applied sciences

MDPI

Review

A Review Toward Deep Learning for High Dynamic Range Reconstruction

Gabriel de Lima Martins *, Josue Lopez-Cabrejos, Julio Martins, Quefren Leher, Gustavo de Souza Ferreti, Lucas Hildelbrano Costa Carvalho, Felipe Bezerra Lima, Thuanne Paixão * and Ana Beatriz Alvarez

PAVIC Laboratory, University of Acre (UFAC), Rio Branco 69915-900, Brazil; josair21@gmail.com (J.L.-C.); julio.sousa@sou.ufac.br (J.M.); quefren.leher@sou.ufac.br (Q.L.); gustavo.ferreti@sou.ufac.br (G.d.S.F.); lucas.hildelbrano@sou.ufac.br (L.H.C.C.); felipe.bezerra@sou.ufac.br (F.B.L.); ana.alvarez@ufac.br (A.B.A.)
* Correspondence: lima.gabriel@sou.ufac.br (G.d.L.M.); thuannepaixao@gmail.com (T.P.)

Abstract: High Dynamic Range (HDR) image reconstruction has gained prominence in a wide range of fields; not only is it implemented in computer vision, but industries such as entertainment and medicine also benefit considerably from this technology due to its ability to capture and reproduce scenes with a greater variety of luminosities, extending conventional levels of perception. This article presents a review of the state of the art of HDR reconstruction methods based on deep learning, ranging from classical approaches that are still expressive and relevant to more recent proposals involving the advent of new architectures. The fundamental role of high-quality datasets and specific metrics in evaluating the performance of HDR algorithms is also discussed, as well as emphasizing the challenges inherent in capturing multiple exposures and dealing with artifacts. Finally, emerging trends and promising directions for overcoming current limitations and expanding the potential of HDR reconstruction in real-world scenarios are highlighted.

Keywords: high dynamic range; transformers; generative adversarial networks; diffusion models; image quality; performance metrics; reconstruction algorithms

Academic Editors: Atsushi Mase and Pedro Couto

Received: 27 February 2025
Revised: 6 May 2025
Accepted: 7 May 2025
Published: 10 May 2025

Citation: Martins, G.d.L.; Lopez-Cabrejos, J.; Martins, J.; Leher, Q.; Ferreti, G.d.S.; Carvalho, L.H.C.; Lima, F.B.; Paixão, T.; Alvarez, A.B. A Review Toward Deep Learning for High Dynamic Range Reconstruction. *Appl. Sci.* **2025**, *15*, 5339. https://doi.org/10.3390/app15105339

1. Introduction

Recent advances in computer vision models, significantly driven by Generative Adversarial Networks (GANs), Transformers, and Diffusion Models (DMs), have contributed to the field of high dynamic range (HDR) image reconstruction, which is an essential technology to improve the quality of digital content visualization [1–3]. The rapid evolution of the architectures in this field prompted this research, which seeks to perform a new synthesis.

Regarding high visual quality images, the human visual system easily perceives variations in brightness and contrast in a scene. Low Dynamic Range (LDR) cameras, typically operating with 8-bit resolution, are unable to capture the wide spectrum of luminosity present in various situations, resulting in loss of detail in high-contrast areas. HDR images, on the other hand, often use floating-point representations of up to 32 bits, allowing them to cover more extensive light values [4]. Thus, HDR images are considered a more faithful representation of a scene, approximating the functioning of the human visual system compared to traditional LDR images.

The advent of new generative models and attention mechanisms offers new capabilities to bridge the gap between LDR and HDR with unprecedented quality and efficiency. However, traditional models still remain relevant and will be addressed alongside the new state-of-the-art models.

More than enhancing the entertainment experience, the applications of HDR images extend to several areas of recognized importance. In the medical field, for example, the visual precision offered by HDR images is essential for more accurate diagnoses and more effective treatments. In radiology, such images can reveal subtle details that would go unnoticed in LDR images, allowing doctors to identify diseases in earlier stages [5]. Similarly, in education, HDR images can improve understanding and learning by providing more realistic and detailed representations of concepts. This aspect proves especially useful in fields like science and engineering, where the visualization of complex data is essential [6]. Also, in the field of visual communication, HDR images are ideal for advertising and graphic design, as they allow information to be conveyed more effectively and attractively.

Despite these numerous benefits, capturing native HDR images, i.e., without the need for software processing, requires specialized hardware, which makes high-quality HDR poorly accessible, as such sensors tend to be expensive and require more complex handling due to their size and weight [7]. As a result, the scientific community has investigated methods to reconstruct HDR images from a single or multiple LDR images. In this perspective, reconstruction refers to the process of creating an HDR version using the information contained in the reference with one or multiple LDR images [8]. These strategies are called single-frame and multi-frame reconstruction, respectively. The first consists of using only one LDR image to generate its HDR counterpart, while multi-frame reconstruction uses more than one image to extract color and luminosity information over a short time interval [9,10]. It is worth noting that the inherent challenges in these reconstruction approaches are being addressed with increasing sophistication by new deep learning paradigms. For example, single-frame reconstruction is an ill-posed problem that requires "hallucinating" lost information, which is a suitable task for generative models like GANs and DMs [8,10,11]. Multi-frame reconstruction, although benefiting from more information, faces critical obstacles such as motion-related artifacts (blur, ghosting, misalignment) due to the movement of objects or the camera between exposures and photometric artifacts (noise, saturation, texture loss) arising from sensor limitations and exposure differences [12–14]. These specific challenges directly motivate the development and comparison of the different deep learning strategies addressed in this research—from Convolutional Neural Networks (CNNs), optical flow, or deformable convolutions for explicit alignment [15,16] to Transformers using attention for implicit alignment and modeling long-range dependencies [17,18] and DMs offering robust generative capabilities for artifact reduction [19]. Additionally, model-induced artifacts, such as checkerboarding patterns or edge halos, also require careful architectural design and loss function selection [20,21].

Early deep learning approaches relied heavily on CNNs, using autoencoder structures [8], feedback mechanisms [22], or multi-exposure generation pipelines [23] for single-frame tasks. Multi-frame CNN methods often incorporated optical flow [15] or attention blocks [13] to handle alignment. In parallel, GANs demonstrated solid results, offering enhanced photorealism through adversarial training to simulate nuances of contrast and light intensity [24–26]. More recently, Vision Transformers (ViTs) introduced powerful self-attention mechanisms, allowing models to focus on relevant image regions and capture global context, proving effective for implicit alignment and feature fusion in the HDR, and offering an alternative to explicit alignment methods [2,17,27].

Following this same line of evolution, DMs emerged as the state of the art in generative models [3], overcoming the limitations of other methods such as stability issues in GANs [25]. Their strength lies in the precise modeling of complex distributions and the iterative refinement of the output, making them highly effective in generating coherent and high-quality images, with potential for reducing artifacts in tasks like HDR reconstruction [19,28].

In the field of HDR video generation, although it is one of the most promising applications, there is relatively less dedicated research due to, among other factors, the inherent challenges of temporal inconsistencies and artifacts arising from the fusion of consecutive frames with alternating exposures [29]. Even so, there are methods that achieve results comparable to those of static image reconstruction, including approaches that employ frame alignment via optical flow [30–32] and attention blocks [33,34].

For the proper development of these models, robust and high-quality datasets are a resource of great importance, making them fundamental for advancing research in HDR image reconstruction. Such datasets provide a solid basis for training and evaluating algorithms, allowing new techniques to be consistently compared and improved [9]. Other important characteristics that will be analyzed in this review include the type of capture (real or synthetic), resolution, number of images, and specific applications, which bring valuable contributions to the development of reconstruction methods for both single-frame and multi-frame images, as well as covering different scenarios, including mobile devices and videos.

Finally, evaluating the quality of HDR images requires the use of specific metrics that take into account the unique characteristics of this type of image. Traditional LDR metrics are not directly applicable to HDR due to differences in dynamic range and visual perception. For this reason, specialized metrics have been created, such as HDR-VDP2 [35] and HDR-VDP3 [36], which allow for a more rigorous analysis. Additionally, LDR metrics can be used to evaluate an HDR image, provided that linearization methods, like PU21 [37] or tone mapping, are used, adapting the image to these metrics. A crucial aspect, often underdiscussed and which this review aims to address, is how well these metrics correlate with subjective human judgments of image quality, as they are based on qualitative analyses with individuals.

In previous research, such as that of [29], a comprehensive review of HDR reconstruction was presented, with its scope limited to a taxonomy of classical techniques based on CNNs and GANs. Furthermore, we also acknowledge the contribution of the paper by [38], which offered an extensive review of HDR Image Quality Assessment (IQA), focusing on subjective and objective quality assessment metrics. Building upon this research, this paper advances this landscape by incorporating emerging architectures, particularly Transformers and Diffusion Models, and by examining how they address key challenges such as explicit versus implicit alignment and artifact reduction. We comparatively discuss recent methods, such as PASTA, DCDR-UNet, DiffHDR, and IREANet, detailing the specific strategies each employs to address issues of saturation, noise, and motion in dynamic scenes. Thus, our review complements the existing literature by compiling and analyzing both specific HDR metrics (e.g., HDR-VDP-3) and robust adaptations of LDR metrics (PU21-SSIM, PQ-FSIM, among others), discussing their relevance for the validation of reconstruction algorithms.

In summary, this review seeks to provide a distinct analysis of the HDR reconstruction field by synthesizing the rapid advances driven by emerging architectures (ViT, DMs) and established ones (CNN, GAN) and their effectiveness in addressing specific HDR challenges, such as alignment, artifacts, and dynamic range expansion. Additionally, the review examines evaluation methodologies, including the perceptual relevance of metrics and datasets. Finally, it highlights emerging trends such as the use of multi-modal inputs, the pursuit of efficient architectures—including the potential of State Space Models (SSMs) like Mamba—and the prospective integration with large vision models.

2. Main Concepts

2.1. Single-Frame Reconstruction

The technique of reconstructing HDR images from a single image, or single frame, aims to recover information lost in areas of high and low exposure, using a low dynamic range

image as input and generating a high dynamic range image as output. The scope of this technique can vary [8,39,40], resulting in a 16-bit depth image [41,42] or tone mapping [23,43].

The challenges of this technique mainly involve the generation of artifacts, i.e., noticeable visual errors [14]. To overcome these problems, a common strategy is to use image restoration networks with blocks that create attention maps capable of identifying overexposed and underexposed regions and mitigating their influence on the result. In this scenario, the most widespread approach is the use of architectures based on convolution blocks [44,45], although there are recent proposals that apply other strategies [46], as well as the use of specific sensors [47].

2.2. Multi-Frame Reconstruction

Reconstructing an HDR image from multiple frames involves capturing several images of the same scene, which are then combined to produce an image with a wider dynamic range than would be possible with a single frame. This approach allows additional luminance information to be collected from different exposures, which is essential for dealing with scenes with high contrast between light and dark areas.

On the other hand, this technique faces major challenges, such as the alignment of objects in relation to the reference image, noise in low-exposure images, and overexposure in high-exposure images. In order to overcome these problems, some authors have proposed specific networks for frame alignment tasks and image restoration networks that deal with natural degradations. These networks can be based on CNNs [14,48], GANs [10,10], Transformers [17,49,50], or diffusion architectures [3,28,51].

Multi-frame methods can also be classified into different subcategories. The following subsections deal with some of the main ones: burst, bracketing, and multi-exposure.

2.2.1. Burst

Burst mode, or continuous mode, involves capturing a rapid sequence of successive images. This technique is particularly useful in low-light conditions, when building high dynamic range images, and when capturing fast-moving scenes.

In the context of HDR reconstruction using deep learning techniques, burst mode can significantly improve image quality by reducing noise and increasing dynamic range. This is achieved by capturing multiple frames with constant exposures and using advanced alignment and noise reduction algorithms [52–54].

2.2.2. Bracketing

The bracketing technique can be evaluated in different ways when it comes to multi-frame or single-frame techniques. The method proposed by Debevec and Malik [55] uses a set of LDR images captured with multiple exposures to recover an HDR image. In the case of single-frame images, ref. [23] proposes an indirect approach to this multi-frame technique, in which, from a single LDR image, a set of LDR images with different exposures is synthetically generated.

2.2.3. Multi-Exposure

The technique based on multi-exposure consists of integrating images with different exposures to form a final full-exposure image [56]. Typically, LDR images are captured with different exposure times so that each image retains detail in specific regions of the scene. The reference image, usually the one with medium exposure, serves as the basis for fusion, while the other images are used to recover missing information in overexposed or underexposed regions [26].

2.3. Tone Mapping

Conventional display devices are limited by the fact that they work with only 8 bits per color channel, which restricts their ability to represent the full spectrum of brightness and contrast of an HDR image [57]. In order for these images to be properly displayed on screens with a reduced dynamic range, we resort to tone mapping, which compresses brightness and contrast values using nonlinear transformations, adapting the HDR content to the hardware limitations [58].

One of the simplest but lowest quality approaches is linear scaling [59], whose only advantage is processing speed. On the other hand, the application of a logarithmic curve [15] stands out for its good relationship between final quality and computational complexity. Other techniques go further, seeking to mimic human perception by adjusting factors such as brightness and light halos [60–62]. Figure 1 illustrates the application of logarithmic tone mapping, described in Equation (1), to an HDR image, showing how the method adjusts its pixel values for viewing on an LDR device.

(**a**) HDR image. (**b**) Tonemapped image.

Figure 1. Result of applying logarithmic tone mapping to an HDR image.

Recently, Zhu et al. [4] proposed a zero-shot scheme to transfer a tone mapping model trained on LDR to the HDR domain using very little data. Their approach decomposes the image into tonal (brightness and intensity) and structural (details and contours) components. By processing each part separately, the technique retains subtle details while effectively adjusting brightness, resulting in a conversion that retains high dynamic range characteristics.

$$I_{LDR} = \frac{\log(1 + \mu\, I_{HDR})}{\log(1 + \mu)} \tag{1}$$

where $\mu = 5000$, I_{LDR} is the resulting LDR image, and I_{HDR} is the input HDR image.

2.4. Artifacts

HDR image reconstruction, both single-frame and multi-frame, is often affected by various visual artifacts that compromise the quality of the final result. These problems can be classified into three broad groups based on their origin: motion-related artifacts (resulting from misalignments or scene changes during multi-frame capture), photometric artifacts (linked to sensor limitations and exposure differences), and model-induced structural artifacts (generated by the reconstruction algorithm itself). Understanding and mitigating these different types of artifacts are crucial for obtaining high-fidelity HDR results. The following subsections detail the most common artifacts within these categories and discuss strategies for their correction.

2.4.1. Motion-Related Artifacts: Blur, Ghosting, and Misalignment

Movements of objects (or the camera itself) during the acquisition of LDR images with different exposures cause a series of artifacts in the reconstructed HDR. Blur occurs when the displacement is slight, causing the object to appear partially “blurred” in all captures

(Figure 2a); ghosting is characterized by more pronounced displacements, resulting in multiple translucent replicas of the same object, as illustrated in Figure 2b. When the spatial discrepancy between scene elements and the background is even greater, so-called misalignment artifacts arise, manifesting as duplications or misaligned contours after fusing the exposures [12,13,63].

Mitigating these phenomena is one of the main challenges of multi-frame HDR reconstruction [9]. Recent strategies include implicit alignment via multi-scale attention modules, fusion guided by long-range dependencies, and adaptive kernels capable of compensating for complex movements, showing promising results in reducing blur, ghosting, and alignment errors [26,64,65].

(**a**) Blur artifact. (**b**) Ghosting artifact.

Figure 2. Examples of motion-related artifacts: (**a**) blur and (**b**) ghosting.

2.4.2. Photometric Artifacts: Noise, Saturation, and Texture Loss

Physical limitations of the sensor and exposure differences between LDR images can introduce random noise, which is more evident in underexposed regions (Figure 3b), and saturation or texture loss in overexposed/underexposed areas (Figure 3a), where the original information is irrecoverable. Deep convolutional networks have been used both for noise removal [66–69] and for plausible texture synthesis through dedicated spatial modules and explicit noise injection at the input to handle varying levels [41,45,70,71]. Such strategies combine feature learning with attention to context to reconstruct fine details and maintain tonal consistency throughout the HDR.

(**a**) Saturation artifact. (**b**) Noise artifact.

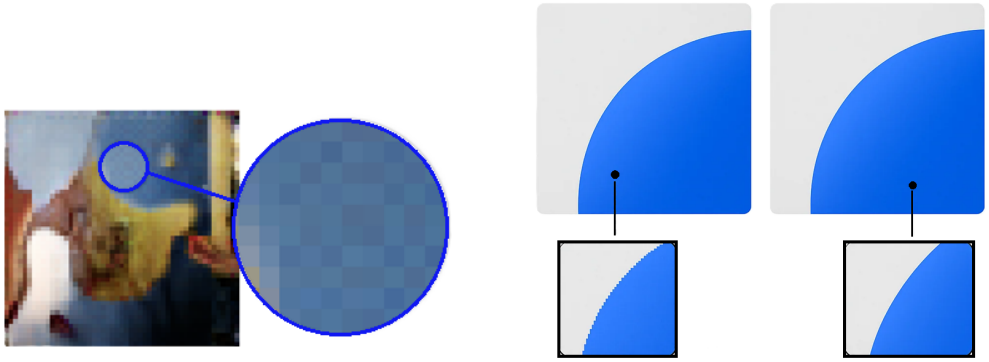
Figure 3. Examples of photometric artifacts: (**a**) saturation and (**b**) noise.

