, EV0, EV+3}* of the middle frame in LDR stacks are selected to obtain the reference HDR image by triangular fusion function [54]. As shown in Fig. 3, the distortions generation as follows:

- *Over-exposure (OE)* : Selecting images with exposure values *{EV+1, EV+2, EV+3}* denotes degradation level 2, and those with *{EV0, EV+2, EV+3}* denotes level 1.
- *Under-exposure (UE)* : Selecting images with exposure values *{EV-3, EV-2, EV-1}* denotes degradation level 2, and those with *{EV-2, EV-1, EV-0}* denotes level 1.
- *Ghosting* : Multi-exposure images from neighboring frames are fused, with one- and two-frame intervals serving as degradation levels 1 and 2, respectively, to obtain ghost images.
- *Motion blur (MB)* : Each LDR image stack is captured statically, making it hard to directly obtain a blurred image. Inspired by [55], a strategy of interpolating is adapted, averaging and then fusing to obtain blurred HDR images, as shown in Fig. 3(c)(d).

2) SIR Scheme: As shown in Fig. 4, seven-exposure levels LDR image stacks are acquired through the Canon R6 Mark 2 pre-set bracketing program. The reference image generation method is same as MEF scheme. In addition to the scenes we captured, another 41 scenes are carefully selected from [56]. The distortions generation details as follows:

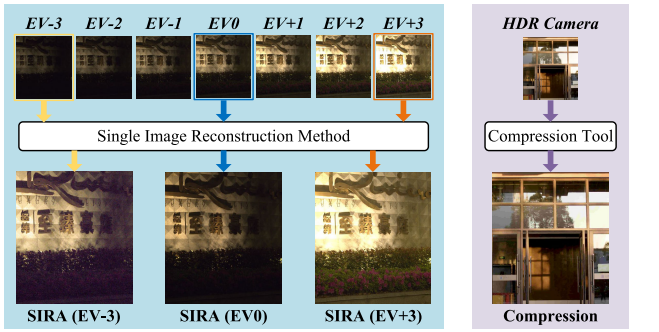
- *Single image reconstruction artifact (SIRA)* : HDRCNN [57] and ExpandNet [58] are selected for each scene as the algorithm to reconstruct the HDR content. Three LDR images with exposure values *{EV-3}*, *{EV0}* and *{EV+3}* are reconstructed to three HDR images.

3) HDR Camera Scheme: Digital film camera, the Blackmagic URSA Mini Pro 4.6K, is used to directly capture HDR images for HDR Camera scheme. The HDR image captured with digital film camera regarded as high quality reference HDR image. The distortions generation as follows:



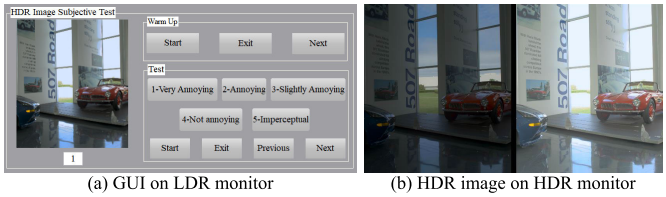
(a) LDR images of MEF scheme (b) Distortion samples (c) Blurred HDR image generation (d) Interpolation process

Fig. 3. (a) LDR image stacks for Multi-exposure Fusion scheme. (b) Over-exposure, under-exposure, ghosting and motion blur distortion samples. (c) Blurred HDR image generation method; in the LDR images row, the larger images are the original images from the LDR stacks, while the smaller images are interpolated frames. It is assumed that adjacent frames are temporally continuous and evenly spaced, achieved through pairwise interpolation. To generate blurred LDR images with different exposure durations and continuous motion, averaging is performed on 1, 8, and 64 consecutive LDR images with different exposures to create blurred LDR images. Finally, these images are fused to obtain the blurred HDR image. (d) is interpolation process that is implemented in pairs and performed 5 times to obtain 31 interpolated frames.



(a) SIR scheme (b) HDR camera scheme

Fig. 4. (a) LDR image stacks of Single Image Reconstruction scheme, and its distortion samples. (b) HDR camera scheme and its distortion sample.



(a) GUI on LDR monitor (b) HDR image on HDR monitor

Fig. 5. (a) The graphical user interface (GUI) on LDR monitor. (b) HDR image on HDR monitor. The HDR monitor shows the reference image on one side and the test image on the other, separated by a 32-pixel-wide black bar in the middle.

• *Compression* : Three degradations are generated with different JPEG-XT configurations, i.e., *ProfileA* with quality factor pairs *[20, 20]*, *ProfileB* with quality factor pairs *[35, 35]*, and *ProfileC* with quality factor pairs *[50, 50]*.

C. Subjective Test Protocol

Subjective test experiments are carried out in a gray-tone laboratory uninterrupted by external light sources and other factors, which is recommended by ITU-R [59]. A *SIM2 HDR* monitor, the full *HD SIM2 HDR47ES4MB* (1920 × 1080, *16:9)*, is used to display the test images. The Dual Stimulation Impairment Scale (DSIS) Variant I method [59] is employed for this test. The test image and the reference image are segmented to 944 × 1080, then pieced together into a image with 32 black pixels in the center. Twenty participants are asked to subjectively rate the test images on a scale ranging from *1* to *5* (*i.e. 1: very annoying, 2: annoying, 3: slightly annoying, 4: perceptible, but not annoying, 5: imperceptible*).

The computer used in the subjective testing experiments is connected to both LDR and HDR monitors. The Graphical User Interface (GUI) and displayed image as shown in Fig. 5. To mitigate the influence of image order, participants are divided into two groups with different reference image positions. To ensure the accuracy of participants' evaluations, a 10-minute warm-up session is conducted prior to the start of the test. During the testing process, participants know the position of the reference image and provide subjective quality assessments within the 6-second display of the HDR images. Each testing session has a duration of 10 minutes, followed by a 15-minute rest period before proceeding to the subsequent round of testing.

In total, 28,180 opinion scores of 1,409 images are collected. Two outliers are detected according to the principles recommended in ITU-R [59]. Each mean opinion score (MOS) is subsequently obtained by averaging the effective opinion scores. Assuming that the scores follow Student's *t*-distribution, each MOS is associated with the 95% confidence interval.

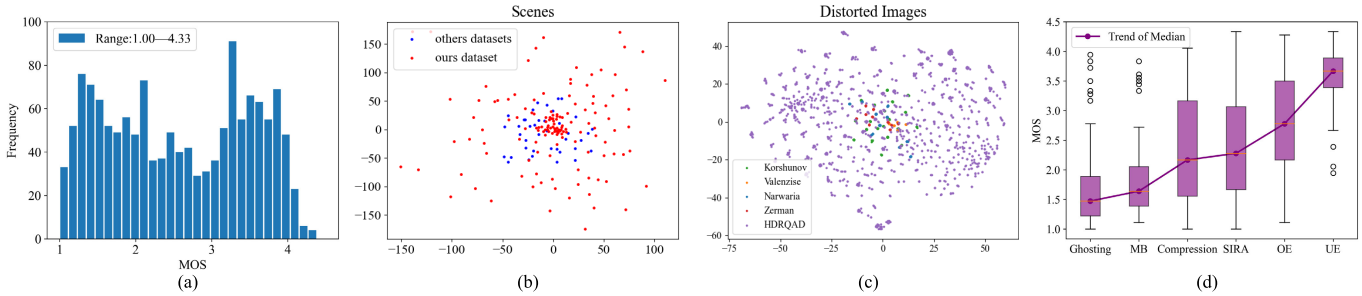


Fig. 6. (a) Histogram of MOS values in HDRQAD. (b) Scene diversity comparison. (c) Distorted content diversity comparison. Luminance map of original and distorted images are reduced to 2 dimensions using T-SNE [44]. (d) Distribution of MOS values for each type of distortion.

Fig. 7. (a) Abnormal dynamic range. The Reference denotes the dynamic range distribution of pristine reference image, while the SIRA or the Over-exposure denotes the dynamic range distribution of corresponding distorted image. (b) Visual artifacts with large overall gap. (c) Non-uniform quality in HDR images.

IV. Dataset And Distortion Analysis

A. Dataset Analysis

As shown in Fig. 6(a), it can be observed that the MOS values of HDRQAD almost span the entire quality axis, which reveals that proposed HDRQAD exemplifies reasonable perceptual quality of various distortions separation recognizability. To quantify its superiority, the proposed HDRQAD is compared with existing HDR-IQA datasets. As presented in Table I, HDRQAD stands out for its diversity of scenes, quantity, and types of distortion. As shown in Fig. 6(b) and Fig. 6(c), HDRQAD exhibits superior diversity and extensive coverage across both image content and distortion types. The MOS distributions of various distortions are shown in Fig. 6(d). The median MOS values show a clear upward trend from Ghosting to UE, which reveals that the influence of these distortions on perceived image quality is undeniable and cannot be overlooked.