2.4.3. Model-Induced Structural Artifacts: Checkerboard and Edge

Some artifacts arise not from the input images but from the neural reconstruction process itself. Improper application of upsampling layers (e.g., transposed convolution)

can create checkerboard patterns or halos, as shown in Figure 4a, in regions of high saturation [20,72], while interpolation errors at high-frequency levels produce edge artifacts (Figure 4b), which are spurious contours around objects [21].

Effective solutions include masking critical features and architectures that avoid or replace alias-generating layers, as well as decoupled predictive kernels guided by soft masks capable of maintaining sharp edges and eliminating periodic patterns [73].



(**a**) Checkerboard artifacts. (**b**) Edge artifacts.

Figure 4. Examples of two common structural artifacts: (**a**) checkerboard patterns arising from improper deconvolutional layers [74]; (**b**) simplified contours caused by high-frequency interpolation failures.

3. Datasets and Evaluation Metrics

3.1. Datasets

The choice of dataset is an important part of the research and development of HDR image reconstruction algorithms. The success of deep learning models and image processing techniques depends largely on the availability and quality of the data used for training and validation [75]. However, obtaining HDR images is a costly and complex process, which requires captures with multiple exposures and the use of specialized equipment. In addition, the varied lighting and movement conditions in dynamic scenes further increase the difficulty of generating large volumes of data with consistent quality [26].

This section presents the main datasets used in HDR image reconstruction research that can be found available for research use and accessible for download; they are classified in a taxonomy that considers their characteristics and capture types. These datasets range from image sequences captured with multiple exposures in real environments to synthetic images generated by simulation or captured on mobile devices. Also discussed are data augmentation methods applied to recent and relevant datasets in the field of HDR imaging research, which seek to increase data variability and mitigate the limitation of small datasets.

Table 1 summarizes the datasets analyzed, with their specifications, main characteristics, and recommended application areas.

Table 1. Summary of HDR image reconstruction datasets.

Reference	Quantity	Resolution	Type	Key Features	Applications
Kalantari et al. [15]	89 sequences	1500 × 1000	Real	Dynamic Scenes, Multi-Exposure, Deghosting	HDR Reconstruction, Machine Learning
Sen et al. [65]	8 sequences	1350 × 900	Real	Extracted from HDR Videos, Continuous Motion, Deghosting	HDR Video Reconstruction
IISc VAL [48]	582 sequences	1000 × 1500	Real	Variety of Scenarios, Moving Objects, Deghosting	Multi-Frame Fusion, Deghosting

Table 1. *Cont.*

Reference	Quantity	Resolution	Type	Key Features	Applications
IISc VAL 7-1 [76]	84 sequences	1000 × 1500	Real	Dynamic Scenes, Reconstruction Method Evaluation	Deghosting, Multi-Frame Fusion
Funt et al. [77]	105 scenes	2142 × 1422	Real	Multiple Exposures, Broad Range of Scenarios, Deghosting	HDR Reconstruction, Deghosting Studies
Tursun et al. [78]	16 sequences	1024 × 682	Real	Indoor and Outdoor Scenes, Motion Focus	HDR Fusion, Deghosting
HDR-Eye [79]	46 images	1920 × 1080	Real	Multi-Frame Images of Natural Scenes, Humans, Stained Glass, Sculptures	Assessment of HDR Algorithms in Complex Scenarios
Laval Indoor [80]	2100 images	2048 × 1024	Real	High-Resolution Indoor Panoramas, Multi-Exposure	HDR Reconstruction in Indoor Environments
Laval Outdoor [80]	205 images	2048 × 1024	Real	High-Resolution Outdoor Panoramas, Multi-Exposure	HDR Reconstruction in Outdoor Environments
Liu et al. [40]	502 sequences	512 × 512	Real/ Synthetic	Indoor and Outdoor Scenes, Reconstruction from a Single LDR Image	Single-Frame HDR Reconstruction
Chen et al. [81]	1352 sequences	4K	Synthetic	LDR to HDR Conversion in Videos, Dynamic and Static Scenarios	HDR Video Compression and Quality Optimization
SI-HDR [82]	181 images	1920 × 1280	Synthetic	Simulation of Different Exposures Using CRF, Multiple Frames	Multi-Frame Fusion, Full Dynamic Range Capture
NTIRE2021 [83]	1761 images	1900 × 1060	Real HDR/ Synthetic LDR	Multi- and Single-Frame, 29 Scenes	HDR Reconstruction from LDR, NTIRE 2021 Challenge
DeepHDRVideo [31]	Real Videos	1080p	Real	Complex Scenes, HDR Algorithm Evaluation in Videos	HDR Video Reconstruction
GTA-HDR [84]	40,000 images	512 × 512	Synthetic	Scenes from GTA V Game, Variety of Lighting and Dynamic Environments	HDR Reconstruction Techniques in Synthetic Environments
Zhang et al. [85]	880 scenes	2304 × 1728	Real	Medical Images, Significant Lighting Variations	Advanced Image Processing, HDR Reconstruction
Google HDR+ [52]	3640 sequences	Various	Real	Burst Mode Images, Multiple Exposures, Less Noise	Mobile HDR Processing
MobileHDR [86]	251 sequences	4K	Real	Captured by Smartphones, Wide Range of Lighting Conditions	Mobile HDR Reconstruction
UPIQ [87]	3779 LDR/380 HDR images	Various	Real	Absolute Photometric and Colorimetric Units, Derived from Known IQA Datasets	HDR and LDR Image Quality Evaluation

3.1.1. Real Multi-Frame HDR Datasets

The dataset by Kalantari et al. [15] has 89 sequences of real LDR images captured in dynamic scenes with their corresponding HDR bracketing images. The images, with resolutions of 1500 × 1000 pixels, were collected in multiple exposures, which makes this dataset ideal for evaluating HDR reconstruction algorithms, especially in scenarios involving movement, such as deghosting. This set is widely used as one of the main references in work involving machine learning applied to HDR.

Similarly, the dataset by Sen et al. [65] contains eight sequences of LDR images extracted from real HDR videos, with a resolution of 1350 × 900 pixels. This dataset stands out for capturing continuous movement in multiple frames, which makes it especially useful for developing algorithms aimed at reconstructing HDR in videos, as well as deghosting. Despite its low number of samples, the complexity of the dynamic scenes makes it a popular choice for HDR reconstruction benchmarks.

Another important resource in multi-frames is the IISc VAL [48] dataset, which provides 582 sequences of real LDR images accompanied by their HDR deghosting versions, with a resolution of 1000 × 1500 pixels. The wide variety of scenarios, both indoors and outdoors, and the presence of moving objects make this dataset particularly suitable for testing multi-frame fusion and deghosting algorithms.

Complementing the above, IISc VAL 7-1 [76] presents 84 sequences of real LDR images, with their corresponding HDR images, also in 1000 × 1500 pixel resolution. This set focuses on evaluating HDR reconstruction methods in dynamic scenes and could be of great help in research into deghosting and multi-frame fusion in dynamic environments.

The Funt et al. [77] dataset contains 105 scenes captured with multiple exposures, resulting in HDR images with a resolution of 2142 × 1422 pixels. The images were taken with a Nikon D700 (Nikon Corporation, Tokyo, Japan) camera and cover a wide range of scenarios, from artificially lit interiors to outdoor landscapes. This is a relevant dataset for studies with high dynamic range images, as well as for testing deghosting algorithms and HDR reconstruction from multiple exposures.

In addition, the dataset by Tursun et al. [78] contains 16 sequences of images captured in indoor and outdoor scenes, with a resolution of 1024 × 682 pixels. Focusing on scenes with movement, this dataset could be an ideal addition for testing deghosting in HDR fusion algorithms.

The HDR-Eye dataset [79] consists of 46 real HDR images captured in various natural scenes, including humans, stained glass windows, sculptures, among others, with a high resolution of 1920 × 1080 pixels. This dataset provides images from multiple frames, making it suitable for evaluating HDR reconstruction algorithms, especially in scenarios involving complex lighting and detailed textures.

The Laval Indoor dataset [80] comprises 2100 high-resolution indoor panoramas captured with a Canon 5D Mark III (Canon Inc., Tokyo, Japan) camera, with a resolution of 2048 × 1024 pixels. The images were collected in multiple exposures to create HDR panoramas of indoor environments, making it a valuable resource for testing HDR reconstruction methods in indoor scenarios.

Similarly, the Laval Outdoor dataset [80] includes 205 high-resolution outdoor panoramas also captured with a Canon 5D Mark III (Canon Inc., Tokyo, Japan) at a resolution of 2048 × 1024 pixels. This dataset offers multi-frame HDR images of outdoor scenes, providing a variety of lighting conditions and environments, which is beneficial for evaluating HDR algorithms in outdoor scenarios.

3.1.2. Synthetic or Simulated HDR Datasets Based on CRF

In the field of HDR reconstruction from synthetic or simulated LDR images, the dataset by Liu et al. [40] provides 502 image sequences, both real and synthetic, with their corresponding HDR images. The images have a resolution of 512 × 512 pixels and cover a variety of indoor and outdoor scenarios. This dataset is often used in research aimed at solving the challenge of HDR reconstruction from a single LDR image, which is a key issue in deep learning.

Another important set is that created by Chen et al. [81], which contains 1352 synthetic image sequences in 4K resolution designed for research into converting LDR to HDR in videos. This dataset was created to explore the performance of neural networks in both dynamic and static scenarios and is a relevant choice of data in studies seeking to optimize the compression and quality of HDR video.

The SI-HDR [82], on the other hand, offers 181 simulated HDR images with different exposures, with a resolution of 1920 × 1280 pixels. This dataset simulates the responses of different cameras using the camera response function (CRF), allowing HDR images to be

generated from multiple frames. SI-HDR proves useful in studies that optimize the fusion of multiple frames and investigate the capture of the full dynamic range of a scene.

Chen et al. [31] present a dataset focused on capturing and reconstructing HDR in videos, called DeepHDRVideo. This dataset consists of real 1080p videos and was created to evaluate HDR video reconstruction algorithms in complex scenarios, offering a robust environment for testing performance and image quality in multiple lighting conditions.

The NTIRE2021 dataset [83] consists of 1761 images, including real HDR images and synthetic LDR images, with a resolution of 1900 × 1060 pixels. It covers 29 scenes and includes data from multiple frames and single frames. This dataset was created for the NTIRE 2021 HDR challenge, aimed at advancing the state of the art in HDR reconstruction from LDR images, and is suitable for training and evaluating HDR reconstruction algorithms using real and synthetic data.

An interesting resource in the context of synthetic data for HDR is GTA-HDR [84], which is a dataset made up of 40,000 HDR images taken from scenes in the game GTA V (Grand Theft Auto V). The images, with a resolution of 512 × 512 pixels, simulate a wide variety of lighting conditions and dynamic environments, making them valuable for testing HDR reconstruction techniques in synthetic and controlled environments.

The dataset by Zhang et al. [85] is another example of a multi-dimensional HDR application, containing 880 scenes of cholangiocarcinoma images obtained by hyperspectral microscopy and high-resolution color images (2304 × 1728 pixels). Originally designed for medical research, this dataset can also be useful in exploring advanced image processing techniques, including HDR reconstruction in scenarios with significant illumination variations.

3.1.3. Datasets of HDR Captured with Mobile Devices

In the field of processing HDR images captured by mobile devices, the Google HDR+ dataset [52] stands out for containing 3640 sequences of images captured in burst mode with smartphones Nexus 6, 5X, and 6P (Google LLC., Mountain View, CA, USA) . The images were obtained from multiple frames with different exposures, which allows for the generation of a final HDR image with less noise and greater dynamic range. This dataset is widely used in research aimed at improving the quality and computational efficiency of HDR image processing on mobile devices.

MobileHDR [86] consists of 251 sequences of LDR images captured by smartphones at 4K resolution accompanied by their corresponding HDR bracketing images. The images were captured in a wide range of lighting conditions, covering both day and night, making this dataset particularly relevant for studies focused on HDR reconstruction on mobile devices.

3.1.4. Datasets Aimed at Image Quality Assessment

UPIQ [87] is a dataset containing 3779 LDR images and 380 HDR images derived from four widely known IQA datasets: LIVE, TID2013, Narwaria, and Korshunov. The images are represented in absolute photometric and colorimetric units, corresponding to the light emitted by a screen. This dataset is widely used in research evaluating the quality of HDR and LDR images and is a reference in the field of image quality.

3.1.5. Data Enhancement for HDR

Data augmentation in the high dynamic range images is an essential technique for increasing the variability of training datasets, especially in image processing and computer vision tasks. HDR images are a particularly costly type of data to obtain, as they require multiple exposures captured in different lighting conditions, as well as specific equipment to ensure high-quality capture. Therefore, data augmentation techniques can be of great

importance in increasing the robustness and the performance of deep learning models, compensating for the scarcity of available data.

The following topics present some recent techniques used to augment HDR datasets.

Neural Augmentation for Saturation Restoration

An approach to restoring saturated regions in low dynamic range images derived from HDR scenes combines model-based methods with data-driven approaches. In this process, synthetic LDR images with different exposures are generated and refined by a neural network before being combined using exposure fusion algorithms. The aim is to improve the recovery of details in shadow and highlight areas by using a neural network to correct the exposures and then synthesize an enhanced HDR image. This technique, presented in [88], has been shown to be effective in recovering saturated regions and increasing the overall quality of the images; however, these synthetic images generated in the process could be a useful resource in data augmentation for HDR.

Local-Area-Based Mixed Data Augmentation

In order to improve HDR video reconstruction, a technique called HDR-LMDA applies localized exposure increases and RGB channel permutation to LDR images. By changing exposures in specific regions and shuffling the color channels of small blocks of the image, the model is forced to learn how to deal with the recovery of incorrect exposures and localized color correction. This method, proposed in [89], has been shown to be effective in increasing performance metrics such as the PSNR (Peak Signal-to-Noise Ratio), offering significant improvements over traditional techniques.

Traditional Data Augmentation Techniques

In addition to these more recent approaches, traditional data augmentation techniques, such as geometric transformations, modifications to the color space (brightness, contrast, and saturation), and the use of generative adversarial networks (GANs), continue to be widely used. These techniques are still an efficient way to avoid overfitting and help improve model performance, especially in scenarios with limited or unbalanced data. Some recent studies, including those of the datasets cited in this section [9,31], demonstrate how these approaches can be combined to improve HDR image reconstruction and increase model robustness.

3.2. HDR Image Quality Assessment

HDR images can represent a significantly wider range of luminance and contrast compared to traditional LDR formats. This has driven much research in HDR processing, such as acquisition/generation [43], compression [90,91], and quality assessment [92]. Image quality assessment is crucial for the development and comparison of algorithms, and the metrics used can be divided into Full-Reference (FR), which require the complete original image; No-Reference (NR), which do not require any reference; and Partial-Reference (PR) or Reduced-Reference (RR), which use a subset of information from the reference. Even so, it is difficult to apply metrics originally designed for LDR to HDR images, as they need to be revised or adapted [93]. The literature includes studies on metrics designed for the HDR, including [94], which provides a comprehensive comparison of multiple metrics; ref. [82], which subjectively evaluates the quality of reconstructions from a single image; and ref. [93], which discusses the performance of full-reference (FR) metrics in HDR scenarios. The following topics present relevant metrics in the HDR reconstruction scenario separated into two basic categories: metrics exclusive to HDR and LDR metrics that can be applied to HDR images.

Table 2 summarizes the metric information that has been described in this section.

Table 2. Summary of characteristics metrics.

Metric	Applicability	Reference Type	Evaluation Type	Domain	Computational Cost
FovVideoVDP [95]	Both	FR	Global	Spatial and Temporal	High
HDR-VQM [96]	HDR	FR	Global	Spatial	Medium
HDR-VDP2 [35]	Both	FR	Global	Spatial	High
HDR-VDP3 [36]	Both	FR	Global	Spatial	High
BRISQUE [97]	LDR	NR	Local	Spatial	Low
NIQE [98]	LDR	NR	Global	Spatial	Low
PIQUE [99]	LDR	NR	Local	Spatial	Medium
NIMA [100]	LDR	NR	Global	Spatial	Medium
LPIPS [101]	LDR	FR	Global	Spatial	Medium
SSIM [102]	LDR	FR	Global	Spatial	Low
MS-SSIM [103]	LDR	FR	Global	Spatial	Low
FSSIM [104]	LDR	FR	Global	Spatial	Low
VSI [105]	LDR	FR	Global	Spatial	Medium
PSNR [29]	LDR	FR	Global	Spatial	Low
MSE [29]	LDR	FR	Global	Spatial	Low
ColorVideoVDP [106]	Both	FR	Global	Spatial and Temporal	High
MUSIQ [107]	LDR	NR	Global and Local	Spatial	Medium

3.2.1. HDR Metrics

This subsection explores metrics aimed exclusively at HDR images and videos. Although they are still less common given the much greater volume of LDR images in the field of computer vision, they are developed to deal with the greater range of luminance and the particularities of visual perception in HDR content. It is important to highlight that metrics based on complex human visual system (HVS) models, such as the VDP family and HDR-VQM, do not rely on a single, simple mathematical formula. Instead, they implement multi-stage computational models simulating visual perception pathways. Their outputs are often visibility maps, which can then be pooled into a single quality score. For the complete detailed mathematical formulations, refer to the original publications cited for each metric.

FovVideoVDP

The Visible Difference Predictor for Wide Field-of-View Video is a video difference metric modeling spatial, temporal, and peripheral HVS aspects. Its computation involves complex stages derived from psychophysical studies: space–time contrast sensitivity functions (CSFs) adapted for foveal and peripheral vision, cortical magnification simulation, and contrast masking models. It processes video frames considering display parameters (luminance, size, resolution) and viewing distance. The output predicts the probability of visible differences over time and space. The full model details are in [95].

VDP

The original Visual Difference Predictor [108] estimates visual differences between two images using an HVS model. It typically involves amplitude nonlinearity (converting luminance to a perceptual domain), contrast sensitivity function (CSF) filtering (simulating optical blur and neural sensitivity to different frequencies), decomposition into frequency and orientation channels (cortical simulation), contrast masking calculations (predicting

how features mask each other), the calculation of detection probabilities for differences in each channel, and finally pooling probabilities into a final visibility map. The detailed steps and parameters are found in [109] and subsequent adaptations such as [110].

ColorVideoVDP

As an extension of FovVideoVDP, ColorVideoVDP evaluates distortions by jointly modeling luminance and color perception within the complex HVS framework used in FovVideoVDP. It incorporates models for color opponency and color contrast sensitivity. It predicts the visibility of spatio-temporal color and luminance distortions. Refer to [106] for the model architecture.

HDR-VQM

The Quality Measure for High Dynamic Range Videos evaluates HDR quality by simulating stages of human visual perception. Its processing workflow typically includes signal preprocessing (e.g., applying perceptual quantizers), spatial decomposition (often using steerable pyramids or wavelets), contrast calculation within the resulting sub-bands, distortion estimation based on differences between features of the reference and test signals in these bands, temporal analysis (modeling eye fixations or temporal aggregation of distortions), and finally, aggregation (pooling) of distortions across bands and time into a single quality score. The complete algorithm details, including the specific decomposition and weighting methods, can be found in [96].

HDR-VDP

The first adaptation of VDP specifically for HDR [111] modifies the core VDP components to handle the wide luminance range. Key aspects include using an HDR-specific CSF that accounts for local light adaptation levels and employing masking models suitable for high contrast ranges. It outputs a probability map of visible differences. The model is described in [110].

HDR-VDP2

A widely used successor, HDR-VDP2, refines the HDR-VDP model using improved calibration against psychophysical data. It simulates key aspects of the human visual system (HVS), including local adaptation, contrast sensitivity, and masking effects across different visual channels. This allows it to predict the visibility of differences between a reference and a test image or to estimate an overall quality score. The calculation involves pooling distortion signals computed internally within the HVS model and potentially mapping this result to align with subjective scales like Mean Opinion Scores (MOSs), often using functions calibrated against subjective datasets. While the underlying HVS model is complex, its calibration aims to provide results that correlate well with human perception. The full model details can be found in [35].

HDR-VDP3

The latest iteration, HDR-VDP3, builds upon HDR-VDP2, incorporating more advanced HVS models, including color difference perception, improved temporal modeling (for video), and better handling of different display characteristics. It aims to provide more accurate predictions of visible differences for both images and videos across various viewing conditions. The detailed architecture and validation are in [36].

3.2.2. Application of LDR Metrics to HDR Images

Although LDR quality metrics are widely used, their direct application to HDR images is not straightforward due to the nonlinearity between color values and visual perception. To address this, the HDR content must be adapted, either via tone mapping or using methods like PU21. PU21, for example, adjusts pixel values to a scale closer to human sensitivity using linearizing perception to facilitate the application of LDR metrics [37]. This makes it an alternative to using tone mapping operators as a preprocessing step when evaluating HDR images with metrics originally developed for LDR.

BRISQUE

The Blind/Referenceless Image Spatial Quality Evaluator is a no-reference metric. Its calculation involves extracting features based on Natural Scene Statistics (NSSs) and using a pretrained regression model (Support Vector Regression—SVR) to predict a quality score from these features. Due to its reliance on a trained machine learning model, it does not have a closed-form mathematical formula that fully characterizes it [97]. Lower values indicate better perceptual quality.

NIQE

The Natural Image Quality Evaluator also operates without a reference, relying on NSSs. It calculates the distance between a multi-variate statistical model fitted to the NSSs features of the test image and a precomputed model from "pristine" images. The metric represents this statistical distance and is not described by a simple algebraic formula [98]. The lower the NIQE value, the better the image quality.

PIQUE

The Perception-based Image Quality Evaluator is another no-reference metric. Its evaluation is based on detecting perceptible distortions (such as noise or blocking) in image blocks using heuristics and thresholds, followed by aggregating the distortion scores from the identified blocks. The calculation is algorithmic and rule-based, lacking a single equation that summarizes it [99]. Lower PIQUE values indicate better visual quality.

NIMA

The Neural Image Assessment metric uses a CNN to predict the statistical distribution (e.g., histogram) of quality scores that humans would assign to the image. The final score is usually the weighted average of this predicted distribution. As it uses a neural network, this metric cannot be represented by a single mathematical formula [100]. Its training on real data at the pixel level aligns this assessment with human perception. The predicted average score (usually from 1 to 10) indicates quality; higher values are better.

LPIPS

The Learned Perceptual Image Patch Similarity, a full-reference metric, measures perceptual similarity by calculating the distance between feature maps extracted by multiple layers of pretrained deep neural networks (like VGG or AlexNet) applied to both the reference and distorted images. The process involves extracting these features via the network and aggregating the distances between them, not being reducible to a simple mathematical formula [101]. Lower LPIPS values indicate greater visual similarity (better quality).

MUSIQ

The Multi-scale Image Quality Transformer evaluates image quality without reference using a Transformer-based architecture. The calculation involves dividing the image into patches at multiple scales, processing these patches through Transformer blocks with multi-

scale attention to capture local and global information, and then aggregating the resulting representations to predict a final quality score. Like other deep learning-based metrics, its complexity prevents a simplified mathematical representation of the calculation [107]. The higher the prediction value, the better the visual quality of the image.

SSIM (Structural Similarity Index)

The Structural Similarity Index (SSIM) [94] calculates the structural similarity between two images—considering luminance, contrast, and structure—providing a measure of how faithful the distorted image is compared to the reference [102]. The mathematical representation of the SSIM is given in Equation (2).

$$SSIM(x,y) = \frac{(2\mu_x\mu_y + C_1)(2\sigma_{xy} + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)} \tag{2}$$

where μ_x and μ_y are the local means of images x and y; σ_x^2 and σ_y^2 are the variances; σ_{xy} is the covariance; and $C_1 = (K_1L)^2$ and $C_2 = (K_2L)^2$ are constants to stabilize the division, with L being the maximum pixel intensity value (e.g., 255 for 8 bits). Values closer to 1 indicate better similarity.

MS-SSIM (Multi-Scale SSIM)

As an extension of the SSIM, the Multi-Scale SSIM (MS-SSIM) [94] evaluates local and global image characteristics at multiple scales, although it places less emphasis on brightness [103]. Its mathematical formulation is found in Equation (3).

$$MS\text{-}SSIM(x,y) = [l_M(x,y)]^{\alpha_M} \prod_{j=1}^{M} [c_j(x,y)]^{\beta_j} [s_j(x,y)]^{\gamma_j} \tag{3}$$

where l_M is the luminance similarity at scale M; c_j and s_j are the contrast and structure comparisons at scale j; and α_M, β_j, and γ_j are weights assigned to each component. Values closer to 1 indicate greater structural similarity.

FSIM (Feature Similarity Index)

The Feature Similarity Index (FSIM) [104] considers phase congruency and gradient magnitude to assess image similarity, emphasizing edges and textures rather than pixel values. It can be determined by Equation (4).

$$FSIM = \frac{\sum_{x\in\Omega} S_L(x) \cdot PC_m(x)}{\sum_{x\in\Omega} PC_m(x)} \tag{4}$$

where $S_L(x)$ is the local similarity based on the phase and gradient.

$PC_m(x) = \max(PC_1(x), PC_2(x))$ is the phase congruency map used as a weight; Ω represents the set of all pixels. Values closer to 1 indicate greater perceptual similarity.

VSI (Visual Saliency-Induced Index)

The Visual Saliency-Induced Index (VSI) [105] incorporates saliency maps to assign greater weight to visually relevant regions, simulating how humans perceive visual quality. Equation (5) shows its mathematical representation:

$$VSI = \frac{\sum_{x\in\Omega} S(x) \cdot VSm(x)}{\sum_{x\in\Omega} VSm(x)} \tag{5}$$

where $S(x)$ represents the combination of saliency, gradient, and chromaticity similarity; $VSm(x) = \max(VS_1(x), VS_2(x))$ highlights salient regions. Values closer to 1 indicate images of higher perceived quality.

MSE (Mean Squared Error)

The Mean Squared Error (MSE) [29] calculates the average of the squared differences in intensity between the pixels of the original and distorted images, quantifying the overall error through Equation (6).

$$MSE = \frac{1}{mn}\sum_{i=1}^{m}\sum_{j=1}^{n}[x(i,j) - y(i,j)]^2 \quad (6)$$

where $x(i,j)$ and $y(i,j)$ are the pixel values of the original and distorted images; $m \times n$ is the image size. Lower MSE values indicate less error and, therefore, better quality.

PSNR (Peak Signal-to-Noise Ratio)

The Peak Signal-to-Noise Ratio (PSNR) [29] measures the quality of reconstructed or lossy compressed images using the difference between the original and distorted images. It is determined by Equation (7):

$$PSNR = 10 \cdot \log_{10}\left(\frac{L^2}{MSE}\right) \quad (7)$$

where L is the maximum possible pixel value (e.g., 255). Higher PSNR values (in dB) indicate better quality.

4. Deep Learning Methods for HDR Reconstruction

4.1. CNN for HDR Reconstruction

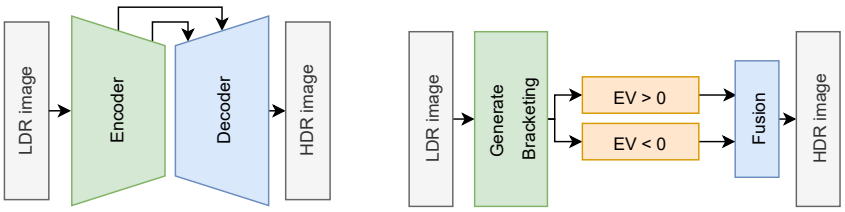
Convolutional neural networks are widely known for their success in machine learning tasks. They consist of multiple convolutional layers that multiply kernels and image input regions, thus extracting relevant features and ensuring network accuracy through a loss function optimized to guide results [112]. The works by [8,15] heralded the advent of CNNs for HDR image reconstruction in approaches that use a single image as input (single-frame) and multiple images as input (multi-frame), respectively.

To perform this task using CNNs, three predominant approaches have emerged in the area: single-frame, multi-frame, and a new hybrid approach combining single-frame and multi-frame. We will discuss and analyze the main HDR reconstruction algorithms based on the CNN below, tracing the evolution of the field to date.

4.1.1. Single LDR to HDR Reconstruction with CNN

Single-frame HDR image reconstruction methods generate their HDR output using only a single exposure as input. From this perspective, Eilersten et al. [8] proposed the HDR-CNN architecture, which utilizes skip connections introduced by the Unet network [44], resulting in a hybrid network with an LDR image encoder and an HDR image decoder. This classic architecture is shown in Figure 5a. Conversely, HDRUnet [41] proposed merging the encoding and decoding stages, enabling 8- and 16-bit image processing simultaneously. This model was designed to reduce noise and quantization error, comprising three subnetworks related to a base network, a conditional network, and a weight network, alongside the Tanh L1 loss function. In a different vein, the HDRTV [113] model introduced a structure organized into three stages: Adaptive Global Color Mapping (AGCM), Local Enhancement (LE), and Highlight Generation (HG). Building on these results, the DCDR-UNet [45] architecture, with its novel Deformable Convolution Residual Block (DCRB)

implementation and use of a VGG and Tanh L1 combined loss function, achieved promising performance in generating high-quality HDR images.



(**a**) HDRCNN architecture. (**b**) DrTMO architecture.

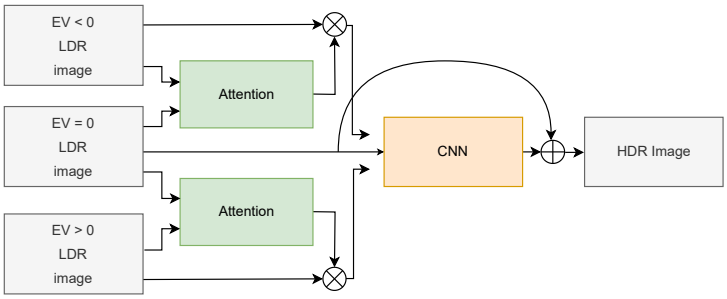
Figure 5. Single-frame architectures.

Under a different methodological approach for generating brackets to form the HDR, the DrTMO [23] architecture establishes a neural network that leverages multiple exposures for training, enabling synthetic exposure generation, as illustrated in Figure 5b. This work served as a foundation for developing subsequent models [9,114–116]. In a similar vein, Phuoc-Hieu Le et al. [9] applied an architecture following the traditional pipeline of generating HDR images. The structure comprises three CNNs: one for encoding and two for decoding (up and down exposure). The goal is to create multiple exposure values (EVs) from a single LDR image. Along these lines, Chen et al. [115] implemented continuous exposure learning, in which the network can produce EVs with continuous values, leading to results with more detail compared to using fixed EVs.

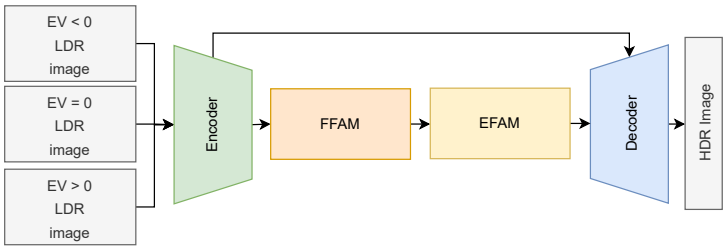
4.1.2. Multi LDR to HDR Reconstruction with CNN

Multi-frame HDR image reconstruction combines multiple images captured with different exposures to produce a final HDR image. By leveraging a variety of exposures, the model can retain more details in bright and dark regions, which helps achieve results closer to real-world scenarios. More recent methods have explored a blend of techniques such as deformable convolutions [13,16] and CNNs [117,118] to address problems like alignment, saturation, and ghosting artifacts. Kalantari et al. [15] propose using the optical flow algorithm [119] to align LDR images and then fuse them with a CNN for HDR image reconstruction. However, optical flow-based approaches can fail under occlusion or abrupt motion, leading to alternative approaches such as Wu et al. [120], which transformed the problem into an LDR-to-HDR conversion without optical flow using an encoder–decoder architecture for reconstruction.

Following this concept, ref. [13] introduced the AHDRNet architecture, which is composed of two neural networks for alignment and fusion. The first network is dedicated to extracting features from the input images by means of convolutional encoders, applying attention maps to nonreference images to identify useful features, thereby reducing ghosting artifacts. The configuration can be seen in Figure 6a. Then, the second fusion network uses these attention-guided features to estimate the HDR image using dilated residual dense blocks (DRDBs) and a global residual learning strategy (GRL). Hence, it enhances the use of extracted features, producing HDR images with consistent and realistic detail.



(**a**) AHDRNet architecture.



(**b**) IREANet architecture.

Figure 6. Multi-frame architectures.

On the other hand, IREANet [121], shown in Figure 6b, stands out for introducing the Flow-guided Feature Alignment Module (FFAM) and the Enhanced Feature Aggregation Module (EFAM). These components enable more effective alignment and reconstruction of multi-exposure LDR images, mitigating motion artifacts and improving final image quality. Moreover, using the BayerAug (Bayer preserving augmentation) strategy for preserving multi-exposure RAW data broadens the model's generalization, positioning IREANet as one of the leading solutions for HDR reconstruction.

Lastly, RepUNet [122] made progress by emphasizing efficiency and robustness using structural reparameterization to create a lightweight and fast model for HDR reconstruction, as shown in Figure 7. Additionally, APNT-Fusion [117] proposes an HDR restoration model via progressive neural fusion guided by attention, preventing ghosting artifacts from motion and efficiently transferring texture in saturated regions.

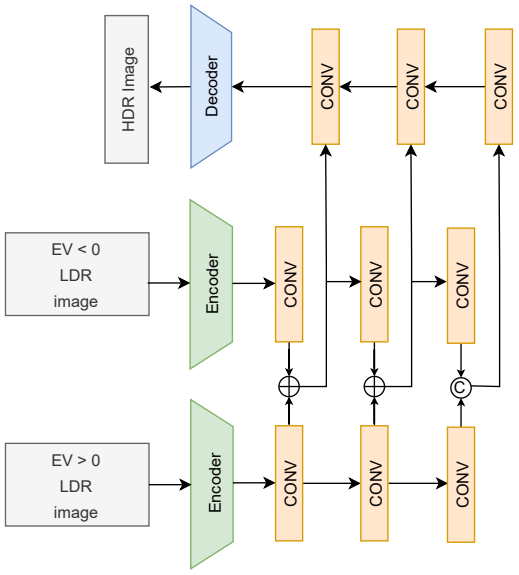


Figure 7. RepUNet architecture.

4.1.3. Single and Multi LDR to HDR Reconstruction with CNN

Li et al. [123] proposed an innovative approach for HDR image reconstruction that integrates the two main existing techniques: single-frame and multi-frame. The architecture is predominantly based on the single-frame technique but uses the multi-frame technique in an auxiliary manner, enabling the capture of more detail in overexposed areas and suppressing ghosting artifacts. The model is divided into four modules, among which SHDR-ESI stands out for integrating single-frame with multi-frame to improve final image quality. SHDR-ESI also plays a key role in providing intermediate data to the SHDR-A-MHDR module, which primarily reduces ghosting and enhances overexposed regions. Additionally, the system utilizes DEM, consisting of an SRM (Self-Representation Module) and an MRM (Mutual-Representation Module), to highlight essential information and effectively merge complementary features. The merging of reference and nonreference features is performed using FIFM, ensuring accurate and detailed HDR reconstruction.