TABLE I

Comparison of Existing HDR Image Quality Databases. The Distortion Types of HDRQAD Consist of Over-Exposure (OE), Under-Exposure (UE), Motion Blur (MB), Single Image Reconstruction Artifact (SIRA), Ghosting and JPEG-XT.' N.A.' Indicates Not Available for HDR-IQA Tasks

Dataset	Scene	Images	Distortion Type	Annotations
Narwaria [10]	10	240	JPEG	MOS
Narwaria [14]	6	216	JPEG2K	MOS
Valenzise [11]	5	260	JPEG, JPEG2K, JPEG-XT	MOS
Korshunov [12]	20	240	JPEG-XT	MOS
Zerman [13]	48	690	JPEG, JPEG2K, JPEG-XT	MOS
Rousselo [26]	8	96	HEVC, Gaussian, CGM	MOS
UPIQ [27]	30	380	JPEG, JPEG2K, JPEG-XT	JOD
Fang [28]	20	180	Ghosting	N.A.
Han [9]	27	432	SIRA	N.A.
HDRQAD	147	1409	OE, UE, MB, SIRA Ghosting, JPEG-XT	MOS

B. Distortion Analysis

In Fig. 7, the effect of various distortions regarding the MOS values and image characteristics is illustrated, where HDR image distortion can be analyzed from two aspects. **From the distortion representation aspect**, these distortions mainly manifest as dynamic range shifts and HDR visual artifacts. As shown in Fig. 7(a), the dynamic range distributions of distortions exhibit substantial heterogeneity, resulting in disparate influences on HDR image quality. Besides, different distortions exhibit pronounced overall gap in visual

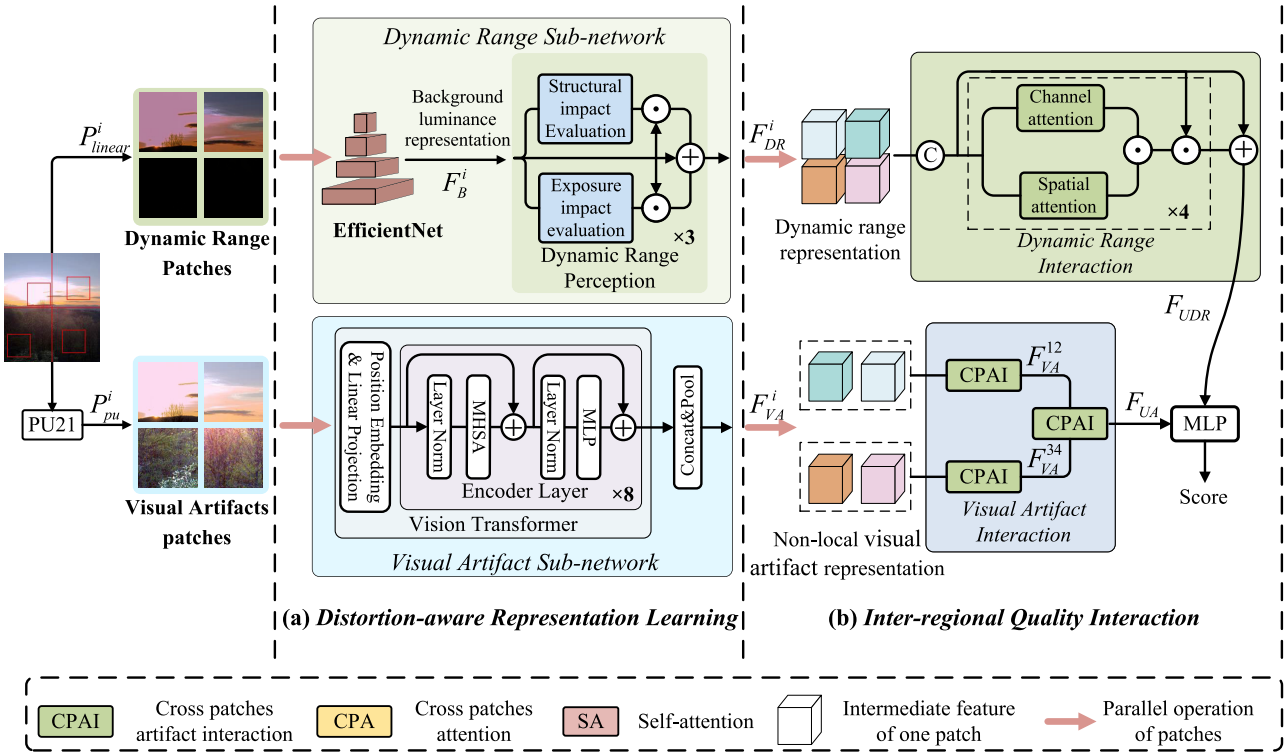


Fig. 8. The architecture of the proposed method.

artifacts. As shown in Fig. 7(b), ghosting, characterized by target motion, contrasts starkly with detail loss associated with over-exposure, showcasing the distinct visual consequences of distortion types. Therefore, dynamic range and HDR visual artifacts emerge as pivotal considerations within the spectrum of distortion content. **From the inter-regional non-uniform quality aspect**, the localized nature of distortion content leads to significant disparities in image quality across intra-image regions, which renders inter-regional quality dependencies critical for representing the overall image quality. For example, as shown in Fig. 7(c), although both low- and high-quality regions are presented in blurred and under-exposed images, they exhibit distinctly different subjective quality. This is because the inter-regional quality dependencies of blurred image fall on low-quality regions, whereas the under-exposure image exhibits the opposite tendency.

From these observations, two key issues can be summarized: 1) HDR distortions are complex and require evaluation from the perspectives of dynamic range and HDR visual artifacts to accurately measure the degree of distortion; 2) The non-uniform quality across different regions within an image makes the quality dependency relationship critical when representing the overall image quality. However, existing HDR-IQA algorithms with simple quality mapping functions are insufficient in effectively addressing these challenges, which highlights an imperative need for innovative HDR-IQA algorithms.

V. Proposed Method

A. Overall Pipeline

As analyzed in Section IV-B, the representations of complex HDR distortions and the phenomena of inter-regional non-uniform quality increase the difficulty of HDR-IQA tasks. To address these issues, a new network is proposed for HDR IQA, which independently learns the representation of distortion content within an individual region and captures the quality dependencies between different regions for better representing HDR image quality. The overall pipelne of the proposed method as shown in Fig. 8. Four representative patches are randomly selected from the quad-divided HDR input image I, and then serve as independent input content in linear domain $\{P^i_{linear}|i = 1, 2, 3, 4\}$ and perceptually uniform space $\{P^i_{pu}|i = 1, 2, 3, 4\}$. Paired P^i_{linear} and P^i_{pu} are fed into Distortion-aware Representation Learning module (DRL) to learn representations of dynamic range anomalies $\{F^i_{DR}|i = 1, 2, 3, 4\}$ and HDR visual artifacts $\{F^i_{VA}|i = 1, 2, 3, 4\}$, thereby accurately identifying distorted content in HDR images. Subsequently, all F^i_{DR} and F^i_{VA} are fed into Inter-Regional Quality Interaction module (IRQI) to capture inter-regional quality dependencies, where the united dynamic range features F_{UDR} and united artifact features F_{UA} are generated by F^i_{DR} and F^i_{VA}, respectively. Finally, the Multilayer Perceptron (MLP) with F_{UDR} and F_{UA} is employed to predict HDR image quality.

B. Distortion-Aware Representation Learning

The representation of HDR distortion content takes into account both the dynamic range and HDR visual artifacts. When the human visual system (HVS) processes the dynamic range of HDR images, the base perception of dynamic range is generated by local background luminance, then local anomalies [19] and geometric structures [60] caused by exposure errors further affect the perception of dynamic range. Besides, the visual artifacts are classified as another

important criterion for evaluating the quality of HDR images. The significant differences in visual artifacts across distortions motivate addressing this issue from a global patch perspective. Combining the above two points, the Distortion-aware Representation Learning module (DRL) consists of a Dynamic Range sub-network (DRN) and a Visual Artifact sub-network (VAN).

1) Dynamic Range Sub-Network: As shown in Fig. 8(a), the Dynamic Range Sub-network is designed in the linear domain because linear content is not influenced by any transforms and is more related to dynamic range. The EfficientNet-B0 [61] is employed to extract the background luminance representation F_B^i. Specifically, the F_B^i is achieved by utilizing convolution and pooling to integrate the outputs of the 3rd, 4th, 6th, and 9th stages of EfficientNet-B0, which improves the network's ability to characterize the dynamic range. Subsequently, a parallel attention mechanism is introduced to evaluate the effect of geometric structure and local exposure. Compared to refined geometric structure, the glare and area cutoff caused by exposure errors require wider range detection. For this reason, a multi-scale Dynamic Range Perception module (DRP) is designed to extract the exposure error and geometric structure priors. The dynamic range features extraction is formulated as follows

$$F_{DR}^i = DRP(F_B^i) \tag{1}$$

where the superscript i is ith patch, F_B^i is background luminance representation and F_{DR}^i is dynamic range representation extracted by DRP. The DRP details as follows

$$\begin{aligned} DRP(B) &= W_{EIE} \odot B + W_{SIE} \odot B + B, \\ W_{SIE} &= \sigma\{\sum_{k=1,s=2}^{3} Convk(Pool(B)) + GAP(B)\}, \\ W_{EIE} &= \sigma\{\sum_{k=5,s=2}^{9} Convk(Pool(B)) + GAP(B)\}, \end{aligned} \tag{2}$$

where B is background luminance representation, W_{EIE} is attentional weight for evaluate exposure impact, W_{SIE} is attentional weight for evaluate structual impact, $\odot$ is Broadcasting product, $Convk$ is Convolution operation with kernel size k and stride 1, GAP is global average pool operation, σ is sigmoid activation function, $Concat$ is concating in channel dimension.