Table 3 presents the metrics-based performance of the aforementioned models.

Table 3. Summary of metrics for each CNN approach.

Model	PSNR-L (dB)	PSNR-μ (dB)	SSIM	SSIM-μ	HDR-VDP2	LPIPS
Single-frame						
HDRUNet [41]	41.61	34.02	—	—	—	—
ArtHDR-Net [124]	35	—	0.93	—	69	—
DCDR-UNet [45]	52.47	54.71	0.9985	—	68.68	0.0026
Multi-frame						
APNT-Fusion [117]	—	43.96	—	0.9957	—	—
IREANet [121]	39.78	—	0.9556	—	—	0.102
RepUNet [122]	44.8081	—	—	—	—	—
Single-frame and Multi-frame						
SAMHDR [123]	52.49	46.98	0.9991	0.9961	69.76	—

4.2. GAN for HDR Reconstruction

This architecture was developed in 2014, but between 2018 and 2021, GANs [25] experienced their most significant surge in interest within the scientific community. In that sense, according to [125], Yann LeCun stated that GANs were the "most interesting idea in the last ten years of machine learning".

GANs have proven to be a valuable tool for generating high-quality data across various fields, including image generation and reconstruction [126,127], video synthesis [128], data augmentation [129], style transfer [130], natural language processing [131], the medical domain [132], remote sensing [133], autonomous vehicles [134], pattern classification with imbalanced data [135], and anomaly detection in time series [24,46].

GANs operate by setting up competition between a generator model, which aims to produce samples that mimic the distribution of the original data, and a discriminator model that seeks to determine whether the sample under analysis is real or fake. This competition drives improvements in both models [25]. Figure 8 illustrates how GANs function, highlighting the interaction between generator and discriminator models.

The main advantage of this architecture lies in the fact that the generator uses a distribution to directly sample from data without needing premodeling, i.e., without making explicit assumptions or approximations about the data distribution before generating new samples, thus theoretically matching the entire distribution of original data [136]. Moreover, compared to other generative architectures, GANs require fewer restrictions to be used with different data types, making them more flexible [125], and their training does not rely

on Markov chains [136]. Hence, GANs yield high-quality results for all data types, often producing images with high sharpness, clarity, and more efficient training compared to other architectures [136].

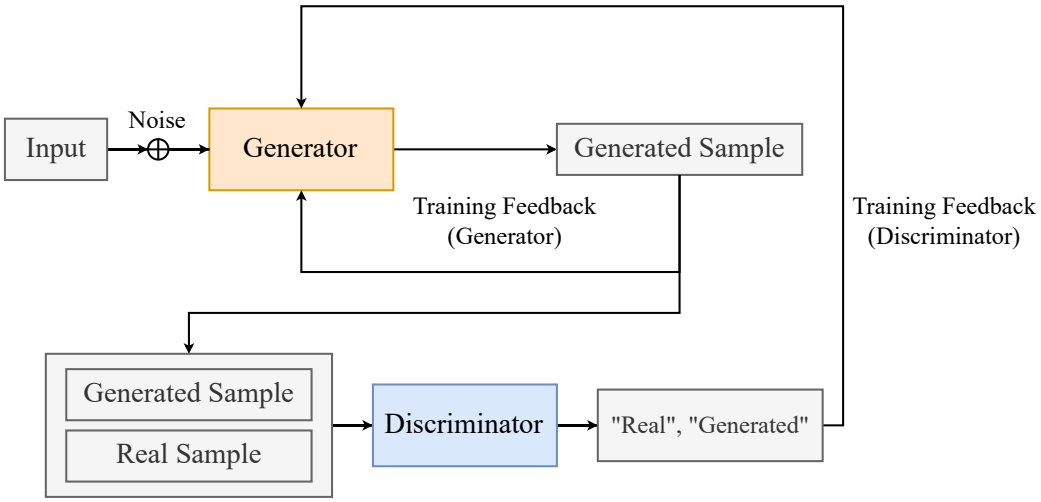


Figure 8. GAN architecture.

In the context of HDR image reconstruction, GAN-based models can generate a distribution of original data on light intensity and local contrast [25,137]. In this sense, numerous studies have been conducted using GANs to aid in recovering regions where image content is missing due to underexposure and overexposure, as well as to address the limitations of existing methods in handling large object movements in a scene.

HDR image reconstruction with GANs can be subdivided into two specific approaches: single-frame and multi-frame.

4.2.1. Single LDR to HDR Reconstruction with GAN

Reconstruction of HDR images from a single LDR input has been tackled by various methodologies, and in this scenario, GANs stand out for their ability to produce high-quality results. Hence, in the approach developed by HDR-cGAN [138], reconstruction is treated as an image-to-image translation problem using a conditional GAN (cGAN) [139], where the LDR image is concatenated with both the generator output and the ground truth (GT) HDR image in the discriminator, ensuring that the HDR output is guided by the LDR input. The generator first passes the LDR input through a Linearization module that converts the nonlinear LDR image into linear irradiance. Next, the Overexposed Region Correction (OERC) module estimates pixel values in saturated regions from the overexposure mask based on the method proposed in [140]. After that, the OERC output is fed into a Refinement module, which is responsible for correcting any remaining irregularities. The discriminator follows the PatchGAN [141] framework, evaluating the authenticity of the generated HDR image at local patches instead of assessing the entire image, which allows for finer corrections in specific regions.

Another notable approach is GlowGAN [10], which consists of three models: a generator, a discriminator, and a camera model. The generator and discriminator are based on StyleGAN-XL [142], and the camera model is the main difference among other architectures, multiplying the generated HDR image by an exposure value to fix the dynamic range and forming an LDR image similarly to how a camera functions. The camera model acts as an intermediary between generator and discriminator, allowing training on datasets containing only LDR images.

These models differ mainly in their objectives. Both HDR-cGAN and GlowGAN can generate HDR images with few artifacts. However, GlowGAN stands out for its ability to train without the need for paired datasets, i.e., without the original HDR images.

4.2.2. Multi LDR to HDR Reconstruction with GAN

A significant amount of research has been carried out aiming to reconstruct HDR images from multiple images of the same reference with different exposures. In this context, GAN-based methods show remarkable results. Among these, the approach proposed for HDR-GAN [26] stands out, consisting of a multi-scale LDR encoder that processes each input with different exposures to extract features at various scales, enabling the network to learn LDR image features for subsequent HDR construction. Then, two processes happen simultaneously: one for merging aligned reference features and another for deep HDR supervision. In the first process, features extracted at multiple scales are fused and aligned with a residual structure in the feature domain. In the second, the process is reinforced by upsampling lower features and concatenating them with higher ones, improving alignment and the final HDR image quality. Lastly, a PatchGAN-based [141] discriminator evaluates the generated image.

In the same perspective, the UPHDR-GAN [143] approach involves a generator that transforms the input image domain into the desired domain. The discriminator follows the PatchGAN [141] approach to distinguish between two images. Thus, a discriminator loss value is computed, and to obtain this value for the generator, a Min-patch module is applied, adding a pooling layer to the discriminator output.

Likewise, MEF-GAN [144] uses a generator based on three blocks: a self-attention block applying the mechanism described in [145], a local detail block to retain details that might be lost in the first block, and a merge block that combines these results to produce the final HDR image. The generated HDR image is then evaluated by a discriminator, yielding a scalar value between 0 and 1 that determines how close the generator output is to the original HDR image.

All the architectures above produce HDR images with good performance in both qualitative and quantitative results. Except for MEF-GAN, the other approaches handle the movement present in the utilized datasets, solving to some extent the misalignment of objects among different exposures. UPHDR-GAN stands out for training without needing paired datasets, while HDR-GAN [26] yields the best image metrics. Lastly, MEF-GAN is notable for its multi-exposure fusion approach that preserves crucial details and context.

The models employing the methods described in the topics above and their metrics are presented in the following sections and in Table 4.

Table 4. Summary of metrics for GAN approaches.

Method	PSNR-l	PSNR-μ	PSNR-PU	SSIM-l	SSIM-μ	SSIM-PU	MSSIM	HDR-VDP2	HDR-VDP3
Single-Frame									
GlowGAN [10]	—	—	31.8	—	—	—	—	—	7.44
HDR-cGAN [138]	17.57	—	—	0.78	—	—	—	51.94	—
Multi-Frame									
HDR-GAN [26]	41.76	43.64	43.2004	0.9869	0.9891	0.9913	—	65.45	—
UPHDR-GAN [143]	43.005	—	42.115	0.988	—	0.986	—	63.542	—
MEF-GAN [144]	68.42	—	—	—	—	—	0.982	—	—

4.3. Transformers for HDR Reconstruction

Models based on attention mechanisms have radically transformed how neural networks process and prioritize information in deep learning tasks. This revolutionary

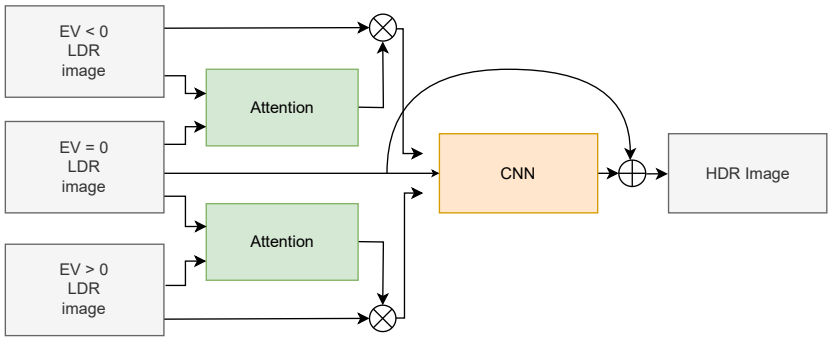
approach was introduced with the proposal of a neural network exclusively based on self-attention mechanisms, known as Transformers [18], marking a milestone in natural language processing (NLP). Such architectures stand out for their ability to focus on different regions of the input, allowing the model to learn and capture complex relationships more effectively and applying a more reliable attention mechanism, but at a higher computational cost, given that self-attention must be iteratively calculated for each input token, implying quadratic complexity.

The Transformer architecture is mainly composed of two fundamental blocks: the encoder and the decoder. The encoder takes the input sequence and transforms it into an encoded representation, while the decoder uses that representation to generate the output sequence. This architecture has proved highly accurate in managing complex linguistic structures, leading to its rapid adoption and expansion into various domains. The initial success of transformers in NLP led researchers to explore their potential in other fields. In particular, they have been adapted to computer vision, showing promising results in a range of tasks. A notable and pioneering example is the Vision Transformer (ViT) [146], which adapts the Transformer encoder architecture for image classification. In the ViT, an image is split into patches treated as tokens, akin to words in a sentence. This approach enables the ViT to grasp intricate patterns and relationships within the image, achieving superior performance in various benchmarks.

Today, Transformer-based models have become key tools in many computer vision applications. Their effectiveness has been demonstrated in tasks such as low-light image enhancement [147,148], content generation [149,150], super-resolution [49,151], and more [152–155]. This versatility underscores the power of Transformers in addressing a broad range of computational challenges.

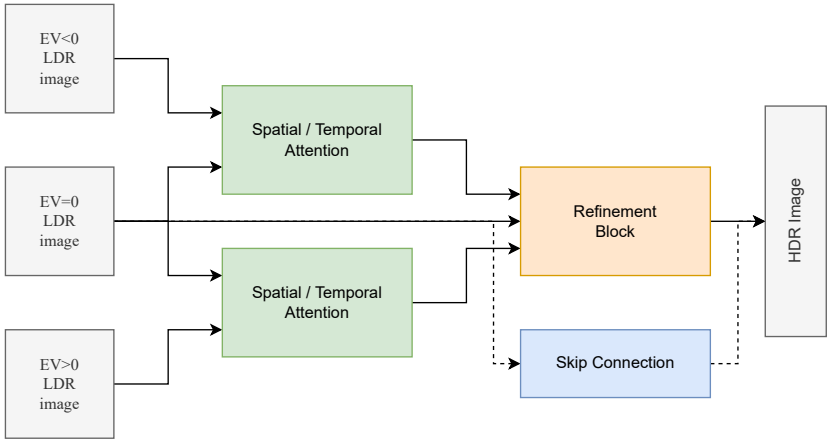
In the specific context of HDR image reconstruction, Transformers have shown significant potential for improving the quality and details of images. The most recent models have leveraged the attention mechanisms inherent to transformers to efficiently handle the various levels of luminance and complex scenes typically found in HDR content. Such progress not only boosts accuracy in HDR reconstruction but also fosters greater generalization across different scene types.

Figure 9 presents the most common approaches to HDR reconstruction, which can be broadly classified into three major methods. The first relies on aligning images with different exposures before the Transformer mechanism, the second focuses on spatial or temporal attention mechanisms for feature extraction, and the third employs methods based on feature concatenation, relying on Transformer mechanisms for the final reconstruction.

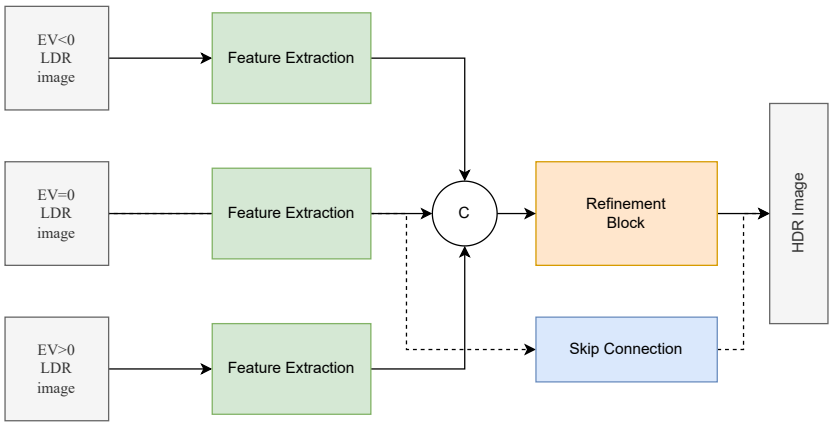


(**a**) Alignment-based transformers methods.

Figure 9. *Cont.*



(**b**) Attention-based transformers methods.



(**c**) Convolution-based transformers methods.

Figure 9. Transformer-based architectures.

4.3.1. Alignment-Based Methods

Some state-of-the-art methods focus on feature extraction with an emphasis on aligning the input frames, using the medium-exposure frame as a reference image for object positions in the scene. Alignment may be performed using convolutional blocks, attention mechanisms, or even entire Transformer architectures, leading to a clearly separable two-step approach (alignment, then feature fusion), as shown in Figure 9a.

The HFT architecture [156] uses a Shallow Feature Alignment (SFA) block for frame alignment. This block concatenates the medium-exposure frame with either the over- or underexposed frame and applies two successive convolutions to produce a single aligned frame. To refine features, the approach employs a Fusion Transformer architecture, extracting multi-scale features and merging them to reconstruct the optimal HDR image with final convolution blocks. The results were validated on the Kalantari [15] and Prabhakar et al. [48] (IISc VAL and IISc VAL 7-1 datasets) works, as well as on images captured by a cell phone to verify the generalization quality of the model. Ablation studies highlight the relevance of the SFA model, Fusion Transformers, and channel-level attention block.

Kim et al. [157] propose a different alignment approach by performing a quadruple Bayesian filter preprocessing of the input LDR images, converting them into a single single-channel image that captures information from all LDR frames in the dataset. Subsequently, the image is split into patches with four-channel depth, from which channel-level attention maps are extracted and fused with the patches via elementwise multiplication. These images with attention are aligned using a multi-scale Transformer architecture and then

refined with convolutions to reconstruct an HDR image. For validation, a synthetic dataset was introduced, along with generalization tests using the Suda et al. [158] dataset.

Another architecture using alignment-based methodology is HyHDRNet [159], aligning frames through ghost attention, carrying out alignment at the patch level by identifying misaligned elements, and performing the corresponding correction. After consecutively repeating these corrections, all frames are concatenated, including the initial reference frame. The extraction of features from the aligned images and their refinement is handled by the SwinIR [160] Transformer architecture. A final convolution plus the sum of the initial features reconstructs the HDR image from three LDR images. The results were validated with the Kalantari and Hu et al. [63] datasets and generalized using Tursun et al. [14] and Sen et al. [65]. Ablation studies highlight the efficiency of its alignment blocks.

4.3.2. Attention-Based Methods

Another approach researchers have adopted is focusing on other aspects of the frames by applying temporal or spatial attention. Unlike alignment blocks, these methods do not seek to align every object in the scene with the reference frame; instead, they emphasize certain regions based on objects from the reference scene. This is achieved by computing attention maps and multiplying that map by the frames where one wants to apply such attention, as illustrated in Figure 9b.

Examples include CA-ViT [17], which applies special attention to the features extracted from the input images using a single convolutional layer for each. After obtaining feature frames with spatial attention, these features are refined with a Transformer architecture called Context Aware Transformer (CA-ViT). The final HDR image is reconstructed with one final convolution. The model was trained on the Kalantari [15] dataset, and its generalization was verified with images from the Sen and Tursun et al. [14] datasets, with ablation studies underlining the importance of the CA-ViT block.

Similarly, the HDT-HDR [161] architecture employs spatial attention mechanisms in features extracted from the input images by a convolutional layer. While it also concatenates the features with the applied attention, it further concatenates the attention maps as additional features, refining them with a transformer architecture named HDT. The final image is reconstructed with a convolutional layer. As with previous models, it was trained on the Kalantari dataset and validated on images from Sen and Tursun et al. [14] for generalization tests.

PASTA [50] also adopts convolution to extract the input image features, applying temporal attention across frames to identify the relevant features for each one. To further refine these features, the model employs a progressive representation and aggregation network, scaling features down by factors of two up to four times using self-attention with Transformers to focus on specific regions and then upsampling back to the original scale and adding the features at the matching scale. A final convolution merges these features to yield a three-channel HDR image. The authors also propose an alternative architecture designed to reduce parameter counts at some cost in quantitative metrics. Their models were trained on the Kalantari dataset and the Tel et al. [27] dataset. Their results highlight the model's efficiency, measuring memory usage and inference time for test images.

4.3.3. Convolution-Based Methods

Some methods seek to refine features exclusively with the Vision Transformer architecture, i.e., they do not employ alignment blocks or spatial or temporal attention blocks for input features. In contrast, the methods described below simply extract image features and concatenate them to carry out all relevant processing in the Vision Transformer blocks, requiring the architecture to handle in some way the importance of the central frame relative to over- and underexposed frames, as shown in Figure 9c.

One example is the architecture proposed by Tel et al. [27], which extracts features from each input frame through convolutional blocks. These features serve as input to the Transformer blocks, being concatenated beforehand. The architecture performs the self-attention typical of Vision Transformers but also the cross-attention between the central frame and the other exposure frames to handle issues like blur, noise, and misalignment. In the final stage, refined features are added to the central frame features, and a final convolution reconstructs the HDR image.

Chi et al. [162] introduced a multi-scale architecture, reducing by a factor of two until reaching eightfold reduction and then upsampling back to the original scale. A convolution block extracts features from the initial frames at each scale, refining them with a vision transformer specialized in multi-exposure feature fusion, including additional convolutions after each transformer block.

Yan et al. [118], using a slightly different approach, propose a self-supervised network for HDR reconstruction called SSHDR. Similar to others, the model extracts features from each input frame using a convolutional layer. The features are then refined through a SwinTransformers [163] architecture with a multi-scale structure. A final convolution merges features into an HDR image. Furthermore, the self-supervised approach enables training with LDR images, both with or without HDR references.

Table 5 summarizes all the cited architectures, highlighting the strengths and limitations grouped under the three main methods. Table 6 shows the results of common HDR metrics for Transformer-based architectures evaluated on the Kalantari et al. [30] dataset.

Table 5. Transformer-based architectures, with their strengths and limitations.

Category	Architecture	Strengths	Limitations
Alignment-based	HFT-HDR [156], KIM et al. [157], HyHDRNet [159]	• Frame alignment • Multi-scale refinement • Patch-level alignment	• Supervised learning • High computational cost
Attention-based	CA-ViT [17], HDT-HDR [161], PASTA [50]	• Spatial or temporal-level attention • Emphasis on global context • Enhanced missing-information reconstruction	• Complex processing • High memory usage
Convolution-based	SCTNet [27], SV-HDR [162], SSHDR [118]	• Multiscale feature fusion • Possibility of self-supervised approach • Better reconstruction	• Requires correctly exposed images • High inference time

Table 6. Summary of metrics for HDR reconstruction models based on Transformers using the Kalantari dataset.

Architecture	PSNR-μ	SSIM-μ	HDR-VDP2
CA-ViT [17]	44.32	0.9916	66.03
SCTNet [27]	44.49	0.9924	66.65
KIM et al. [157] [1]	40.10	0.9619	75.57
HDT-HDR [161]	44.36	66.08	—
HFT-HDR [156]	44.45	0.988	—
PASTA [50]	44.53	0.9918	65.92
HyHDRNet [159]	44.64	0.9915	66.05
SV-HDR [162] [1]	37.30	0.9826	—
SSHDR [118]	41.97	0.9895	67.77

[1] Models that did not use Kalantari dataset.

4.4. Diffusion Models for HDR Reconstruction

Diffusion Models [28] constitute a class of probabilistic models employing a parameterized Markov chain to produce data samples matching the training set distribution after a finite time. DMs define a diffusion process in which Gaussian noise is gradually added to data until the original signal is destroyed. The model then learns to reverse that diffusion process, enabling the generation of samples resembling the training data.

Formally, the initial stage, called the forward diffusion process, takes a data sample x_0 and maps it through a series of intermediate latent variables $x_1, x_2, \ldots, x_T$ [164]. In this process, Gaussian noise ϵ is added step by step in each stage t following a variance schedule $\{\beta_t \in (0,1)\}_{t=1}^{T}$. The forward diffusion process is given by Equation (8).

$$q(x_t \mid x_{t-1}) = \mathcal{N}(x_t; \sqrt{1-\beta_t} \cdot x_{t-1}, \beta_t \cdot I) \tag{8}$$

This process results in a sequence of latent variables that become increasingly noisy until the original information is nearly completely lost in X_T.

The second stage, called the reverse diffusion process, involves learning a model that reverses this diffusion. The goal is to model the conditional probability of retrieving X_{T-1} from X_T by using Gaussian conditional distributions. Each transition in the reverse process is parameterized by a neural network estimating the mean $\mu_\theta(x_t, t)$ and variance β_t [165]. The reverse diffusion process is defined by Equation (9).

$$p_\theta(x_{t-1} \mid x_t) = \mathcal{N}(x_{t-1}; \mu_\theta(x_t, t), \beta_t \cdot I) \tag{9}$$

By iteratively applying these transitions, the model can reconstruct the original sample x_0 from noisy x_T. Throughout this process, the model learns to generate new samples consistent with the training data distribution [166].

DMs demonstrate substantial potential for HDR image reconstruction, especially for their ability to learn from noisy data and generate coherent samples. By handling uncertainties and variations in data, DMs stand out in scenarios where multiple LDR images are combined to produce a single HDR image. The iterative structure of DMs allows for continuous adjustment during reconstruction, refining pixel estimates and reducing unwanted artifacts (ghost or luminous artifacts). However, this same iterative feature, when used for large models to estimate entire images, becomes inefficient and impairs real-time usage.

This method can be implemented as single-frame, multi-frame, or image enhancement.

4.4.1. Diffusion Models for Single-Frame

Bemana et al. [167] proposed an approach to generate HDR images from a single LDR input image. This approach creates a sequence of LDR images, referred to as brackets, which, when merged, produce an HDR image. Initially, pretrained DMs are used to produce multiple exposures from a single LDR image. Next, an exposure consistency term couples the different generated exposures, ensuring that the LDR images remain consistent with one another. Additionally, the approach enables conditioning the HDR content generated in saturated regions using text conditioning and image/histogram guidance combinations, offering more control over image quality and the final tonal distribution of the HDR image.

Similarly, Dalal et al. [11] propose a Denoising Diffusion Probabilistic Model (DDPM) architecture using classifier-free guidance. The proposed architecture uses an autoencoder to produce a latent representation of the LDR input image for conditioning the HDR reconstruction. Furthermore, the authors introduce a hybrid loss function, weighting reconstruction, and perceptual and exposure losses to improve convergence of the proposed architecture.

On another note, Goswami et al. [168] proposed the Diffusion Inverse Tone Mapping (DITMO) approach, similarly aiming to generate HDR images from a single LDR. The proposed method uses a saturation mask to identify saturated areas and a semantic mask to classify objects in the image into different classes. With that information, the method applies a diffusion-based inpainting process that hallucinates details in saturated regions, employing a set of prompts associated with the classes identified by semantic segmentation. Finally, multiple versions of the image are generated at different exposures and merged to form the final HDR image.

4.4.2. Diffusion Models for Multi-Frame

The DiffHDR [19] architecture addresses generating HDR images from multiple LDR exposures, focusing on mitigating ghost artifacts resulting from scenes with large motion and saturation. HDR deghosting is treated as a conditional generative modeling task by adopting Diffusion Models, including a Feature Condition Generator (FCG) and a Domain Feature Align module, as shown in Figure 10. Furthermore, to mitigate the semantic confusion caused by saturation in LDR images, the authors introduced a Sliding-window Noise Estimator to smoothly sample noise in a patch-based approach. This method allows the model to focus on local contextual information, avoiding artifact introduction from ignoring correlation among adjacent patches.

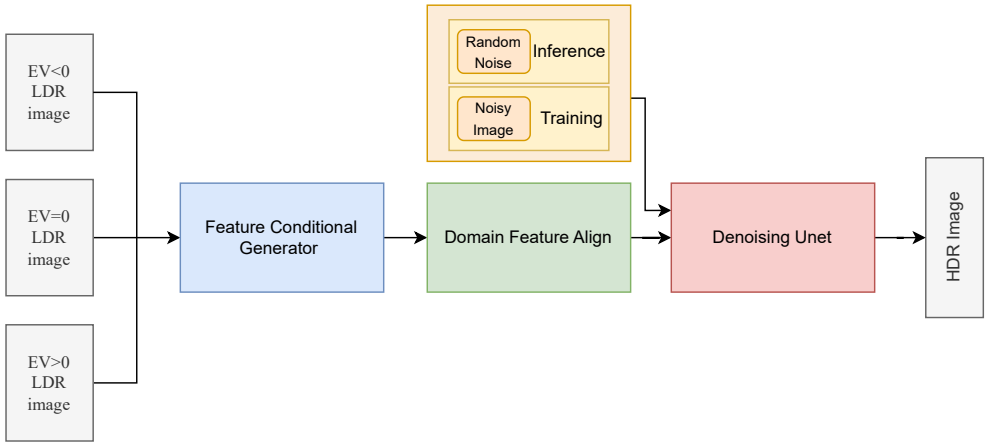


Figure 10. Simplified DiffHDR approach architecture.

Hu et al. [169] proposed the Low-Frequency Aware Diffusion Model (LF-Diff) approach to recover HDR images without ghosting artifacts from multiple LDR images at different exposures with significant motion. The LF-Diff architecture consists of an alignment module, a model extracting low-frequency information, and a refinement module combining this information with the output of a denoising UNet, which is a key component in DMs. The main distinction of LF-Diff lies in its usage of low-frequency information to guide HDR reconstruction.

4.4.3. Diffusion Models for Image Enhancement

Image enhancement is the process of improving image quality without losing information, involving modifications of one or more image attributes such as contrast, brightness, or noise. The LS-Sagiri architecture proposed by Li et al. [170] employs a two-module approach to enhance LDR images. In the first module, a Transformer-based model [18], Latent-SwinIR, that adjusts brightness and contrast, harmonizing color distribution. In the second module, the generative refinement module, a ControlNet-inspired [171] approach with an extra parallel encoder is used. This parallel encoder works alongside the Denoising UNet from RePaint [172], applying binary masks to highlight overexposed and

underexposed regions and merging the latent input features with the refined feature map. Additionally, the generative refinement module can function as a plug-and-play module, allowing improved LDR enhancement for existing models.

On the other hand, Wang et al. [173] proposed integrating a diffusion model with a physics-based exposure model for low-light image enhancement to predict a normally exposed image from an underexposed one. The authors noticed that starting the restoration directly from the underexposed noisy image rather than from purely Gaussian noise reduced network complexity and the number of inference steps required. Furthermore, this approach allowed for restoring images from any intermediate step of the diffusion process, reducing the need for repeated denoising. The article also introduced an adaptive residual layer enabling the proposed model to dynamically adjust its denoising strategies based on local image features, taking into account the signal-to-noise ratio.

Table 7 summarizes all the different approaches and metrics presented in this section.

Table 7. Summary of performance for Diffusion Models.

Metric	LF-Diff	DiffHDR	LS-Sagiri	Zero-Shot	Exp. Diff.	Cond. Diff.	DITMO
PSNR-l	44.76	44.11	—	—	—	16.97	—
SSIM-l	0.9919	0.9911	—	—	—	—	—
PSNR-μ	42.59	41.73	—	—	—	81	—
SSIM-μ	0.9906	0.9885	—	—	—	—	—
HDR VDP-2	66.54	65.52	—	—	—	—	—
HDR VDP-2.2	—	—	—	—	—	52.29	—
BRISQUE	—	—	19.72	—	—	—	—
NIQE	—	—	20.31	2.34	—	—	—
TMQI	—	—	—	0.8915	—	—	—
MANIQA	—	—	0.57	—	—	—	—
CLIP-IQA	—	—	0.67	—	—	—	—
NoR-VDP++	—	—	—	—	—	—	54.337
PU21-Pique	—	—	—	—	—	—	8.789

4.5. HDR Video Reconstruction

Some existing methods for direct HDR video capture rely on dedicated hardware solutions for that task, such as line-by-line exposure/ISO or internal/external beam splitters [174]. A more practical alternative for reconstructing HDR video is, just as with HDR image reconstruction, to use consecutive frames from an LDR video with alternating exposures to generate the frames for the HDR video. Classical approaches to accomplish this were initiated by Kang et al. [175], who introduced an algorithm to align neighboring frames with alternating exposures to a reference frame with an intermediate exposure. However, the motion among frames present in videos often leads to ghosting artifacts in this method. Therefore, Mangiat and Gibson [176] improved this approach by introducing block-based motion estimation and refining the result using color similarity and filtering. Kalantari et al. [177] introduced an optimization approach based on patches that synthesizes images at multiple exposures to reconstruct an HDR frame. These approaches usually require considerable processing time for the optimization and still show ghost artifacts from alignment inaccuracies.

With the progress in deep learning, various CNN-based HDR video reconstruction models have been introduced, representing a significant leap in the quality of resulting videos and mainly employing optical flow or attention blocks. Table 8 shows a quantitative

comparison of state-of-the-art HDR video reconstruction methods, including three optical flow-based models [30–32] and two attention-based models [33,34].

Table 8. Performance of HDR video reconstruction methods on the DeepHDRVideo dataset.

Methods	2-Exposure				3-Exposure			
	PSNR$_T$	**SSIM$_T$**	**HDR-VDP-2**	**Time (ms)**	**PSNR$_T$**	**SSIM$_T$**	**HDR-VDP-2**	**Time (ms)**
Kalantari19 [30]	39.91	0.9329	71.11	200	38.78	0.9331	65.73	220
Chen [31]	<u>42.48</u>	**0.9620**	<u>74.80</u>	522	39.44	<u>0.9569</u>	67.76	540
LAN-HDR [33]	41.59	0.9472	71.34	415	40.48	0.9504	<u>68.61</u>	525
HDRFlow [32]	**43.25**	0.9520	**77.29**	35	<u>40.56</u>	0.9535	**72.42**	50
HDR-V-Diff [34]	42.07	<u>0.9604</u>	70.88	–	**40.82**	**0.9581**	68.16	–

Bold: best performance; <u>Underlined</u>: second best.

4.5.1. Optical Flow-Based Approaches

Currently, state-of-the-art methods for HDR video reconstruction mostly comprise models relying on optical flow alignment. Kalantari et al. [30] introduced the first end-to-end model of this kind, consisting of a flow network to align the LDR frames and a weight network to merge them. Chen et al. [31] advanced alignment by incorporating deformable convolutions after optical flow alignment. To further reduce flow estimation errors in these methods, Xu et al. [32] proposed an approach with a more robust optical flow estimation, surpassing other SoTA methods in public benchmarks [31] while also being the first to reconstruct HDR video in real time. Figure 11 illustrates this model architecture, featuring an optical flow network with large multi-size convolution kernels, an HDR-domain alignment loss to supervise the flow network, and a sophisticated training scheme combining real video datasets (Vimeo-90K) and synthetic ones (Sintel).

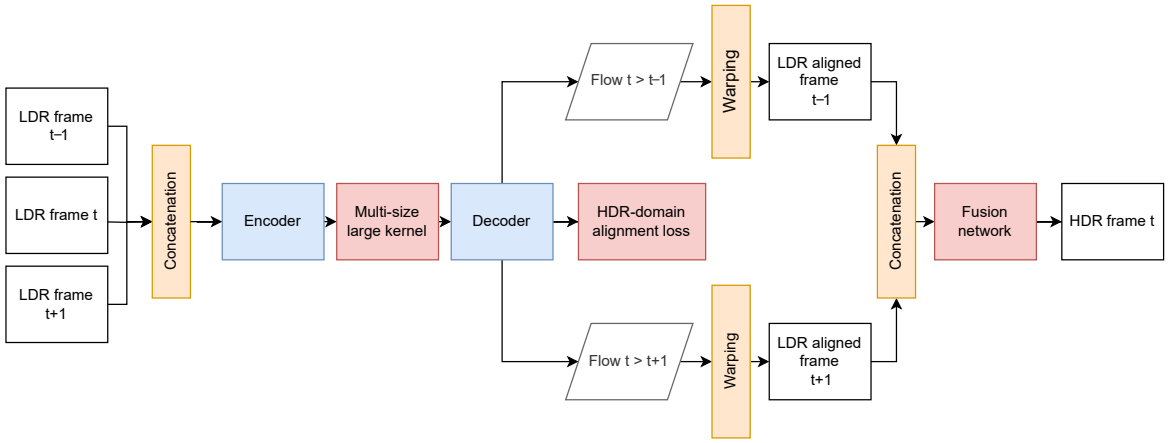


Figure 11. HDRFlow architecture.

4.5.2. Attention-Based Approaches

Seeking to tackle limitations from flow estimation errors, which cause ghosting in frames with complex motion or extreme saturation, some attention-based HDR video reconstruction methods have emerged, including SoTA approaches such as the network by Chung et al. [33] named LAN-HDR. This network aligns frames by computing a luminance-based attention score to register the motion between adjacent LDR frames and the reference frame. Guan et al. [34] recently proposed an HDR video reconstruction model using diffusion, which is illustrated in Figure 12. The architecture is composed of a latent HDR diffusion model that learns the prior distribution of individual HDR frames, a module for

temporally consistent alignment that learns multi-scale feature alignment among frames, and a reconstruction module that, with the help of a Zero-init Cross-Attention (ZiCA) block, effectively integrates the learned information.