2) Visual Artifact Sub-Network: In the Visual Artifact Subnetwork, as shown in Fig. 8(a), the HDR images are first converted to perceptual space with the PU21 [19], which is a global transform and is beneficial for extracting features from a global perspective. The tailored ViT-B/16 [62] is choosed as visual artifact branch network to extract visual artifact prior. The outputs of 5th-8th stage of encoder layers are integrated as patches non-local visual artifact priors, which can be formulated as follows

$$F_{VA}^i = Conv1(Concat\left\{F_j^{'i}\right\}_{j=5}^{8}) \tag{3}$$

where $F_j^{'i}$ is output of jth stage of ViT and F_{VA}^i is nonl-local texture prior of ith patch.

Finally, the four patches' representations of the dynamic range $\{F_{DR}^i | i = 1, 2, 3, 4\}$ and the non-local visual artifact $\{F_{VA}^i | i = 1, 2, 3, 4\}$ are respectively extracted by the dynamic range sub-network and the visual artifact sub-network.

C. Inter-Regional Quality Interaction

While the DRL network proposed in the previous section effectively learns the representation of HDR distortions, the inherent localized nature of these distortions results in non-uniform HDR image quality distribution. Directly utilizing distortion features from different regions to represent HDR image quality can lead to performance degradation. To address this problem, Inter-regional Quality Interaction with the dynamic range interaction and the visual artifact interaction modules is proposed, which captures inter-region quality dependencies by leveraging features associated with dynamic range and visual artifacts, respectively. This approach enables more accurate representation of HDR image quality, which is detailed as follows.

1) Dynamic Range Interaction: The dynamic range interaction is highly related to crucial HVS-inspired features, such as object surface luminance, glare, and inter-regional luminance differences, which are important for accurately representing image quality. To this end, the Dynamic Range Interaction module (DRI) is proposed, which employs channel attention and spatial attention to respectively capture inter-regional dynamic range relationships and select effective features for representing the inter-regional quality dependencies associated with dynamic range.

As shown in Fig. 9, the proposed DRI consists of dynamic range interaction blocks. Given the features to be interacted $\{F_{DR}^i | i = 1, 2, 3, 4\}$, the dynamic range interaction calculation process based on attention mechanism is as follows

$$\begin{cases} F_{UDR}^0 = Conv1(Concat\left\{F_{DR}^i\right\}_{i=1}^{4}) \\ W_k = DRIB(F_{UDR}^{k-1})) \quad , k = 1, 2, 3, 4 \\ F_{UDR}^k = W_k \odot F_{UDR}^0 \quad , k = 1, 2, 3, 4 \end{cases} \tag{4}$$

$$F_{UDR} = F_{UDR}^4 \oplus F_{UDR}^0 \tag{5}$$

where $\oplus$ is broadcasting addition operaction, $DRIB$ is dynamic range interaction block, F_{UDR}^0 is the preliminary interaction result of input features F_{DR}^i, F_{UDR}^k is the intermediate result and F_{UDR} is the final result of the interaction of F_{DR}^i using the attention mechanism. W_k are the interaction weight matrics updated with intermediate attentions, which can be expressed as

$$DRIB(M) = \sigma(C(M) \odot S(M)) \tag{6}$$

where M is intermediate result and σ is sigmoid activation function. $C(M)$ is channel attention for capturing interaction between patches and $S(M)$ is spatial attention for finding out significant dynamic range features.

2) Visual Artifact Interaction: As HDR visual artifacts are significant during global observation, the interaction between artifacts in different regions should be considered from a global perspective. Inspired by [63], the Visual Artifact Interaction module (VAI) is proposed, which adapts self-attention

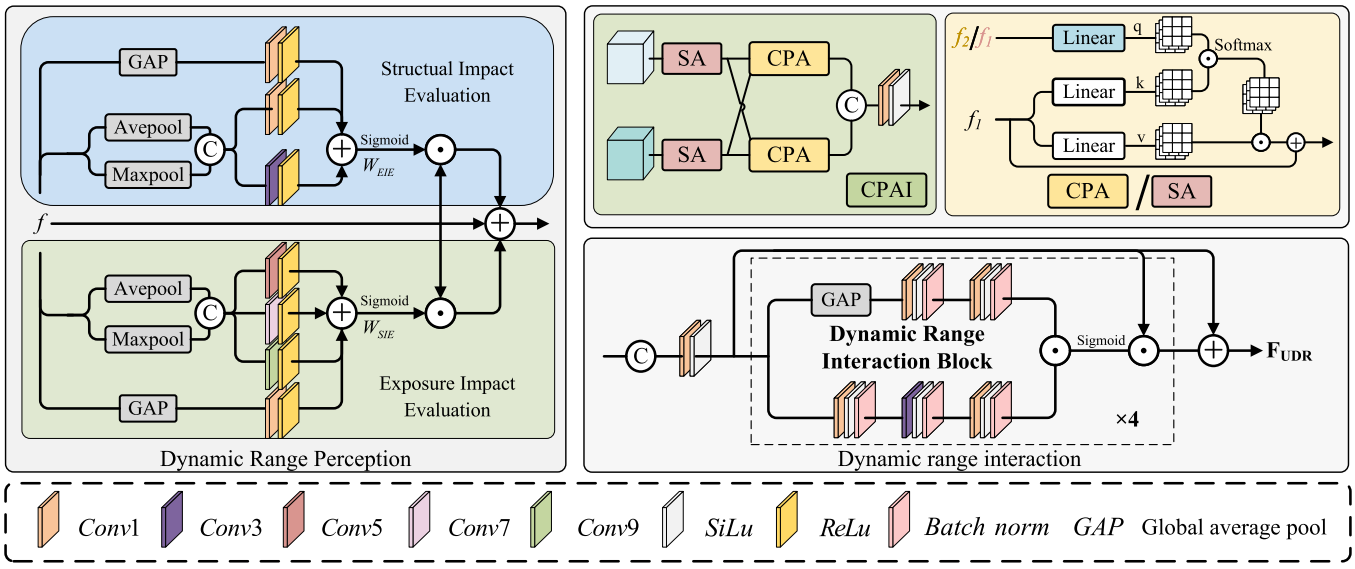


Fig. 9. The details of designed modules.

to filter salient features (e.g., texture, edge and skeletons of object) and cross-attention to capture inter-regional dependencies associated with visual artifacts.

As shown in Fig. 9, the proposed VAI consists of three Cross Patches Artifact Interaction blocks (CPAI). The CPAI contains two Cross-attention (CA) blocks, two Self-attention (SA) blocks and a 1×1 convolution layer. These attentional blocks consist of Query (Q), Key (K) and Value (V), and the difference between CA and SA mainly lies in the source of Q. Specifically, the Q, K, V of SA come from the same features, while the Q of CA comes from other features. The texture interaction calculation process based on the attention mechanism is as follows

$$F_{UA} = CPAI(F_{VA}^{12}, F_{VA}^{34}) \tag{7}$$

where *CPAI* is cross-patches texture interaction block, F_{VA}^{12} and F_{VA}^{34} are intermediate interaction result of input features $\left(F_{VA}^{1}, F_{VA}^{2}\right)$ and $\left(F_{VA}^{3}, F_{VA}^{4}\right)$ by using *CPAI*, which can be expressed as follow

$$\begin{aligned} CPAI(F_1, F_2) &= Conv1(Concat(F_{CA1}, F_{CA2})), \\ F_{CA1} &= CPA(SA(F_1), SA(F_2)), \\ F_{CA2} &= CPA(SA(F_2), SA(F_1)). \end{aligned} \tag{8}$$

where (F_1, F_2) is input features pairs, F_{CA1} and F_{CA1} are results of (F_1, F_2) utilizing cross-patch attention, *SA* is self-attention block, *CPA* is cross-patches attention block. Note that, the first input of CA is soure of K and V and the second input is source of Q.

D. Quality Prediction and Loss Function

Given a priori features F_{prior}, a feature score map and a feature weight map are generated from F_{prior}, which is achieved by two independent MLPs. The final HDR image score is calculated by weighting the score map with the weight map.

$$\hat{q} = \frac{\sum s \odot m}{\sum m} \tag{9}$$

where $\hat{q}$ is predicted score, s is score map, m is weight map, $\sum$ is summation operation, $\odot$ is Hadamard product.

The loss function is Mean Square Error (MSE).

TABLE II

THE SETTINGS OF DATASETS ADOPTED FOR EXPERIMENTS. "-" INDICATES THAT NO PARTITIONING IS APPLIED TO NARWARIA [10]

	Overall		Train Set		Test Set	
Dataset	Scenes	Images	Scenes	Images	Scenes	Images
HDRQAD	147	1409	114	1108	33	311
Korshunov [12]	20	240	15	192	5	48
Zerman [13]	10	100	8	80	2	20
UPIQ [27]	30	380	24	304	6	76
Narwaria [10]	10	150	-	-	-	-

VI. EXPERIMENT

A. Experiment Protocol

1) Dataset: There are five datasets adopted, including our HDRQAD and four publicly available datasets [10], [12], [13], [27]. As presented in Table II, the adopted HDR datasets are utilized either in their entirety or partitioned into separate training and testing sets. For dataset UPIQ, the HDR portion of the data is adapted, which consists of 380 distorted images with 30 scenes. The datasets HDRQAD, Korshunov, Zerman and UPIQ are divided into 80% for training and 20% for testing, which is performed according to original scenes to ensure the independence of image content. For the HDRQAD dataset, since the original image content is already within the range of [0, 1], we directly use the raw data as input. For the other datasets, considering that the image content does not have a unified range and compression artifacts may introduce outliers, we perform normalization based on their respective reference

TABLE III
QUANTITATIVE COMPARISON FOR TRAINING ON THE EXISTING DATASET [10], [27] OR HDRQAD, WHILE EVALUATING ON HDRQAD, [10] OR [27]