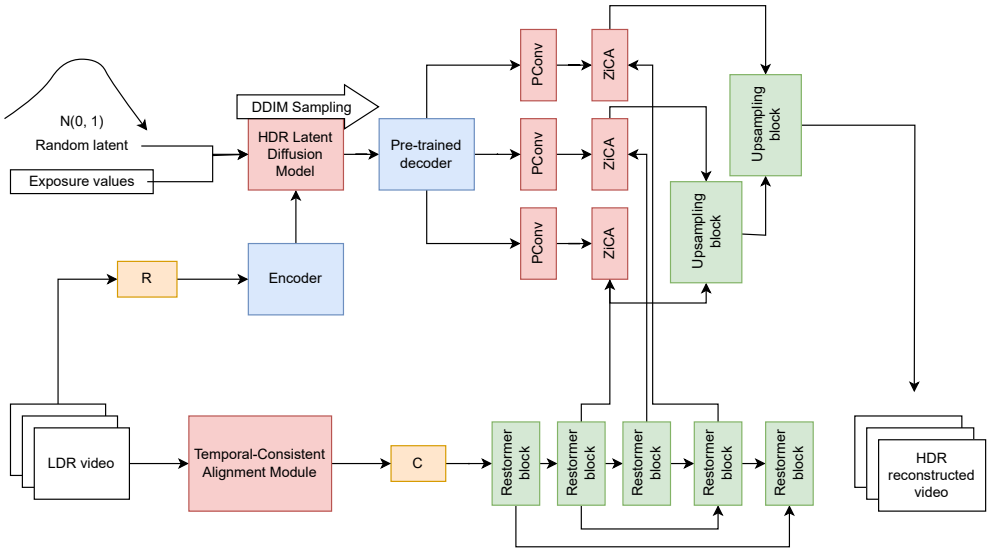


Figure 12. HDR-V-Diff architecture.

Table 9 summarizes the main advantages and disadvantages of HDR video reconstruction methods discussed in the previous subsections.

Table 9. HDR video reconstruction models, with their strengths and limitations.

Category	Model	Strengths	Limitations
Optical Flow-Based	HFT-HDR [31], KIM et al. [157], HDRFlow [32]	• Low inference time • Highly effective in static or less dynamic scenes • Patch-level alignment	• Susceptible to distortions due to inaccurate tracking under large movements • High computational cost
Attention-Based	LAN-HDR [33], HDR-V-Diff [34]	• Spatial and temporal attention • Considers global image context • Improved reconstruction in missing-content areas	• Higher resource usage • High memory consumption

5. Discussions and Future Perspectives

It is evident from the analysis of the current landscape that HDR reconstruction, whether from a single frame (single-frame) or multiple frames (multi-frame), has advanced significantly through innovations in areas such as network architectures (CNNs, GANs, Transformers, and Diffusion Models), improved evaluation metrics, and the creation of more robust datasets. This section discusses the main insights, current challenges, emerging strategies, and future perspectives guiding the field.

5.1. State of the Art

The state of the art in HDR reconstruction has progressed considerably, featuring methods ranging from classic approaches based on autoencoders and optical flow to more complex techniques combining attention mechanisms and GANs. Below is a review of the CNN-based models:

- CNN-Based Models: CNN-based approaches have long dominated the field and remain highly relevant. Works like HDRCNN [8] and variants such as HDRUnet,

HDRTV, and DCDR-UNet highlight the use of convolutional architectures, often complemented by skip connections and deformable modules, to recover lost details in overexposed and underexposed areas [41,45].

The HDRCNN architecture uses a loss function that separates illuminance and reflectance components, allowing for the adjustment of the relative importance of illuminance through a weighting parameter. Subsequent models, regardless of the employed architecture, frequently adopt variations of the L1 loss function, or even L1 itself, due to its proven effectiveness and consistent performance in image reconstruction tasks. The models we have examined are reviewed below:

- GAN-Based Models: The application of GANs, as seen in HDR-cGAN [138] and GlowGAN [10], shows potential for producing photorealistic images thanks to the adversarial generator's ability to simulate nuances of exposure and contrast.
  Typically, GAN models for HDR reconstruction analyzed in this review use a combination of loss functions. The classic adversarial loss (based on the minimax formulation or variations like LSGAN or WGAN) encourages the generator to produce visually plausible results that deceive the discriminator. However, to ensure fidelity to the input image content (LDR) and the reference (HDR ground truth, when available), this adversarial loss is often combined with pixelwise reconstruction losses (like L1, which is preferred over L2 to avoid excessive smoothing) and/or perceptual losses (like LPIPS), which measure similarity in the feature spaces of pretrained networks, better capturing the structural and textural similarity perceived by the human eye [26,138]. The relative weighting between these losses is a crucial hyperparameter for balancing realism and fidelity.
- Transformer-Based Models: Initially prominent in NLP, Transformers have been adapted for HDR image reconstruction in computer vision. Models like CA-ViT [17] and HDT-HDR [161] employ self-attention mechanisms to enhance alignment and fuse multiple exposures, achieving more accurate and detailed reconstructions. Typically, the loss functions used in Transformer-based models for HDR image reconstruction combine the classic loss function in the logarithmic domain with perceptual loss, resulting in the form $L_1 + \alpha L_p$ [17,27,157,159,162], aiming to optimize both objective fidelity and the perceived visual quality of the reconstruction. However, although there are proposals that employ loss functions based on Sobel filters aiming to maintain structural details, the results obtained so far have not proven promising [156].
- Diffusion Models (DMs): Recently, new diffusion-based architectures have emerged to address stability and training issues found in other generative methods. By iteratively adjusting pixel estimates, these models allow for the progressive refinement of reconstructed images. In general, Diffusion Models applied to HDR reconstruction use a loss function based on the L2 loss. The standard diffusion loss, different from the traditional approach that uses information from the input image and the neural network output, compares the noise added to the input data with the noise predicted by the network [3,25].

In HDR video, approaches based on optical flow and attention blocks have effectively reduced ghosting artifacts and maintained temporal consistency [30,33,34].

5.2. Current Challenges and Limitations

Despite notable advances, multiple challenges remain:

- Data Availability and Quality: The scarcity of extensive and varied datasets, especially those capturing dynamic scenes with complex motion and diverse lighting, hinders the training of robust models capable of generalizing to out-of-distribution (OOD) data [65,78].

- Computational Complexity and Real-Time HDR: Advanced architectures, particularly Transformer- and DM-based models, demand significant computational power and memory, potentially hindering their use in real-time scenarios or on resource-constrained devices like smartphones, drones, and AR/VR [10]. Achieving real-time or near real-time performance is a critical requirement for many practical applications, demanding efficient model designs and evaluation strategies that consider not only quality metrics but also inference speed, computational cost (e.g., FLOPS), and memory consumption. In this regard, some works have advanced toward low-cost efficient reconstruction like [32,178].
- Artifact Handling: Artifacts, including blur, ghosting, halos, noise, and misalignment, remain formidable obstacles. Although alignment and fusion techniques have improved, extreme movements and saturation can still produce visual imperfections in HDR images [12,63].
- Real-World Generalization: Models trained on synthetic or rigidly controlled datasets may fail to generalize to real-world scenarios where lighting and motion conditions are highly variable [84].
- Metric Adaptation and Correlation: Metrics originally designed for LDR must be revised or redesigned for HDR images, complicating the interpretation of results and the standardization of evaluations. Furthermore, ensuring that objective metrics accurately correlate with subjective human perception of HDR quality remains an ongoing challenge.

5.3. Ongoing Solutions and Advancements

Several emerging strategies show promise in mitigating current challenges:

- Hybrid Architectures: Combining CNNs, Transformers, and DMs leverages the strengths of each. Hybrid models like PASTA [50] integrate temporal and spatial attention mechanisms, enhancing alignment and multi-exposure fusion while reducing artifacts and improving visual quality.
- Data Augmentation Techniques: Methods like neural data augmentation [88] and HDR-LMDA [89] expand the diversity of datasets, increasing model robustness against variations in exposure and lighting. They simulate different capture scenarios, improving generalization capabilities.
- Self-Supervised and Unsupervised Models: Methods that reduce reliance on large annotated datasets are gaining attention. Self-supervised models can learn representations from unlabeled data, simplifying training in data-scarce environments [50]. Approaches based on Large Language Models (LLMs), such as LLM-HDR, explore self-supervised learning for unpaired LDR-to-HDR translation [179].
- Computational Optimization: Efforts to develop lighter and more efficient models, such as RepUNet [122], aim to reduce computational costs without substantially compromising HDR image quality. Techniques like pruning, quantization, and knowledge distillation also yield more compact models suitable for real-time applications. Hardware-based solutions, such as those using FPGAs and optimized lookup tables, also represent viable paths to achieve real-time HDR video processing [178].
- Integration of Multi-modal Data: Incorporating depth, polarization, event cameras, or other modalities can enhance HDR reconstruction by providing deeper scene perception and facilitating exposure alignment and fusion.
- Better Loss Functions: New hybrid loss functions considering perceptual and structural aspects of HDR images are being introduced to guide model training more effectively. Combining adversarial, perceptual, and reconstruction losses helps re-

duce artifacts [41]. Losses guided by semantic information or human knowledge (e.g., LLM-based loss) are also being explored [179].

5.4. Applications and Interdisciplinary Connections

Advances in HDR reconstruction have implications beyond image processing:

- Medicine: Images with higher exposure fidelity can reveal critical details for earlier and more accurate diagnoses, aiding medical imaging.
- VR/AR: HDR reconstruction algorithms can enhance immersion and realism in virtual environments, benefiting applications in entertainment, training, and simulation.
- Entertainment and Media: Televisions, monitors, and cameras using HDR technology provide richer visual experiences, increasing the impact of cinematic and advertising content.
- Autonomous Vehicles: Handling extreme lighting variations is crucial for vehicle vision systems, improving environmental interpretation under adverse conditions.

5.5. Trends and Future Ideas

Future directions in HDR reconstruction promise exciting avenues:

- Multi-modal Models: Integrating information from different sensor modalities, such as depth (from LiDAR/Stereo), polarization, or event cameras, can result in more robust and accurate HDR reconstructions. Models leveraging multiple input modalities could capture extra scene nuances, boosting HDR image quality, especially in complex dynamic scenarios.
- Self-Supervised and Unsupervised Learning: Techniques that do not require massive annotated datasets are attracting greater interest. Self-supervised methods can enable HDR reconstruction from unlabeled data, facilitating training in data-scarce settings [50]. Unsupervised approaches, including those using cycle consistency or domain adaptation, are particularly valuable for unpaired LDR–HDR datasets [179].
- Computational Optimization and Energy Efficiency: Developing lighter and more efficient models that maintain high HDR reconstruction quality while reducing resource usage is a growing research area. Pruning, quantization, and knowledge distillation will play a key role in creating models suitable for mobile and real-time applications [122]. The exploration of new and efficient architectures, such as SSMs, notably Mamba, which offer linear scaling and effective modeling of long-range dependencies, could provide significant efficiency gains for high-resolution images and video compared to traditional Transformers.
- Large Language/Vision Models (LLMs/LVMs): Integrating the perception and semantic understanding capabilities of LLMs and LVMs presents an innovative direction. These models can provide high-level guidance for reconstruction (e.g., identifying objects/regions, suggesting enhancement styles), enable zero-shot or few-shot adaptation to new scenarios, or serve as powerful priors or feature extractors within HDR pipelines, especially for human-knowledge-guided unpaired data reconstruction [179].
- Standardization and Evaluation Protocols: Establishing consistent standards for HDR data collection, labeling, and evaluation—both in terms of metrics (including their correlation with human perception and real-time suitability) and benchmark datasets covering diverse real-world scenarios—would facilitate fair comparisons and promote collaboration within the scientific community [94].
- Advanced Transformers and Attention Mechanisms: The ongoing adaptation and refinement of Transformers in vision tasks are poised to further enhance HDR reconstruction. More sophisticated attention mechanisms, considering relationships between multiple exposures and overall scene structure, may lead to more detailed reconstructions with fewer artifacts [17].

- Enhanced Diffusion Models: DMs are growing as a viable alternative for generating high-quality HDR images. Future studies will likely focus on reducing their computational complexity, making them more accessible for real-time use and deployment on resource-constrained devices [28,34].
- Adaptive and Controllable Reconstruction: Future models may offer more user control, allowing for adjustments to tone mapping style, artifact tolerance, or region-specific enhancement interactively, which could be possibly guided by semantic scene understanding.
- Vision Mamba: This is an approach that maintains the high performance characteristic of Transformers but with lower computational cost. Originally presented as Mamba, the technique was extended to computer vision tasks, resulting in the variant called Vision Mamba. This approach has demonstrated superior performance compared to Transformers in various tasks, such as detection, segmentation [180], and image restoration [181]. Future work may investigate the impact of Vision Mamba on the task of HDR image reconstruction, especially in the feature enhancement stage, positioning it as a promising candidate to replace Transformer-based architectures, which are more complex and costly.

6. Conclusions

This review provided an overview of HDR image reconstruction methods, spanning classical CNN approaches to more recent GAN, Transformer, and diffusion-based techniques. Beyond offering a taxonomy organizing these methods, we addressed their applications, limitations, and how emerging advancements in data, metrics, architectures, and strategies can expand the potential of each research line, bringing remarkable progress toward making HDR more accessible and computationally economical. In summary, HDR reconstruction continues to evolve rapidly, and future perspectives suggest convergent approaches, both fully supervised and auto/unsupervised, to achieve superior performance in real-world scenarios. It is expected that combining ideas like some of those discussed here, coupled with efforts for more efficient and standardized solutions, will accelerate the maturity of these technologies and open new horizons for HDR applications across increasingly diverse domains.

Author Contributions: Conceptualization, G.d.L.M., J.L.-C., J.M., Q.L., G.d.S.F., L.H.C.C., F.B.L., T.P. and A.B.A.; methodology, G.d.L.M., J.L.-C., J.M., Q.L., G.d.S.F., L.H.C.C., F.B.L., T.P. and A.B.A.; resources, A.B.A.; writing—original draft preparation, G.d.L.M., J.L.-C., J.M., Q.L., G.d.S.F., L.H.C.C., F.B.L., T.P. and A.B.A.; writing—review and editing G.d.L.M., T.P. and A.B.A.; supervision, A.B.A.; project administration,T.P. and A.B.A.; funding acquisition, A.B.A. All authors have read and agreed to the published version of the manuscript.

Funding: This research was financially supported by the PAVIC Laboratory, which benefited from SUFRAMA fiscal incentives under Brazilian Law No. 8387/1991.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Acknowledgments: The authors gratefully acknowledge support from the PAVIC Laboratory (Pesquisa Aplicada em Visão e Inteligência Computacional) at the University of Acre, Brazil.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Ferreti, G.d.S.; Paixão, T.; Alvarez, A.B. Generative Adversarial Network-Based Lightweight High-Dynamic-Range Image Reconstruction Model. *Appl. Sci.* **2025**, *15*, 4801. [CrossRef]
2. Lopez-Cabrejos, J.; Paixão, T.; Alvarez, A.B.; Luque, D.B. An Efficient and Low-Complexity Transformer-Based Deep Learning Framework for High-Dynamic-Range Image Reconstruction. *Sensors* **2025**, *25*, 1497. [CrossRef] [PubMed]
3. Fuest, M.; Ma, P.; Gui, M.; Fischer, J.S.; Hu, V.T.; Ommer, B. Diffusion Models and Representation Learning: A Survey. *arXiv* **2024**, arXiv:2407.00783.
4. Zhu, R.; Xu, S.; Liu, P.; Li, S.; Lu, Y.; Niu, D.; Liu, Z.; Meng, Z.; Li, Z.; Chen, X.; et al. Zero-Shot Structure-Preserving Diffusion Model for High Dynamic Range Tone Mapping. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 16–22 June 2024; pp. 26130–26139.
5. Beckmann, M.; Krahmer, F.; Bhandari, A. HDR tomography via modulo Radon transform. In Proceedings of the 2020 IEEE International Conference on Image Processing (ICIP), Abu Dhabi, United Arab Emirates, 25–28 October 2020; IEEE: New York, NY, USA, 2020; pp. 3025–3029.
6. Abebe, M. High Dynamic Range Imaging. In *Fundamentals and Applications of Colour Engineering*; Springer: Cham, Switzerland, 2023; pp. 293–310.
7. Takayanagi, I.; Kuroda, R. HDR CMOS image sensors for automotive applications. *IEEE Trans. Electron Devices* **2022**, *69*, 2815–2823. [CrossRef]
8. Eilertsen, G.; Kronander, J.; Denes, G.; Mantiuk, R.K.; Unger, J. HDR image reconstruction from a single exposure using deep CNNs. *ACM Trans. Graph.* **2017**, *36*, 178. [CrossRef]
9. Le, P.; Le, Q.; Nguyen, R.; Hua, B. Single-Image HDR Reconstruction by Multi-Exposure Generation. In Proceedings of the 2023 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV), Los Alamitos, CA, USA, 2–7 January 2023; pp. 4052–4061. [CrossRef]
10. Wang, C.; Serrano, A.; Pan, X.; Chen, B.; Myszkowski, K.; Seidel, H.P.; Theobalt, C.; Leimkühler, T. Glowgan: Unsupervised learning of hdr images from ldr images in the wild. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Paris, France, 2–6 October 2023; pp. 10509–10519.
11. Dalal, D.; Vashishtha, G.; Singh, P.; Raman, S. Single Image LDR to HDR Conversion using Conditional Diffusion. In Proceedings of the 2023 IEEE International Conference on Image Processing (ICIP), Kuala Lumpur, Malaysia, 8–11 October 2023.
12. Prabhakar, K.R.; Agrawal, S.; Singh, D.K.; Ashwath, B.; Babu, R.V. Towards practical and efficient high-resolution HDR deghosting with CNN. In Proceedings of the Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, 23–28 August 2020; Proceedings, Part XXI 16; Springer: Cham, Switzerland, 2020; pp. 497–513.
13. Yan, Q.; Gong, D.; Shi, Q.; Hengel, A.v.d.; Shen, C.; Reid, I.; Zhang, Y. Attention-guided network for ghost-free high dynamic range imaging. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 15–20 June 2019; pp. 1751–1760.
14. Tursun, O.T.; Akyüz, A.O.; Erdem, A.; Erdem, E. The state of the art in HDR deghosting: A survey and evaluation. In *Computer Graphics Forum*; Wiley Online Library: Hoboken, NJ, USA, 2015; Volume 34, pp. 683–707.
15. Kalantari, N.K.; Ramamoorthi, R. Deep high dynamic range imaging of dynamic scenes. *ACM Trans. Graph.* **2017**, *36*, 144. [CrossRef]
16. Liu, Z.; Lin, W.; Li, X.; Rao, Q.; Jiang, T.; Han, M.; Fan, H.; Sun, J.; Liu, S. ADNet: Attention-guided Deformable Convolutional Network for High Dynamic Range Imaging. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Nashville, TN, USA, 20–25 June 2021.
17. Liu, Z.; Wang, Y.; Zeng, B.; Liu, S. Ghost-free high dynamic range imaging with context-aware transformer. In Proceedings of the European Conference on Computer Vision, Tel Aviv, Israel, 23–27 October 2022; Springer: Cham, Switzerland, 2022; pp. 344–360.
18. Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A.N.; Kaiser, Ł.; Polosukhin, I. Attention is all you need. *arXiv* **2017**, arXiv:1706.03762.
19. Yan, Q.; Hu, T.; Sun, Y.; Tang, H.; Zhu, Y.; Dong, W.; Van Gool, L.; Zhang, Y. Towards high-quality hdr deghosting with conditional diffusion models. *IEEE Trans. Circuits Syst. Video Technol.* **2023**, *34*, 4011–4026. [CrossRef]
20. Kinoshita, Y.; Kiya, H. Checkerboard-Artifact-Free Image-Enhancement Network Considering Local and Global Features. In Proceedings of the 2020 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC), Auckland, New Zealand, 7–10 December 2020.
21. Santos, M.S.; Tsang, I.R.; Kalantari, N. Single image HDR reconstruction using a CNN with masked features and perceptual loss. *ACM Trans. Graph. (TOG)* **2020**, *39*, 80:1–80:10. [CrossRef]
22. Khan, Z.; Khanna, M.; Raman, S. FHDR: HDR Image Reconstruction from a Single LDR Image using Feedback Network. In Proceedings of the 2019 IEEE Global Conference on Signal and Information Processing (GlobalSIP), Ottawa, ON, Canada, 11–14 November 2019; pp. 1–5. [CrossRef]
23. Endo, Y.; Kanamori, Y.; Mitani, J. Deep Reverse Tone Mapping. *ACM Trans. Graph.* **2017**, *36*, 1–10. [CrossRef]