Train dataset	UPIQ [27]			Narwaria [10]			HDRQAD					
Test dataset	HDRQAD			HDRQAD			UPIQ [27]			Narwaria [10]		
Method	SRCC	PLCC	KRCC	SRCC	PLCC	KRCC	SRCC	PLCC	KRCC	SRCC	PLCC	KRCC
PU21-HyperNet [16]	0.2811	0.3420	0.1985	0.4037	0.4616	0.2928	0.7588	0.7505	0.5591	0.8989	0.8846	0.6989
PU21-VCRNet [41]	0.3144	0.2885	0.2230	0.2388	0.3543	0.1673	0.7202	0.7379	0.5317	0.8202	0.8135	0.6633
PU21-TempQT [43]	0.3151	0.3923	0.2218	0.3705	0.4539	0.2661	0.7275	0.7021	0.5317	0.8369	0.8232	0.6502
BHDRIQA [24]	0.2700	0.2728	0.1829	0.3123	0.3438	0.2164	0.7388	0.7397	0.5444	0.8147	0.8117	0.6328

images, which is similar to the preprocessing procedure of PU21 [19]. The normalization formula is as follows:

$$I_{norm} = Clip(\frac{I}{Max_{ref}}, 0, 1) \tag{10}$$

where $Clip(., 0, 1)$ denotes clipping the values that out of the range of 0 to 1, the I represent the input HDR image and Max_{ref} is maximum values of the reference image of I. Besides, if PU21 transformation is required, a luminance coefficient is multiplied, and the PU values scale to the range of 0 to 1. The PU21 transformation formula is as follows:

$$I_{PU21} = \frac{PU21(C * I_{norm})}{Max_{PU21} + eps} \tag{11}$$

where $PU21(.)$ denotes the PU21 transformation, I_{norm} denotes the normalized HDR image, $C = 10^4$ denotes the luminance coefficient, $Max_{PU21} = 595.39$ denotes the max value of PU21 and $eps = 10^{-3}$ denotes the constant that prevents abnormal values.

2) Implementation Details: The proposed system is implemented on a computer with an Intel Xeon Silver 4210R Processor, 192G RAM, and a Nvidia RTX2080Ti GPU with Pytorch 1.8.1 and CUDA 10.2. EfficientNet-B0 [61] and ViT-B/16 [62] with patch size P set to 16 are chosen as pre-trained models. The EfficientNet-B0 is pre-trained on ImageNet-21k and fine-tuned on our dataset, while ViT-B/16 is pre-trained on ImageNet-21k and fine-tuned on ImageNet-1k. The weight of EfficientNet-B0 is freezed in final training. Each image is cropped 20 times and randomly horizontally flipped each image with a given probability 0.5. We use Adam optimizer with cosine annealing learning rate with the parameters T_{max} and eta_{min} set to 50 and 0 for 15 epochs. Learning rate and mini-batch size are set to 1e-5 and 12. Representative patches from four sub-images of each test image are selected by randomly cropping 20 times during testing, and the 20 prediction results are averaged for each test image. The final performance of each experiment is tested 5 times and averaged. We perform the experiment 10 times with different seeds and report the best metrics for algorithms.

3) Performance Evaluation: The performance evaluation used is Spearman Rank order Correlation Coefficient (SRCC), Pearson Linear Correlation Coefficient (PLCC) and Kendall's Rank Correlation Coefficient (KRCC). As suggested in VQEG [68], a non-linear logistic regression function is employed to map the predicted scores to the MOS values.

B. Evaluation of Proposed Dataset

To evaluate the effectiveness of our HDRQAD, the HDRQAD is compared with the existing dataset [10], [27]. The representative IQA models [16], [24], [41], [43] are trained on the HDRQAD and compared datasets, and the performance of these models is tedsted on the HDRQAD and dataset [10], [27]. Besides, the pre-trained LDR models [16], [41], [43] are tested on the HDRQAD and compared datasets. In these experiments, the training and testing sets are each adopted as the corresponding entire dataset to perform a comprehensive evaluation of the HDRQAD dataset.

Quantitative results of HDRQAD and other datasets are shown in Table III. The proposed HDRQAD demonstrates superior generalization capabilities. As Table III demonstrates, the models trained on datasets [10], [27] struggle to achieve SRCC and PLCC results above 0.5 on HDRQAD, while the same models trained on HDRQAD consistently exceed 0.7, with some surpassing 0.85 on datasets [10], [27]. The models trained on existing datasets exhibit limited generalization capabilities to other distortion types, while the model trained on our dataset not only effectively generalizes to existing datasets but also has the ability to assess other HDR distortions.

Quantitative results of pre-trained LDR models [16], [41], [43] wiht PU21 [19] on HDRQAD and others datasets are shown in Table IV. The results further validate the diversity of distortions in HDRQAD, as pre-trained LDR models struggle more to address the distorted images in HDRQAD compared to other datasets. The evaluation metrics SRCC, PLCC and KRCC for HDRQAD are the lowest overall, with all scores remaining below 0.31. In contrast, the SRCC and PLCC metrics for the UPIQ and the Narwaria [10] exceed 0.65 for the TempQT pre-trained on LIVE [66], which are significantly higher than the metrics obtained when tested on HDRQAD. These experimental results further demonstrate that the proposed database addresses distortion types that existing databases fail to capture.

C. Evaluation of Proposed Method

The proposed method is compared with seventeen state-of-the-art quality metrics, which can be divided into 1) **HDR-IQA algorithms** that contain BHDRIQA [24], HDRQA-DISTS [47], HDR-VDP2.2 [21], DIGMS [22] and LGFM [23]; and 2) **general IQA algorithms** that consist of NIQE [39], PIQE [40], Brisque [15], HyperNet [16], VCRNet [41], TReS [42], TempQT [43], PSNR, SSIM [35], VIF [36], FSIM [37] and GFM [38]. For general IQA algorithms, HDR images are

TABLE IV

QUANTITATIVE COMPARISON OF DIFFERENT PRE-TRAINED LDR MODELS ON [10], [27] OR HDRQAD. THE " * " INDICATES MODELS PRE-TRAINED ON LDR DATASETS. BLUE INDICATES THE WORST PERFORMANCE AMONG THE SAME TYPE OF METRICS, REFLECTING THE CHALLENGES OF THE DATASET

	Test HDR dataset	Narwaria [10]			UPIQ [27]			HDRQAD		
Method	Train LDR dataset	SRCC	PLCC	KRCC	SRCC	PLCC	KRCC	SRCC	PLCC	KRCC
PU21-HyperNet* [16]	KonIQ [64]	0.4216	0.4702	0.2981	0.4968	0.5093	0.3564	0.0959	0.1143	0.0658
PU21-VCRNet* [41]	CSIQ [65]	0.0281	0.0488	0.0176	0.0206	0.0353	0.0171	0.1806	0.1846	0.1248
	TID2013 [67]	0.1166	0.1381	0.0797	0.0867	0.0652	0.0605	0.0300	0.0115	0.0204
	LIVE [66]	0.3088	0.1818	0.2130	0.1887	0.0806	0.1382	0.2469	0.2313	0.1688
PU21-TempQT* [43]	CSIQ [65]	0.3950	0.3865	0.2756	0.3408	0.3391	0.2480	0.2568	0.2489	0.1779
	TID2013 [67]	0.2568	0.2217	0.1752	0.2515	0.2237	0.1736	0.2515	0.2197	0.1710
	LIVE [66]	0.6836	0.6694	0.4927	0.7686	0.7690	0.5796	0.3064	0.2974	0.2103

TABLE V

PERFORMANCE COMPARISON IN TERMS OF SRCC, PLCC AND KRCC OF THE PROPOSED METHODS AGAINST 17 EXISTING METHODS ON HDRQAD AND THREE HDR IQA DATASETS. THE TOP-1 RESULTS IN FR METHODS ARE HIGHLIGHTED IN BOLD. THE TOP-1 AND TOP2 RESULTS IN NR METHODS ARE HIGHLIGHTED IN RED AND BLUE, RESPECTIVELY

Dataset	HDRQAD			Korshunov [12]			UPIQ [27]			Zerman [13]		
Method	SRCC	PLCC	KRCC	SRCC	PLCC	KRCC	SRCC	PLCC	KRCC	SRCC	PLCC	KRCC
PU21-PSNR (FR)	0.1005	0.2020	0.1740	0.9038	0.8916	0.7427	0.7680	0.7974	0.5790	0.8917	0.8031	0.7476
PU21-SSIM (FR) [35]	0.2886	0.2389	0.1948	0.8608	0.8677	0.7140	0.6567	0.7134	0.5005	0.9542	0.9140	0.8281
PU21-VIF (FR) [36]	0.5342	0.5605	0.3699	0.9390	0.9629	0.7876	0.9030	0.9136	0.7350	0.9565	**0.9854**	**0.8706**
PU21-FSIM (FR) [37]	0.6726	0.6787	0.4954	0.8740	0.8359	0.7583	0.7915	0.7730	0.6025	0.9546	0.9713	0.8706
PU21-GFM (FR) [38]	0.6104	0.5352	0.4265	0.8726	0.8643	0.7246	0.7915	0.7740	0.6030	**0.9614**	0.9487	0.8496
HDRQA-DISTS (FR) [47]	0.5364	0.5317	0.3677	0.9434	0.9561	0.8140	0.8657	0.8594	0.6914	0.9321	0.9235	0.8532
HDR-VDP2.2 (FR) [21]	**0.7739**	**0.7617**	**0.5859**	**0.9644**	0.9292	0.8467	0.7217	0.7144	0.5600	0.9355	0.8442	0.7896
DIGMS (FR) [22]	0.3379	0.2452	0.2419	0.9565	0.9609	**0.8620**	0.9453	**0.9470**	**0.8290**	0.8941	0.8275	0.7540
LGFM (FR) [23]	0.6069	0.6279	0.4163	0.9570	**0.9785**	0.8250	**0.9497**	0.9443	0.8010	0.9551	0.9731	0.8423
PU21-NIQE (NR) [39]	0.3310	0.3643	0.2291	0.8037	0.8564	0.6360	0.8286	0.7543	0.6333	0.8438	0.8193	0.6630
PU21-PIQE (NR) [40]	0.3167	0.3689	0.2297	0.7051	0.7188	0.5327	0.7743	0.7559	0.6064	0.8881	0.9223	0.7705
PU21-Brisque (NR) [15]	0.2734	0.2800	0.1920	0.7666	0.8135	0.6410	0.6567	0.6577	0.4756	0.7148	0.7090	0.5986
PU21-HyperNet (NR) [16]	0.8669	0.8678	0.6863	0.9524	0.9737	0.8177	0.9367	0.9328	0.7822	0.9408	0.9630	0.8126
PU21-VCRNet (NR) [41]	0.8483	0.8552	0.6601	0.9315	0.8872	0.7806	0.9317	0.9290	0.7704	0.9553	0.9597	0.8844
PU21-TReS (NR) [42]	0.8598	0.8670	0.6789	0.9356	0.9567	0.7934	0.9332	0.9288	0.7781	0.8568	0.8627	0.7789
PU21-TempQT (NR) [43]	0.8735	0.8793	0.6955	0.9287	0.9573	0.7702	0.9549	0.9512	0.8186	0.9038	0.8851	0.7368
BHDRIQA (NR) [24]	0.7901	0.7856	0.5911	0.9405	0.9431	0.7952	0.8746	0.8720	0.6870	0.9143	0.9390	0.7789
Ours (NR)	0.9090	0.9143	0.7434	0.9677	0.9813	0.8602	0.9550	0.9561	0.8193	0.9573	0.9650	0.8358

transformed into the perceptual space by using PU21 [19]. The experimental settings are as follows.

- *Quantitative Results*: The training and testing sets are split in an 80%/20% ratio to validate the performance of the proposed method on a single dataset. And the non-learned methods are only tested on test set, while the learned methods are trained on trained set and tested on test set.
- *Cross-Dataset Testing*: The training and testing datasets are the corresponding entire dataset to evaluate the generalization ability of the proposed method across different datasets.
- *Ablation Experiments*: The training and testing sets are split in an 80%/20% ratio to validate the effectiveness of each component of the proposed method.

1) Quantitative Results: Table V shows the performance comparison of different IQA models. The proposed method achieve the best overall performance on these datasets. As Table V demonstrates, the proposed method achieves the best performance on HDRQAD, with SRCC, PLCC, and KRCC scores approximately 3.6%, 3.5%, and 5% higher than the suboptimal method (TempQT), respectively. Attributed to the large-scale HDRQAD, deep learning-based IQA models, such as TempQT and TReS, are revitalized, allowing them to be adapted for HDR-IQA tasks. However, they demonstrate only moderate performance because they do not account for the HDR distortions characteristics. Traditional methods, such as VIF and LGFM, fail to predict image quality on HDRQAD because they rely on handcrafted features, limiting their ability to handle complex distortions. The inferior performance of FR metrics on HDRQAD can be attributed to their lack of consideration for the impact of non-uniform image quality distribution on overall quality. In contrast, the proposed method, which considers the characteristics of HDR distortions and non-uniform quality in HDR images, performs much better than the other models.

2) Result Visualization: In Fig. 10, the scatter plots of the MOS values against the predicted scores of all HDR IQA metrics are presented. The concentrated distribution of data points result from the proposed method has minimal bias and noise, which indicates it can stably and accurately assess HDR image quality across various distortions.

TABLE VI

SRCC, PLCC AND KRCC ON THE CROSS-DATASET VALIDATION. [12] AND [13] ARE SELECTED AS TEST DATASETS FOR THE ABLATION EXPERIMENTS. FOR THE UPIQ TRAINING SET, ONLY [13] IS USED IN THE TEST SET BECAUSE UPIQ ALREADY INCLUDES [12]; FOR THE [13] TRAINING SET, [12] SERVE AS THE TEST SET. THE TEST SET IS REFINED TO REMOVE DUPLICATE CONTENT FROM THE TRAINING SET. THE " * " INDICATES MODELS PRE-TRAINED ON LDR DATASETS. SPECIFICALLY, HyperNet IS PRE-TRAINED ON THE KONIQ [64], WHILE VCRNet, TReS, AND TempQT ARE ALSO PRE-TRAINED ON LIVE [67]. THE "-" INDICATES UNAVAILABILITY. THE TOP-1 RESULTS ARE HIGHLIGHTED IN RED AND THE TOP-2 RESULTS ARE HIGHLIGHTED IN BLUE

Train dataset	HDRQAD						Zerman [13]			UPIQ [27]		
Test dataset	Korshunov [12]			Zerman [13]			Korshunov [12]			Zerman [13]		
Method	SRCC	PLCC	KRCC	SRCC	PLCC	KRCC	SRCC	PLCC	KRCC	SRCC	PLCC	KRCC
PU21-HyperNet* [16]	0.3552	0.4250	0.2478	0.4028	0.3418	0.2817	-	-	-	-	-	-
PU21-VCRNet* [41]	0.3643	0.3477	0.2529	0.1776	0.1320	0.1165	-	-	-	-	-	-
PU21-TempQT* [43]	0.7437	0.7720	0.5659	0.5215	0.5464	0.3586	-	-	-	-	-	-
PU21-HyperNet [16]	0.7921	0.8270	0.6096	0.4656	0.4803	0.3231	0.8315	0.8366	0.6589	0.7556	0.7654	0.5330
PU21-VCRNet [41]	0.6328	0.6776	0.4659	0.3986	0.4365	0.2037	0.8647	0.8110	0.6665	0.6914	0.5781	0.5112
PU21-TempQT [43]	0.7908	0.8117	0.6010	0.6418	0.6739	0.4567	0.8667	0.8306	0.6724	0.7715	0.7889	0.5425
BHDRIQA [24]	0.8103	0.8310	0.6359	0.5905	0.6185	0.4126	0.7935	0.7603	0.6030	0.5815	0.5215	0.4199
Ours	0.8452	0.8733	0.6618	0.7206	0.7617	0.5167	0.8677	0.8394	0.6758	0.7625	0.7705	0.5367

TABLE VII

SRCC, PLCC AND KRCC ON THE CROSS-DATASET VALIDATION. THERE ARE TWO SETTINGS: TRAINING ON THE KORSHUNOV [12] AND TESTING ON THE NARWARIA [10] AND ZERMAN [13]; TRAINING ON THE NARWARIA [10] AND TESTING ON THE KORSHUNOV [12] AND ZERMAN [13]. THE TEST SET IS REFINED TO REMOVE DUPLICATE CONTENT FROM THE TRAINING SET. THE TOP-1 RESULTS ARE HIGHLIGHTED IN RED AND THE TOP-2 RESULTS ARE HIGHLIGHTED IN BLUE

Train dataset	Korshunov [12]						Narwaria [10]					
Test dataset	Narwaria [10]			Zerman [13]			Korshunov [12]			Zerman [13]		
Method	SRCC	PLCC	KRCC	SRCC	PLCC	KRCC	SRCC	PLCC	KRCC	SRCC	PLCC	KRCC
PU21-HyperNet [16]	0.8121	0.8172	0.6360	0.6514	0.6987	0.4563	0.8651	0.8631	0.6639	0.7588	0.7791	0.5689
PU21-VCRNet [41]	0.8359	0.8384	0.6444	0.6885	0.6738	0.4856	0.8645	0.8674	0.6661	0.6815	0.6609	0.4804
PU21-TempQT [43]	0.8071	0.8188	0.6265	0.7471	0.7490	0.5367	0.8067	0.8186	0.6277	0.6961	0.7094	0.5099
BHDRIQA [24]	0.8091	0.8198	0.6200	0.5649	0.6011	0.4062	0.8091	0.8198	0.6210	0.7522	0.7376	0.5612
Ours	0.8410	0.8445	0.6578	0.7695	0.7559	0.5649	0.8657	0.8703	0.6780	0.7901	0.8250	0.5892

3) Cross-Dataset Testing: The performance of cross-dataset testing is shown in Tables VI and VII. It can be observed that the proposed method exhibits the best overall performance, which reflects the proposed method has strong generation ability. Other methods fail to consider the characteristics of HDR distortions or inter-regional quality dependencies, resulting in suboptimal generalization ability. In contrast, the proposed method effectively identifies HDR distortion patterns and captures inter-regional quality dependencies, achieving the best generalization ability.

The experimental results trained on HDRQAD indicate some degree of performance reduction. To strengthen the reliability of the conclusions, we compare the results of pre-trained LDR models on datasets Korshunov [12] and Zerman [13]. As shown in Table VI, the proposed method outperforms not only the retrained methods but also the pre-trained methods, further demonstrating its superiority.

4) Ablation Experiment: Several ablation experiments are presented to study the effects of the proposed HDR-IQA algorithm on HDRQAD. Firstly, the effectiveness of each component in the proposed algorithm is validated. Subsequently, the performance improvement of the proposed method has been validated to stem from the design of the distortion module rather than an increase in input content.