24. Chakraborty, T.; S, U.R.K.; Naik, S.M.; Panja, M.; Manvitha, B. Ten years of generative adversarial nets (GANs): A survey of the state-of-the-art. *Mach. Learn. Sci. Technol.* **2024**, *5*, 011001. [CrossRef]
25. Goodfellow, I.J.; Pouget-Abadie, J.; Mirza, M.; Xu, B.; Warde-Farley, D.; Ozair, S.; Courville, A.; Bengio, Y. Generative Adversarial Networks. *Commun. ACM* **2020**, *63*, 139–144. [CrossRef]
26. Niu, Y.; Wu, J.; Liu, W.; Guo, W.; Lau, R.W. Hdr-gan: Hdr image reconstruction from multi-exposed ldr images with large motions. *IEEE Trans. Image Process.* **2021**, *30*, 3885–3896. [CrossRef] [PubMed]
27. Tel, S.; Wu, Z.; Zhang, Y.; Heyrman, B.; Demonceaux, C.; Timofte, R.; Ginhac, D. Alignment-free HDR Deghosting with Semantics Consistent Transformer. *arXiv* **2023**, arXiv:2305.18135
28. Yang, L.; Zhang, Z.; Song, Y.; Hong, S.; Xu, R.; Zhao, Y.; Zhang, W.; Cui, B.; Yang, M.H. Diffusion models: A comprehensive survey of methods and applications. *ACM Comput. Surv.* **2023**, *56*, 1–39. [CrossRef]
29. Wang, L.; Yoon, K.J. Deep learning for hdr imaging: State-of-the-art and future trends. *IEEE Trans. Pattern Anal. Mach. Intell.* **2021**, *44*, 8874–8895. [CrossRef]
30. Kalantari, N.K.; Ramamoorthi, R. Deep HDR video from sequences with alternating exposures. In *Computer Graphics Forum*; Wiley: Hoboken, NJ, USA, 2019; pp. 193–205.
31. Chen, G.; Chen, C.; Guo, S.; Liang, Z.; Wong, K.Y.K.; Zhang, L. HDR Video Reconstruction: A Coarse-to-fine Network and A Real-world Benchmark Dataset. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Montreal, QC, Canada, 10–17 October 2021.
32. Xu, G.; Wang, Y.; Gu, J.; Xue, T.; Yang, X. HDRFlow: Real-Time HDR Video Reconstruction with Large Motions. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 16–22 June 2024; pp. 24851–24860.
33. Chung, H.; Cho, N.I. LAN-HDR: Luminance-based Alignment Network for High Dynamic Range Video Reconstruction. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Paris, France, 2–6 October 2023.
34. Guan, Y.; Xu, R.; Yao, M.; Gao, R.; Wang, L.; Xiong, Z. Diffusion-Promoted HDR Video Reconstruction. *arXiv* **2024**, arXiv:2406.08204.
35. Mantiuk, R.; Kim, K.J.; Rempel, A.G.; Heidrich, W. HDR-VDP-2: A calibrated visual metric for visibility and quality predictions in all luminance conditions. *ACM Trans. Graph. (TOG)* **2011**, *30*, 1–14. [CrossRef]
36. Mantiuk, R.K.; Hammou, D.; Hanji, P. HDR-VDP-3: A multi-metric for predicting image differences, quality and contrast distortions in high dynamic range and regular content. *arXiv* **2023**, arXiv:2304.13625.
37. Azimi, M.; Mantiuk, R.K. PU21: A novel perceptually uniform encoding for adapting existing quality metrics for HDR. In Proceedings of the 2021 Picture Coding Symposium (PCS), Bristol, UK, 29 June–2 July 2021; IEEE: New York, NY, USA, 2021; pp. 1–5. [CrossRef]
38. Liu, Y.; Tian, Y.; Wang, S.; Zhang, X.; Kwong, S. Overview of High-Dynamic-Range Image Quality Assessment. *J. Imaging* **2024**, *10*, 243. [CrossRef]
39. Marnerides, D.; Bashford-Rogers, T.; Hatchett, J.; Debattista, K. ExpandNet: A Deep Convolutional Neural Network for High Dynamic Range Expansion from Low Dynamic Range Content. *Comput. Graph. Forum* **2018**, *37*, 37–49. [CrossRef]
40. Liu, Y.L.; Lai, W.S.; Chen, Y.S.; Kao, Y.L.; Yang, M.H.; Chuang, Y.Y.; Huang, J.B. Single-Image HDR Reconstruction by Learning to Reverse the Camera Pipeline. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 13–19 June 2020.
41. Chen, X.; Liu, Y.; Zhang, Z.; Qiao, Y.; Dong, C. HDRUNet: Single Image HDR Reconstruction With Denoising and Dequantization. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops, Seattle, WA, USA, 17–18 June 2021; pp. 354–363. [CrossRef]
42. Yao, Z.; Bi, J.; Deng, W.; He, W.; Wang, Z.; Kuang, X.; Zhou, M.; Gao, Q.; Tong, T. DEUNet: Dual-encoder UNet for simultaneous denoising and reconstruction of single HDR image. *Comput. Graph.* **2024**, *119*, 103882. [CrossRef]
43. Banterle, F.; Ledda, P.; Debattista, K.; Chalmers, A. Inverse tone mapping. In Proceedings of the GRAPHITE '06: Proceedings of the 4th International Conference on Computer Graphics and Interactive Techniques in Australasia and Southeast Asia, Kuala Lumpur, Malaysia, 29 November–2 December 2005; pp. 349–356. [CrossRef]
44. Ronneberger, O.; Fischer, P.; Brox, T. U-Net: Convolutional Networks for Biomedical Image Segmentation. In Proceedings of the 18th International Conference, Munich, Germany, 5–9 October 2015.
45. Kim, J.; Zhu, Z.; Bau, T.; Liu, C. DCDR-UNet: Deformable Convolution Based Detail Restoration via U-shape Network for Single Image HDR Reconstruction. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops, Seattle, WA, USA, 17–18 June 2024; pp. 5909–5918. [CrossRef]
46. Li, Y.; Peng, X.; Zhang, J.; Li, Z.; Wen, M. DCT-GAN: Dilated convolutional transformer-based GAN for time series anomaly detection. *IEEE Trans. Knowl. Data Eng.* **2021**, *35*, 3632–3644. [CrossRef]
47. Kronander, J.; Gustavson, S.; Bonnet, G.; Ynnerman, A.; Unger, J. A unified framework for multi-sensor HDR video reconstruction. *Signal Process. Image Commun.* **2014**, *29*, 203–215. [CrossRef]

48. Prabhakar, K.R.; Arora, R.; Swaminathan, A.; Singh, K.P.; Babu, R.V. A fast, scalable, and reliable deghosting method for extreme exposure fusion. In Proceedings of the 2019 IEEE International Conference on Computational Photography (ICCP), Tokyo, Japan, 15–17 May 2019; IEEE: New York, NY, USA, 2019; pp. 1–8.
49. Lu, Z.; Li, J.; Liu, H.; Huang, C.; Zhang, L.; Zeng, T. Transformer for single image super-resolution. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, New Orleans, LA, USA, 18–24 June 2022; pp. 457–466.
50. Liu, X.; Li, A.; Wu, Z.; Du, Y.; Zhang, L.; Zhang, Y.; Timofte, R.; Zhu, C. PASTA: Towards Flexible and Efficient HDR Imaging Via Progressively Aggregated Spatio-Temporal Aligment. *arXiv* **2024**, arXiv:2403.10376.
51. Dhariwal, P.; Nichol, A. Diffusion models beat gans on image synthesis. *Adv. Neural Inf. Process. Syst.* **2021**, *34*, 8780–8794. [CrossRef]
52. Hasinoff, S.W.; Sharlet, D.; Geiss, R.; Adams, A.; Barron, J.T.; Kainz, F.; Chen, J.; Levoy, M. Burst photography for high dynamic range and low-light imaging on mobile cameras. *ACM Trans. Graph. (TOG)* **2016**, *35*, 1–12. [CrossRef]
53. Wronski, B.; Garcia-Dorado, I.; Ernst, M.; Kelly, D.; Krainin, M.; Liang, C.K.; Levoy, M.; Milanfar, P. Handheld multi-frame super-resolution. *ACM Trans. Graph. (TOG)* **2019**, *38*, 1–18. [CrossRef]
54. Karadeniz, A.S.; Erdem, E.; Erdem, A. Burst Photography for Learning to Enhance Extremely Dark Images. *IEEE Trans. Image Process.* **2020**, *30*, 9372–9385. [CrossRef]
55. Debevec, P.E.; Malik, J., Recovering High Dynamic Range Radiance Maps from Photographs. In *Seminal Graphics Papers: Pushing the Boundaries, Volume 2*, 1st ed.; Association for Computing Machinery: New York, NY, USA, 2023. [CrossRef]
56. Xu, F.; Liu, J.; Song, Y.; Sun, H.; Wang, X. Multi-exposure image fusion techniques: A comprehensive review. *Remote Sens.* **2022**, *14*, 771. [CrossRef]
57. Ou, Y.; Ambalathankandy, P.; Ikebe, M.; Takamaeda, S.; Motomura, M.; Asai, T. Real-time tone mapping: A state of the art report. *arXiv* **2020**, arXiv:2003.03074. [CrossRef]
58. Han, X.; Khan, I.R.; Rahardja, S. High dynamic range image tone mapping: Literature review and performance benchmark. *Digit. Signal Process.* **2023**, *137*, 104015. [CrossRef]
59. Ward, G. A contrast-based scalefactor for luminance display. In *Graphics Gems IV*; Academic Press Professional Inc: New York, NY, USA, 1994 ; pp. 415–421. [CrossRef] [PubMed]
60. Land, E.H.; McCann, J.J. Lightness and retinex theory. *J. Opt. Soc. Am.* **1971**, *61*, 1–11. [CrossRef]
61. Pattanaik, S.N.; Ferwerda, J.A.; Fairchild, M.D.; Greenberg, D.P. A multiscale model of adaptation and spatial vision for realistic image display. In Proceedings of the 25th Annual Conference on Computer Graphics and Interactive Techniques, Orlando, FL, USA, 19–24 July 1998; pp. 287–298.
62. Reinhard, E.; Stark, M.; Shirley, P.; Ferwerda, J. Photographic tone reproduction for digital images. In *Seminal Graphics Papers: Pushing the Boundaries, Volume 2*; Association for Computing Machinery: New York, NY, USA, 2023; pp. 661–670.
63. Hu, J.; Gallo, O.; Pulli, K.; Sun, X. HDR deghosting: How to deal with saturation? In Proceedings of the IEEE conference on computer vision and pattern recognition, Portland, OR, USA, 23–28 June 2013; pp. 1163–1170. [CrossRef]
64. Yoon, H.; Uddin, S.M.N.; Jung, Y.J. Multi-Scale Attention-Guided Non-Local Network for HDR Image Reconstruction. *Sensors* **2022**, *22*, 7044. [CrossRef]
65. Sen, P.; Kalantari, N.K.; Yaesoubi, M.; Darabi, S.; Goldman, D.B.; Shechtman, E. Robust patch-based hdr reconstruction of dynamic scenes. *ACM Trans. Graph.* **2012**, *31*, 203-1. [CrossRef]
66. Zheng, M.; Zhi, K.; Zeng, J.; Tian, C.; You, L. A hybrid CNN for image denoising. *J. Artif. Intell. Technol.* **2022**, *2*, 93–99. [CrossRef]
67. Tian, C.; Xu, Y.; Fei, L.; Wang, J.; Wen, J.; Luo, N. Enhanced CNN for image denoising. *CAAI Trans. Intell. Technol.* **2019**, *4*, 17–23. [CrossRef]
68. Zhang, K.; Zuo, W.; Chen, Y.; Meng, D.; Zhang, L. Beyond a gaussian denoiser: Residual learning of deep cnn for image denoising. *IEEE Trans. Image Process.* **2017**, *26*, 3142–3155. [CrossRef]
69. Mao, X.; Shen, C.; Yang, Y.B. Image restoration using very deep convolutional encoder-decoder networks with symmetric skip connections. In Proceedings of the NIPS'16: Proceedings of the 30th International Conference on Neural Information Processing Systems, Barcelona, Spain, 5–10 December 2016. [CrossRef]
70. Zhang, K.; Zuo, W.; Zhang, L. FFDNet: Toward a fast and flexible solution for CNN-based image denoising. *IEEE Trans. Image Process.* **2018**, *27*, 4608–4622. [CrossRef]
71. He, J.; Dong, C.; Qiao, Y. Modulating image restoration with continual levels via adaptive feature modification layers. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 15–20 June 2019; pp. 11056–11064. [CrossRef]
72. Odena, A.; Dumoulin, V.; Olah, C. Deconvolution and Checkerboard Artifacts. *Distill* **2016**, *1*, e3. [CrossRef]
73. Cao, G.; Zhou, F.; Liu, K.; Wang, A.; Fan, L. A Decoupled Kernel Prediction Network Guided by Soft Mask for Single Image HDR Reconstruction. *ACM Trans. Multimed. Comput. Commun. Appl.* **2022**, *19*, 1–23. [CrossRef]
74. Dumoulin, V.; Belghazi, I.; Poole, B.; Mastropietro, O.; Lamb, A.; Arjovsky, M.; Courville, A. Adversarially Learned Inference. *arXiv* **2016**, arXiv:1606.00704. [CrossRef]