To verify the effectiveness of the proposed DRL and IRQI, we compare the SRCC and PLCC of the following combinations: 1) DRL, which verifies the effectiveness of DRN and VAN; 2) Part of DRL+IRQI, which verifies the effectiveness of VAN and VAI; 3) DRL+IRQI, which verifies the complementarity of each component. As shown in Table VIII, both DRN and VAN individually exhibit good prediction accuracy. Compared to DRN, DRN+VAN achieves improvements of 2.2% and 3.1% in SRCC and PLCC, respectively. The performance of part of DRL+IRQI has comprehensively surpassed that of DRL, demonstrating the effectiveness of VAI and DRI. Finally, the complete DRL+IRQI achieves the best performance with 0.9% and 0.7% improvements in SRCC and PLCC, respectively, over the suboptimal components, demonstrating the complementarity of each component.

TABLE VIII

SRCC AND PLCC RESULT FROM THE ABLATION EXPERIMENTS ON HDRQAD. PART DENOTES THE ABLATION IS ONLY PERFORMED FOR DYNAMIC RANGE DISTORTION OR HDR VISUAL ARTIFACTS

	DRN	VAN	DRI	VAI	SRCC	PLCC
DRL	✓	×	×	×	0.8719	0.8681
	×	✓	×	×	0.8922	0.8955
	✓	✓	×	×	0.8944	0.8993
DRL + IRQI (Part)	✓	×	✓	×	0.8882	0.8927
	×	✓	×	✓	0.8998	0.9073
DRL + IRQI	✓	✓	✓	✓	0.9090	0.9143

To verify that the performance improvement results from accurate distortion representation rather than additional input

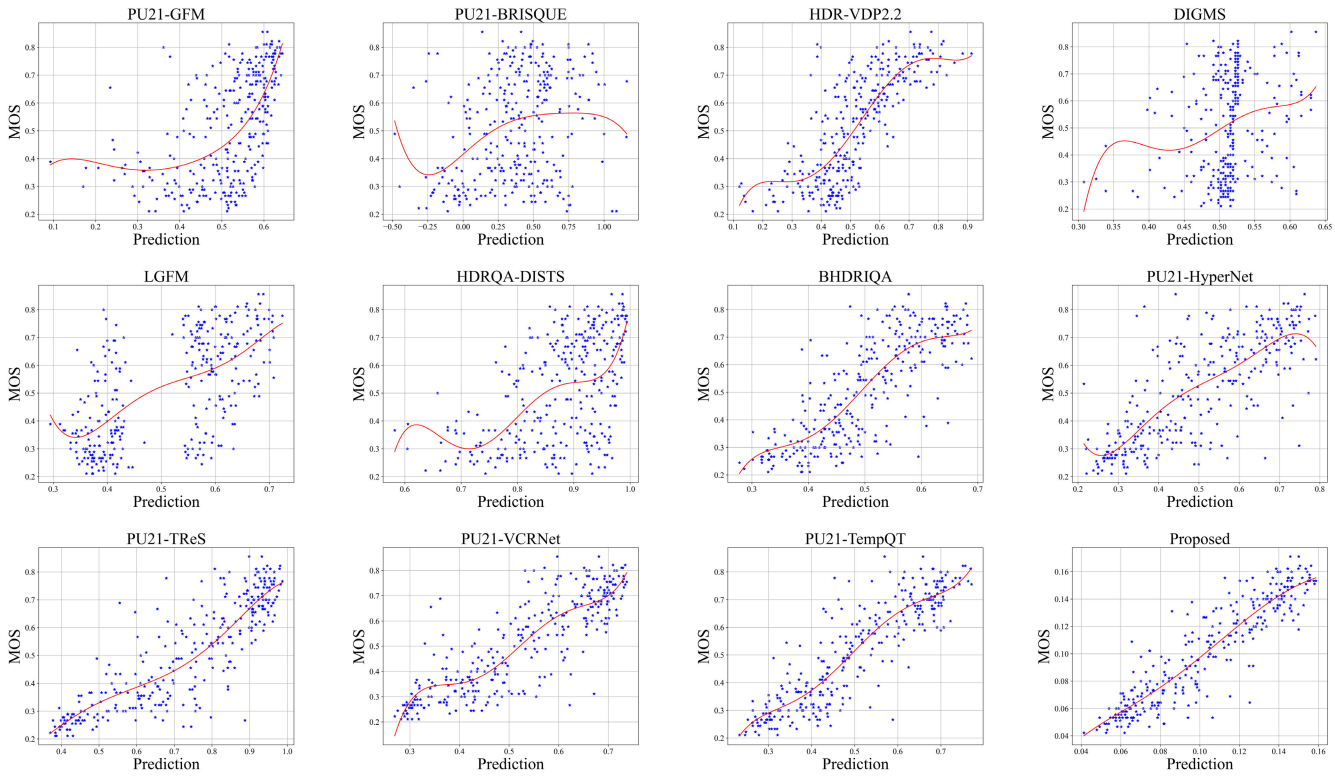


Fig. 10. Scatter plots of the objective scores against the subjective MOS values on HDRQAD testing set.

TABLE IX

SRCC AND PLCC RESULT FROM THE DISTORTION REPRESENTATION ABLATION EXPERIMENTS WITH A SINGLE PATCH ON HDRQAD

EfficientNet	DRP	VAN	SRCC	PLCC
✓	×	×	0.8170	0.8198
✓	✓	×	0.8288	0.8351
×	×	✓	0.8775	0.8851
✓	✓	✓	0.8944	0.8993

content, we conducted ablation experiments on a single patch, excluding IRQI components. We compare the SRCC and PLCC of the following combinations: 1) EfficientNet, which represents the influence of background luminance; 2) DRP, which represents the influence of exposure error; 3) VAN, thich represents the influence of HDR visual artifacts. As shown in Table IX, compared to EfficientNet, EfficientNet+DRP shows improvements of 1.1% and 1.5% in SRCC and PLCC, respectively. The SRCC and PLCC of EfficientNet+DRP+VAN both achieves 0.89, with improvements of 1.7% and 1.4% over VAN. Notably, EfficientNet+DRP+VAN outperforms the suboptimal method (TempQT, as shown in Table V) by 2.1% and 2.0% in SRCC and PLCC, respectively, verifying the effectiveness of distortion representation in the proposed method.

5) Computational Complexity: As shown in Table X, we compare the complexity of NR-IQA methods, including SRCC, time (s), VRAM (G), FLOPs (G) and parameters (M). HDR images are tested on a computer with an Intel Xeon Silver 4210R processor and an RTX 2080 Ti GPU. It can be observed that the proposed method has a relatively longer inference time, primarily because it considers more image content to better identify distortion patterns and predict image quality. Specifically, during the inference of a HDR image, the proposed method's Distortion Representation Learning (DRL) module processes four distinct image patches. Subsequently, the Inter-Region Quality Interaction (IRQI) module handles the feature representations of these patches. Compared to other methods with a single stage that only processes a single patch, the proposed method requires more inference time. Although the proposed method has a relatively longer inference time, it achieves the best SRCC performance, surpassing the suboptimal method by 3.5%. Moreover, the VRAM, FLOPs, and parameter requirements of the proposed method remain within acceptable levels. These characteristics collectively indicate that the proposed method offers significant potential for practical applications.

TABLE X

COMPUTATIONAL COMPLEXITY OF NR-IQA METHODS

	SRCC	Time (s)	VRAM (G)	FLOPs (G)	Params (M)
HyperNet [16]	0.8699	0.93	1.33	4.33	27.38
VCRNet [41]	0.8483	1.27	1.36	10.26	14.93
TempQT [43]	0.8735	1.01	4.10	143.13	240.12
BHDRIQA [24]	0.7901	1.74	1.21	0.071	8.90
Ours	0.9090	1.98	1.93	36.93	74.80

D. Limitation

The proposed method does not explicitly account for semantic information, leading to discrepancies between predicted scores and MOS values in scenarios where semantic interpretability significantly influences human perception. As shown in Fig. 11, the predicted scores for all three images

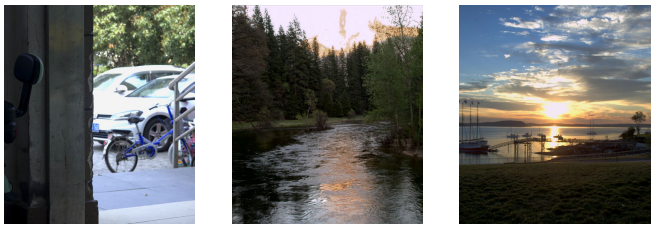


Fig. 11. Some failure cases of proposed method. MOS denotes the mean opinion score. Pre denotes the predicted score of proposed method.

are higher than their corresponding MOS values. These images exhibit a clear foreground but abnormal exposure in the background, potentially leading participants to assign relatively lower quality scores. This phenomenon aligns with the characteristics of human visual perception, that meaningful content is evaluated holistically rather than in isolated components. For instance, in a natural scene, certain artificial or natural elements, in conjunction with the background, collectively construct a scene with comprehensive semantic information. Therefore, in such cases, the semantic interpretability of distorted content plays a critical role in participants' judgment of image quality. However, the proposed method integrates distorted content and inter-regional quality dependencies but does not explicitly account for semantics, resulting in predicted scores that may not align with MOS values. These observations indicate that semantic information has a significant impact on the representation of HDR image quality. In our future work, we aim to explore mechanisms that interpret the perceptual significance of semantic content and contextual information, thereby further enhancing the consistency between model predictions and human perception.

VII. Conclusion

To advance the development of HDR IQA, we constructe a HDR-IQA dataset with the widest range distortions and the largest scale, named HDRQAD, which acquires 1409 HDR images by considering the distortions introduced in three HDR imaging schemes. The HDRQAD covers plentiful natural scenes and typical HDR quality degradation conditions. In contrast to most existing HDR-IQA methods designed just for compression distortion, an end-to-end network is proposed that effectively captures the representation of HDR distortions and addresses the issue of regional non-uniformity in quality, enabling precise characterization of the overall image quality. The experimental results prove the superiority of proposed HDRQAD and demonstrate that the proposed network achieves state-of-the-art performance.

References

[1] A. Chalmers and K. Debattista, "HDR video past, present and future: A perspective," *Signal Process., Image Commun.*, vol. 54, pp. 49–55, May 2017.

[2] Y. Shu, L. Shen, X. Hu, M. Li, and Z. Zhou, "Towards real-world HDR video reconstruction: A large-scale benchmark dataset and a two-stage alignment network," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2024, pp. 2879–2888.

[3] R. Li et al., "UPHDR-GAN: Generative adversarial network for high dynamic range imaging with unpaired data," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 32, no. 11, pp. 7532–7546, Nov. 2022.

[4] X. Hu, L. Shen, M. Jiang, R. Ma, and P. An, "LA-HDR: Light adaptive HDR reconstruction framework for single LDR image considering varied light conditions," *IEEE Trans. Multimedia*, vol. 25, pp. 4814–4829, 2023.

[5] Y. Xu, Z. Liu, X. Wu, W. Chen, C. Wen, and Z. Li, "Deep joint demosaicing and high dynamic range imaging within a single shot," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 32, no. 7, pp. 4255–4270, Jul. 2022.

[6] H. Rebecq, R. Ranftl, V. Koltun, and D. Scaramuzza, "High speed and high dynamic range video with an event camera," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 43, no. 6, pp. 1964–1980, Jun. 2021.

[7] J. Han et al., "Hybrid high dynamic range imaging fusing neuromorphic and conventional images," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 45, no. 7, pp. 8553–8565, Jul. 2023.

[8] F. Xu, J. Liu, Y. Song, H. Sun, and X. Wang, "Multi-exposure image fusion techniques: A comprehensive review," *Remote Sens.*, vol. 14, no. 3, p. 771, Feb. 2022.

[9] P. Hanji, R. Mantiuk, G. Eilertsen, S. Hajisharif, and J. Unger, "Comparison of single image HDR reconstruction methods—The caveats of quality assessment," in *Proc. ACM SIGGRAPH*, 2022, pp. 1–8.

[10] M. Narwaria, M. P. Da Silva, P. Le Callet, and R. Pepion, "Tone mapping-based high-dynamic-range image compression: Study of optimization criterion and perceptual quality," *Opt. Eng.*, vol. 52, no. 10, Oct. 2013, Art. no. 102008.

[11] G. Valenzise, F. De Simone, P. Lauga, and F. Dufaux, "Performance evaluation of objective quality metrics for HDR image compression," in *Proc. Appl. Digit. Image Process. XXXVII*, vol. 9217. Bellingham, WA, USA: SPIE, 2014, pp. 78–87.

[12] P. Korshunov, P. Hanhart, T. Richter, A. Artusi, R. Mantiuk, and T. Ebrahimi, "Subjective quality assessment database of HDR images compressed with JPEG XT," in *Proc. 7th Int. Workshop Quality Multimedia Exp. (QoMEX)*, May 2015, pp. 1–6.

[13] E. Zerman, G. Valenzise, and F. Dufaux, "An extensive performance evaluation of full-reference HDR image quality metrics," *Qual. User Exper.*, vol. 2, no. 1, pp. 1–16, Dec. 2017.

[14] M. Narwaria, M. P. Da Silva, P. Le Callet, and R. Pépion, "Impact of tone mapping in high dynamic range image compression," in *Proc. Int. Workshop Video Process. Qual. Metrics Consum. Electron.*, 2014, pp. 1–7.

[15] A. Mittal, A. K. Moorthy, and A. C. Bovik, "No-reference image quality assessment in the spatial domain," *IEEE Trans. Image Process.*, vol. 21, no. 12, pp. 4695–4708, Dec. 2012.

[16] S. Su et al., "Blindly assess image quality in the wild guided by a self-adaptive hyper network," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2020, pp. 3667–3676.

[17] T. O. Aydin, R. Mantiuk, and H.-P. Seidel, "Extending quality metrics to full luminance range images," in *Human Vision and Electronic Imaging XIII*, vol. 6806. Bellingham, WA, USA: SPIE, 2008, Art. no. 6806B1.

[18] S. Miller, M. Nezamabadi, and S. Daly, "Perceptual signal coding for more efficient usage of bit codes," *SMPTE Motion Imag. J.*, vol. 122, no. 4, pp. 52–59, May 2013.

[19] R. K. Mantiuk and M. Azimi, "PU21: A novel perceptually uniform encoding for adapting existing quality metrics for HDR," in *Proc. Picture Coding Symp. (PCS)*, Jun. 2021, pp. 1–5.

[20] R. Mantiuk, K. J. Kim, A. G. Rempel, and W. Heidrich, "HDR-VDP-2: A calibrated visual metric for visibility and quality predictions in all luminance conditions," *ACM Trans. Graph.*, vol. 30, no. 4, pp. 1–14, Jul. 2011.

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

[22] K. Zhang, Y. Fang, W. Chen, Y. Xu, and T. Zhao, "A display-independent quality assessment for HDR images," *IEEE Signal Process. Lett.*, vol. 29, pp. 464–468, 2022.

[23] Y. Liu, Z. Ni, S. Wang, H. Wang, and S. Kwong, "High dynamic range image quality assessment based on frequency disparity," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 33, no. 8, pp. 4435–4440, Aug. 2023.

[24] S. Jia, Y. Zhang, D. Agrafiotis, and D. Bull, "Blind high dynamic range image quality assessment using deep learning," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Sep. 2017, pp. 765–769.

[25] N. K. Kottayil, G. Valenzise, F. Dufaux, and I. Cheng, "Blind quality estimation by disentangling perceptual and noisy features in high dynamic range images," *IEEE Trans. Image Process.*, vol. 27, no. 3, pp. 1512–1525, Mar. 2018.

[26] M. Rousselot, É. Auffret, X. Ducloux, O. L. Meur, and R. Cozot, "Impacts of viewing conditions on HDR-VDP2," in *Proc. 26th Eur. Signal Process. Conf. (EUSIPCO)*, Sep. 2018, pp. 1442–1446.

[27] A. Mikhailiuk, M. Pérez-Ortiz, D. Yue, W. Suen, and R. K. Mantiuk, "Consolidated dataset and metrics for high-dynamic-range image quality," *IEEE Trans. Multimedia*, vol. 24, pp. 2125–2138, 2022.

[28] Y. Fang, H. Zhu, K. Ma, and Z. Wang, "Perceptual quality assessment of HDR deghosting algorithms," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Sep. 2017, pp. 3165–3169.

[29] S. Seo, S. Ki, and M. Kim, "A novel just-noticeable-difference-based saliency-channel attention residual network for full-reference image quality predictions," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 31, no. 7, pp. 2602–2616, Jul. 2021.

[30] K. Zhang, T. Zhao, W. Chen, Y. Niu, J. Hu, and W. Lin, "Perception-driven similarity-clarity tradeoff for image super-resolution quality assessment," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 34, no. 7, pp. 5897–5907, Jul. 2024.

[31] Z. Huang and S. Liu, "Perceptual hashing with visual content understanding for reduced-reference screen content image quality assessment," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 31, no. 7, pp. 2808–2823, Jul. 2021.

[32] M. Yu, Z. Tang, X. Zhang, B. Zhong, and X. Zhang, "Perceptual hashing with complementary color wavelet transform and compressed sensing for reduced-reference image quality assessment," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 32, no. 11, pp. 7559–7574, Nov. 2022.

[33] H. Li, L. Liao, C. Chen, X. Fan, W. Zuo, and W. Lin, "Continual learning of no-reference image quality assessment with channel modulation kernel," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 34, no. 12, pp. 13029–13043, Dec. 2024.

[34] H. Wang et al., "Blind image quality assessment via adaptive graph attention," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 34, no. 10, pp. 10299–10309, Oct. 2024.

[35] W. Zhou, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *IEEE Trans. Image Process.*, vol. 13, no. 4, pp. 600–612, Apr. 2004.

[36] H. R. Sheikh, A. C. Bovik, and G. De Veciana, "An information fidelity criterion for image quality assessment using natural scene statistics," *IEEE Trans. Image Process.*, vol. 14, no. 12, pp. 2117–2128, Dec. 2005.

[37] L. Zhang, L. Zhang, X. Mou, and D. Zhang, "FSIM: A feature similarity index for image quality assessment," *IEEE Trans. Image Process.*, vol. 20, no. 8, pp. 2378–2386, Aug. 2011.

[38] Z. Ni, H. Zeng, L. Ma, J. Hou, J. Chen, and K. Ma, "A Gabor feature-based quality assessment model for the screen content images," *IEEE Trans. Image Process.*, vol. 27, no. 9, pp. 4516–4528, Sep. 2018.

[39] A. Mittal, R. Soundararajan, and A. C. Bovik, "Making a 'completely blind' image quality analyzer," *IEEE Signal Process. Lett.*, vol. 20, no. 3, pp. 209–212, Mar. 2013.

[40] N. Venkatanath, D. Venkatanath, M. Chandrasekhar, S. S. Channappayya, and S. S. Medasani, "Blind image quality evaluation using perception based features," in *Proc. 21st Nat. Conf. Commun. (NCC)*, Feb. 2015, pp. 1–6.

[41] Z. Pan, F. Yuan, J. Lei, Y. Fang, X. Shao, and S. Kwong, "VCRNet: Visual compensation restoration network for no-reference image quality assessment," *IEEE Trans. Image Process.*, vol. 31, pp. 1613–1627, 2022.

[42] S. A. Golestaneh, S. Dadsetan, and K. M. Kitani, "No-reference image quality assessment via transformers, relative ranking, and self-consistency," in *Proc. IEEE/CVF Winter Conf. Appl. Comput. Vis.*, Jan. 2022, pp. 1220–1230.

[43] J. Shi, P. Gao, and A. Smolic, "Blind image quality assessment via transformer predicted error map and perceptual quality token," *IEEE Trans. Multimedia*, vol. 26, pp. 4641–4651, 2024.

[44] L. Van der Maaten and G. E. Hinton, "Visualizing data using t-SNE," *J. Mach. Learn. Res.*, vol. 9, no. 86, pp. 2579–2605, Jan. 2008.

[45] R. K. Mantiuk et al., "Practical color contrast sensitivity functions for luminance levels up to 10000 cd/m^2," in *Proc. Color Imag. Conf.*, 2020, pp. 1–6.

[46] Z. Shang et al., "A study of subjective and objective quality assessment of HDR videos," *IEEE Trans. Image Process.*, vol. 33, pp. 42–57, 2024.

[47] P. Cao, R. K. Mantiuk, and K. Ma, "Perceptual assessment and optimization of HDR image rendering," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2024, pp. 22433–22443.

[48] R. K. Mantiuk, D. Hammou, and P. Hanji, "HDR-VDP-3: A multi-metric for predicting image differences, quality and contrast distortions in high dynamic range and regular content," 2023, *arXiv:2304.13625*.

[49] M. Narwaria, M. Perreira Da Silva, and P. Le Callet, "HDR-VQM: An objective quality measure for high dynamic range video," *Signal Process., Image Commun.*, vol. 35, pp. 46–60, Jul. 2015.

[50] M. Rousselot, O. Meur, R. Cozot, and X. Ducloux, "Quality assessment of HDR/WCG images using HDR uniform color spaces," *J. Imag.*, vol. 5, no. 1, p. 18, Jan. 2019.

[51] A. Choudhury, R. Wanat, J. Pytlarz, and S. Daly, "Image quality evaluation for high dynamic range and wide color gamut applications using visual spatial processing of color differences," *Color Res. Appl.*, vol. 46, no. 1, pp. 46–64, 2021.

[52] F. Banterle, A. Artusi, A. Moreo, and F. Carrara, "NoR-VDPNet: A no-reference high dynamic range quality metric trained on HDR-VDP 2," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Oct. 2020, pp. 126–130.

[53] D. Kundu, D. Ghadiyaram, A. C. Bovik, and B. L. Evans, "No-reference image quality assessment for high dynamic range images," in *Proc. 50th Asilomar Conf. Signals, Syst. Comput.*, Nov. 2016, pp. 1847–1852.

[54] N. K. Kalantari and R. Ramamoorthi, "Deep high dynamic range imaging of dynamic scenes," *ACM Trans. Graph.*, vol. 36, no. 4, pp. 1–144, 2017.

[55] T. Brooks and J. T. Barron, "Learning to synthesize motion blur," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 6840–6848.

[56] M. D. Fairchild, "The HDR photographic survey," in *Proc. Color Imag. Conf.*, vol. 15, 2007, pp. 233–238.

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

[58] D. Marnerides, T. Bashford-Rogers, J. Hatchett, and K. Debattista, "ExpandNet: A deep convolutional neural network for high dynamic range expansion from low dynamic range content," *Comput. Graph. Forum*, vol. 37, no. 2, pp. 37–49, May 2018.

[59] B. Series, *Methodology for the subjective assessment of the quality of Television Pictures*, document ITU-R BT.500-13, Recommendation ITU-R BT, vol. 500, no. 13, 2012.

[60] V. Hulusic, G. Valenzise, E. Provenzi, K. Debattista, and F. Dufaux, "Perceived dynamic range of HDR images," in *Proc. 8th Int. Conf. Quality Multimedia Exper. (QoMEX)*, Jun. 2016, pp. 1–6.

[61] M. Tan and Q. Le, "EfficientNet: Rethinking model scaling for convolutional neural networks," in *Proc. Int. Conf. Mach. Learn.*, 2019, pp. 6105–6114.

[62] A. Dosovitskiy et al., "An image is worth 16×16 words: Transformers for image recognition at scale," 2020, *arXiv:2010.11929*.

[63] L. Zheng, Y. Luo, Z. Zhou, J. Ling, and G. Yue, "CDINet: Content distortion interaction network for blind image quality assessment," *IEEE Trans. Multimedia*, vol. 26, pp. 7089–7100, 2024.

[64] V. Hosu, H. Lin, T. Sziranyi, and D. Saupe, "KonIQ-10k: An ecologically valid database for deep learning of blind image quality assessment," *IEEE Trans. Image Process.*, vol. 29, pp. 4041–4056, 2020.

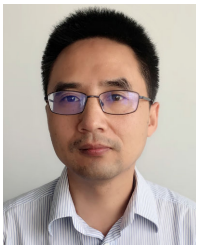
[65] D. M. Chandler, "Most apparent distortion: Full-reference image quality assessment and the role of strategy," *J. Electron. Imag.*, vol. 19, no. 1, Jan. 2010, Art. no. 011006.

[66] H. R. Sheikh, M. F. Sabir, and A. C. Bovik, "A statistical evaluation of recent full reference image quality assessment algorithms," *IEEE Trans. Image Process.*, vol. 15, no. 11, pp. 3440–3451, Nov. 2006.

[67] N. Ponomarenko et al., "Image database TID2013: Peculiarities, results and perspectives," *Signal Process., Image Commun.*, vol. 30, pp. 57–77, Jan. 2015.

[68] V. Q. E. Group et al., "Final report from the video quality experts group on the validation of objective models of video quality assessment," in *Final Report From the Video Quality Experts Group on the Validation of Objective Models of Video Quality Assessment March*. Ottawa, ON, Canada: VQEG Meeting, 2000.

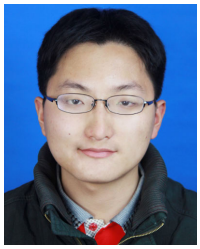
Zihao Zhou received the B.S. degree from the School of Railway Transportation, Shanghai Institute of Technology, Shanghai, China, in 2022. He is currently pursuing the M.E. degree with the Communication and Information Systems Laboratory, Shanghai University, Shanghai. His research interests include high dynamic range image processing, image quality assessment, and deep learning.



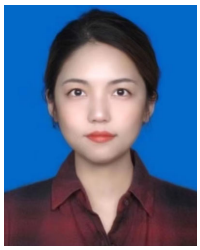
Liquan Shen received the B.S. degree in automation control from Henan Polytechnic University, Henan, China, in 2001, and the M.E. and Ph.D. degrees in communication and information systems from Shanghai University, Shanghai, China, in 2005 and 2008, respectively. Since 2008, he has been with the Faculty of the School of Communication and Information Engineering, Shanghai University, where he is currently a Professor. From 2013 to 2014, he was with the Department of Electrical and Computer Engineering, University of Florida, Gainesville, FL, USA, as a Visiting Professor. He has authored or co-authored more than 100 refereed technical papers in international journals and conferences in the field of video coding and image processing and also holds ten patents in the areas of image or video coding and communications. His research interests include versatile video coding (VVC), perceptual coding, video codec optimization, 3DTV, and video quality assessment.

Jun Lei received the M.E. degree in mechanical design and theory from Zhejiang University, Hangzhou, China, in 2004, and the Ph.D. degree in computer science and information communication from the University of Göttingen, Göttingen, Germany, in 2008. She is currently a Senior Engineer with China Mobile (Hangzhou) Information Technology Company Ltd., working in the Home Visual Communication Product Department. She has authored or co-authored numerous technical papers in international journals and conferences in the fields of broadband video communication and intelligent multimedia systems and has also contributed to the development of international standards under ITU-T. Her research interests include peer-to-peer networks, deep learning applications in multimedia communication, and video streaming technology.

Zhaoyi Tian received the B.S. degree from Jiangsu University, Zhenjiang, China, in 2021. He is currently pursuing the Ph.D. degree with the Communication and Information Systems Laboratory, Shanghai University, Shanghai, China. His research interests include image processing, video coding, and deep learning.



Xiangyu Hu received the M.S. degree from Shanghai University, Shanghai, China, in 2018, where he is currently pursuing the Ph.D. degree with the Communication and Information Systems Laboratory. His research interests include high dynamic range content processing, video encoding, and machine learning.



Shiwei Wang received the B.S. degree from the School of Communication and Information Engineering, Shanghai University, Shanghai, China, in 2020, where she is currently pursuing the Ph.D. degree. Her research interests include video enhancement, video coding, and deep learning.

Yang Chen received the B.S. degree from the School of Communication and Information Engineering, Shanghai University, Shanghai, China, in 2023, where he is currently pursuing the M.E. degree with the Communication and Information Systems Laboratory. His research interests include high dynamic range image processing, image coding, and deep learning.