75. Alzubaidi, L.; Bai, J.; Al-Sabaawi, A.; Santamaría, J.I.; Albahri, A.S.; Al-dabbagh, B.S.N.; Fadhel, M.A.; Manoufali, M.; Zhang, J.; Al-timemy, A.H.; et al. A survey on deep learning tools dealing with data scarcity: Definitions, challenges, solutions, tips, and applications. *J. Big Data* **2023**, *10*, 1–82. [CrossRef]
76. Prabhakar, K.R.; Babu, R.V. High Dynamic Range Deghosting Dataset. 2020. Available online: https://val.cds.iisc.ac.in/HDR/HDRD/ (accessed on 12 December 2024). [CrossRef]
77. Funt, B.V.; Shi, L. The Rehabilitation of MaxRGB. In Proceedings of the International Conference on Communications in Computing, Cape Town, South Africa, 23–27 May 2010.
78. Tursun, O.T.; Akyüz, A.O.; Erdem, A.; Erdem, E. An Objective Deghosting Quality Metric for HDR Images. *Comput. Graph. Forum* **2016**, *35*, 139–152. [CrossRef]
79. Nemoto, H.; Korshunov, P.; Hanhart, P.; Ebrahimi, T. Visual attention in LDR and HDR images. In Proceedings of the 9th International Workshop on Video Processing and Quality Metrics for Consumer Electronics (VPQM), Chandler, AZ, USA, 5–6 February 2015.
80. Gardner, M.A.; Sunkavalli, K.; Yumer, E.; Shen, X.; Gambaretto, E.; Gagné, C.; Lalonde, J.F. Learning to Predict Indoor Illumination from a Single Image. *arXiv* **2017**, arXiv:1704.00090. [CrossRef]
81. Chen, X.; Li, Z.; Zhang, Z.; Ren, J.S.; Liu, Y.; He, J.; Qiao, Y.; Zhou, J.; Dong, C. Towards Efficient SDRTV-to-HDRTV by Learning from Image Formation. *arXiv* **2023**, arXiv:2309.04084.
82. Hanji, P.; Mantiuk, R.; Eilertsen, G.; Hajisharif, S.; Unger, J. Comparison of single image HDR reconstruction methods—The caveats of quality assessment. In Proceedings of the SIGGRAPH '22: ACM SIGGRAPH 2022 Conference Proceedings, Vancouver, BC, Canada, 7–11 August 2022. [CrossRef]
83. Perez-Pellitero, E.; Catley-Chandar, S.; Leonardis, A.; Timofte, R.; Wang, X.; Li, Y.; Wang, T.; Song, F.; Liu, Z.; Lin, W.; et al. NTIRE 2021 Challenge on High Dynamic Range Imaging: Dataset, Methods and Results. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops, Nashville, TN, USA, 19–25 June 2021; pp. 410–425. [CrossRef]
84. Barua, H.B.; Stefanov, K.; Wong, K.; Dhall, A.; Krishnasamy, G. GTA-HDR: A Large-Scale Synthetic Dataset for HDR Image Reconstruction. *arXiv* **2024**, arXiv:2403.17837.
85. Zhang, Q.; Li, Q.; Yu, G.; Sun, L.; Zhou, M.; Chu, J. A Multidimensional Choledoch Database and Benchmarks for Cholangiocarcinoma Diagnosis. *IEEE Access* **2019**, *7*, 149414–149421. [CrossRef]
86. Liu, S.; Zhang, X.; Sun, L.; Liang, Z.; Zeng, H.; Zhang, L. Joint hdr denoising and fusion: A real-world mobile hdr image dataset. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Vancouver, BC, Canada, 17–24 June 2023; pp. 13966–13975. [CrossRef]
87. Mikhailiuk, A.; Perez-Ortiz, M.; Yue, D.; Suen, W.; Mantiuk, R.K. Consolidated dataset and metrics for high-dynamic-range image quality. *IEEE Trans. Multimed.* **2021**, *24*, 2125–2138. [CrossRef]
88. Zheng, C.; Ying, W.; Wu, S.; Li, Z. Neural Augmentation-Based Saturation Restoration for LDR Images of HDR Scenes. *IEEE Trans. Instrum. Meas.* **2023**, *72*, 4506011. [CrossRef]
89. Zhao, F.; Liu, Q.; Ikenaga, T. HDR-LMDA: A Local Area-Based Mixed Data Augmentation Method for Hdr Video Reconstruction. In Proceedings of the 2023 IEEE International Conference on Image Processing (ICIP), Kuala Lumpur, Malaysia, 8–11 October 2023; pp. 2020–2024. [PubMed]
90. Richter, T. On the standardization of the JPEG XT image compression. In Proceedings of the 2013 Picture Coding Symposium (PCS), San Jose, CA, USA, 8–11 December 2013; IEEE: New York, NY, USA, 2013; pp. 37–40.
91. Mai, Z.; Mansour, H.; Mantiuk, R.; Nasiopoulos, P.; Ward, R.; Heidrich, W. Optimizing a tone curve for backward-compatible high dynamic range image and video compression. *IEEE Trans. Image Process.* **2010**, *20*, 1558–1571. [CrossRef]
92. Valenzise, G.; De Simone, F.; Lauga, P.; Dufaux, F. Performance evaluation of objective quality metrics for HDR image compression. In Proceedings of the Applications of Digital Image Processing XXXVII, San Diego, CA, USA, 18–21 August 2014; SPIE: Bellingham, WA, USA, 2014; Volume 9217, pp. 78–87. [CrossRef]
93. Zerman, E.; Valenzise, G.; Dufaux, F. An extensive performance evaluation of full-reference HDR image quality metrics. *Qual. User Exp.* **2017**, *2*, 1–16. [CrossRef]
94. Hanhart, P.; Bernardo, M.V.; Pereira, M.G.; Pinheiro, A.M.; Ebrahimi, T. Benchmarking of objective quality metrics for HDR image quality assessment. *EURASIP J. Image Video Process.* **2015**, *2015*, 1–18. [CrossRef]
95. Mantiuk, R.K.; Denes, G.; Chapiro, A.; Kaplanyan, A.; Rufo, G.; Bachy, R.; Lian, T.; Patney, A. FovVideoVDP: A visible difference predictor for wide field-of-view video. *ACM Trans. Graph.* **2021**, *40*, 1–19. [CrossRef]
96. Narwaria, M.; Da Silva, M.P.; Le Callet, P. HDR-VQM: An objective quality measure for high dynamic range video. *Signal Process. Image Commun.* **2015**, *35*, 46–60. [CrossRef]
97. Mittal, A.; Moorthy, A.K.; Bovik, A.C. No-reference image quality assessment in the spatial domain. *IEEE Trans. Image Process.* **2012**, *21*, 4695–4708. [CrossRef]
98. Mittal, A.; Soundararajan, R.; Bovik, A.C. Making a "completely blind" image quality analyzer. *IEEE Signal Process. Lett.* **2012**, *20*, 209–212. [CrossRef]

99. Mantiuk, R.; Daly, S.; Kerofsky, L. Display adaptive tone mapping. In *ACM SIGGRAPH 2008 Papers*; ACM: New York, NY, USA, 2008; pp. 1–10.
100. Talebi, H.; Milanfar, P. NIMA: Neural image assessment. *IEEE Trans. Image Process.* **2018**, *27*, 3998–4011. [CrossRef]
101. Zhang, R.; Isola, P.; Efros, A.A.; Shechtman, E.; Wang, O. The unreasonable effectiveness of deep features as a perceptual metric. In Proceedings of the IEEE conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 586–595.
102. Wang, Z.; Bovik, A.C.; Sheikh, H.R.; Simoncelli, E.P. Image quality assessment: From error visibility to structural similarity. *IEEE Trans. Image Process.* **2004**, *13*, 600–612. [CrossRef]
103. Wang, Z.; Simoncelli, E.P.; Bovik, A.C. Multiscale structural similarity for image quality assessment. In Proceedings of the Thrity-Seventh Asilomar Conference on Signals, Systems & Computers, Pacific Grove, CA, USA, 9–12 November 2003; IEEE: New York, NY, USA, 2003; Volume 2, pp. 1398–1402. [CrossRef] [PubMed]
104. Zhang, L.; Zhang, L.; Mou, X.; Zhang, D. FSIM: A feature similarity index for image quality assessment. *IEEE Trans. Image Process.* **2011**, *20*, 2378–2386. [CrossRef]
105. Zhang, L.; Shen, Y.; Li, H. VSI: A visual saliency-induced index for perceptual image quality assessment. *IEEE Trans. Image Process.* **2014**, *23*, 4270–4281. [CrossRef]
106. Mantiuk, R.K.; Hanji, P.; Ashraf, M.; Asano, Y.; Chapiro, A. ColorVideoVDP: A visual difference predictor for image, video and display distortions. *arXiv* **2024**, arXiv:2401.11485.
107. Ke, J.; Wang, Q.; Wang, Y.; Milanfar, P.; Yang, F. MUSIQ: Multi-Scale Image Quality Transformer. In Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), Montreal, QC, Canada, 10–17 October 2021; pp. 5148–5157.
108. Mantiuk, R.; Myszkowski, K.; Seidel, H.P. Visible difference predicator for high dynamic range images. In Proceedings of the 2004 IEEE International Conference on Systems, Man and Cybernetics (IEEE Cat. No. 04CH37583), The Hague, The Netherlands, 10–13 October 2004; IEEE: New York, NY, USA, 2004; Volume 3, pp. 2763–2769.
109. Daly, S.J. Visible differences predictor: An algorithm for the assessment of image fidelity. In *Human Vision, Visual Processing, and Digital Display III*; SPIE: Bellingham, WA, USA, 1992.
110. Mantiuk, R.; Daly, S.J.; Myszkowski, K.; Seidel, H.P. Predicting visible differences in high dynamic range images: Model and its calibration. In Proceedings of the Human Vision and Electronic Imaging X, San Jose, CA, USA, 17 January 2005; SPIE: Bellingham, WA, USA, 2005; Volume 5666, pp. 204–214.
111. Korshunov, P.; Hanhart, P.; Richter, T.; Artusi, A.; Mantiuk, R.; Ebrahimi, T. Subjective quality assessment database of HDR images compressed with JPEG XT. In Proceedings of the 2015 Seventh International Workshop on Quality of Multimedia Experience (QoMEX), Costa Navarino, Greece, 26–29 May 2015; IEEE: New York, NY, USA, 2015; pp. 1–6.
112. O'Shea, K. An introduction to convolutional neural networks. *arXiv* **2015**, arXiv:1511.08458.
113. Chen, X.; Zhang, Z.; Ren, J.S.; Tian, L.; Qiao, Y.; Dong, C. A new journey from SDRTV to HDRTV. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Montreal, BC, Canada, 11–17 October 2021; pp. 4500–4509.
114. Lee, S.; Hwan An, G.; Kang, S.J. Deep recursive hdri: Inverse tone mapping using generative adversarial networks. In Proceedings of the European Conference on Computer Vision (ECCV), Munich, Germany, 8–14 September 2018; pp. 596–611. [CrossRef]
115. Chen, S.K.; Yen, H.L.; Liu, Y.L.; Chen, M.H.; Hu, H.N.; Peng, W.H.; Lin, Y.Y. CEVR: Learning Continuous Exposure Value Representations for Single-Image HDR Reconstruction. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Paris, France, 2–6 October 2023.
116. Kim, J.H.; Lee, S.; Kang, S.J. End-to-End Differentiable Learning to HDR Image Synthesis for Multi-exposure Images. In Proceedings of the AAAI Conference on Artificial Intelligence, Virtual, 2–9 February 2021.
117. Chen, J.; Yang, Z.; Chan, T.N.; Li, H.; Hou, J.; Chau, L.P. Attention-guided progressive neural texture fusion for high dynamic range image restoration. *IEEE Trans. Image Process.* **2022**, *31*, 2661–2672. [CrossRef]
118. Yan, Q.; Zhang, S.; Chen, W.; Tang, H.; Zhu, Y.; Sun, J.; Van Gool, L.; Zhang, Y. Smae: Few-shot learning for hdr deghosting with saturation-aware masked autoencoders. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Vancouver, BC, Canada, 17–24 June 2023; pp. 5775–5784.
119. Liu, C. Beyond Pixels: Exploring New Representations and Applications for Motion Analysis. Ph.D. Thesis, Massachusetts Institute of Technology, Cambridge, MA, USA, 2009.
120. Wu, S.; Xu, J.; Tai, Y.W.; Tang, C.K. Deep High Dynamic Range Imaging with Large Foreground Motions. In Proceedings of the European Conference on Computer Vision (ECCV), Munich, Germany, 8–14 September 2018. [CrossRef]
121. Lin, W.; Liu, Z.; Jiang, C.; Han, M.; Jiang, T.; Liu, S. Improving Bracket Image Restoration and Enhancement with Flow-guided Alignment and Enhanced Feature Aggregation. *arXiv* **2024**, arXiv:2404.10358. [CrossRef]
122. Yang, Q.; Liu, Y.; Chen, Q.; Yue, H.; Li, K.; Yang, J. Efficient HDR Reconstruction from Real-World Raw Images. *arXiv* **2024**, arXiv:2306.10311.
123. Li, H.; Yang, Z.; Zhang, Y.; Tao, D.; Yu, Z. Single-Image HDR Reconstruction Assisted Ghost Suppression and Detail Preservation Network for Multi-Exposure HDR Imaging. *IEEE Trans. Comput. Imaging* **2024**, *10*, 429–445. [CrossRef]

124. Barua, H.B.; Krishnasamy, G.; Wong, K.; Stefanov, K.; Dhall, A. ArtHDR-Net: Perceptually Realistic and Accurate HDR Content Creation. In Proceedings of the 2023 Asia Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC), Taipei, China, 31 October–3 November 2023; pp. 806–812. [CrossRef]
125. Gui, J.; Sun, Z.; Wen, Y.; Tao, D.; Ye, J. A Review on Generative Adversarial Networks: Algorithms, Theory, and Applications. *IEEE Trans. Knowl. Data Eng.* **2023**, *35*, 3313–3332. [CrossRef]
126. Karras, T.; Laine, S.; Aila, T. A style-based generator architecture for generative adversarial networks. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, Long Beach, CA, USA, 15–20 June 2019; pp. 4401–4410.
127. Ledig, C.; Theis, L.; Huszar, F.; Caballero, J.; Cunningham, A.; Acosta, A.; Aitken, A.; Tejani, A.; Totz, J.; Wang, Z.; et al. Photo-Realistic Single Image Super-Resolution Using a Generative Adversarial Network. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017.
128. Wang, Y.; Bilinski, P.; Bremond, F.; Dantcheva, A. Imaginator: Conditional spatio-temporal gan for video generation. In Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, Snowmass Village, CO, USA, 1–5 March 2020; pp. 1160–1169.
129. Frid-Adar, M.; Klang, E.; Amitai, M.; Goldberger, J.; Greenspan, H. Synthetic data augmentation using GAN for improved liver lesion classification. In Proceedings of the 2018 IEEE 15th international symposium on biomedical imaging (ISBI 2018), Washington, DC, USA, 4–7 April 2018; IEEE: New York, NY, USA, 2018; pp. 289–293. [CrossRef]
130. Zhu, J.Y.; Park, T.; Isola, P.; Efros, A.A. Unpaired image-to-image translation using cycle-consistent adversarial networks. In Proceedings of the IEEE International Conference on Computer Vision, Venice, Italy, 22–29 October 2017; pp. 2223–2232.
131. Zhang, Y.; Gan, Z.; Carin, L. Generating text via adversarial training. *NIPS Workshop Advers. Train.* **2016**, *21*, 21–32.
132. Guibas, J.T.; Virdi, T.S.; Li, P.S. Synthetic medical images from dual generative adversarial networks. *arXiv* **2017**, arXiv:1709.01872. [CrossRef]
133. Wang, T.; Trugman, D.; Lin, Y. SeismoGen: Seismic waveform synthesis using GAN with application to seismic data augmentation. *J. Geophys. Res. Solid Earth* **2021**, *126*, e2020JB020077. [CrossRef]
134. Pan, X.; You, Y.; Wang, Z.; Lu, C. Virtual to real reinforcement learning for autonomous driving. *arXiv* **2017**, arXiv:1704.03952.
135. Dam, T.; Ferdaus, M.M.; Pratama, M.; Anavatti, S.G.; Jayavelu, S.; Abbass, H. Latent preserving generative adversarial network for imbalance classification. In Proceedings of the 2022 IEEE International Conference on Image Processing (ICIP), Bordeaux, France, 16–19 October 2022; IEEE: New York, NY, USA, 2022; pp. 3712–3716.
136. Gonog, L.; Zhou, Y. A Review: Generative Adversarial Networks. In Proceedings of the 2019 14th IEEE Conference on Industrial Electronics and Applications (ICIEA), Xi'an, China, 19–21 June 2019; pp. 505–510. [CrossRef]
137. Wang, L.; Cho, W.; Yoon, K.J. Deceiving image-to-image translation networks for autonomous driving with adversarial perturbations. *IEEE Robot. Autom. Lett.* **2020**, *5*, 1421–1428. [CrossRef]
138. Raipurkar, P.; Pal, R.; Raman, S. HDR-cGAN: Single LDR to HDR Image Translation using Conditional GAN. *arXiv* **2021**, arXiv:2110.01660.
139. Mirza, M.; Osindero, S. Conditional Generative Adversarial Nets. *arXiv* **2014**, arXiv:1411.1784. [CrossRef]
140. Jain, D.; Raman, S. Deep over and under exposed region detection. In Proceedings of the Computer Vision and Image Processing: 5th International Conference, CVIP 2020, Prayagraj, India, 4–6 December 2020; Revised Selected Papers, Part III 5; Springer: Cham, Switzerland, 2021; pp. 34–45. [CrossRef]
141. Isola, P.; Zhu, J.; Zhou, T.; Efros, A.A. Image-to-Image Translation with Conditional Adversarial Networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017.
142. Sauer, A.; Schwarz, K.; Geiger, A. Stylegan-xl: Scaling stylegan to large diverse datasets. In Proceedings of the ACM SIGGRAPH 2022 Conference Proceedings, Vancouver, BC, Canada, 7–11 August 2022; pp. 1–10.
143. Li, R.; Wang, C.; Wang, J.; Liu, G.; Zhang, H.Y.; Zeng, B.; Liu, S. UPHDR-GAN: Generative Adversarial Network for High Dynamic Range Imaging With Unpaired Data. *IEEE Trans. Circuits Syst. Video Technol.* **2022**, *32*, 7532–7546. [CrossRef]
144. Xu, H.; Ma, J.; Zhang, X.P. MEF-GAN: Multi-Exposure Image Fusion via Generative Adversarial Networks. *IEEE Trans. Image Process.* **2020**, *29*, 7203–7216. [CrossRef]
145. Wang, X.; Girshick, R.B.; Gupta, A.; He, K. Non-local Neural Networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018.
146. Dosovitskiy, A.; Beyer, L.; Kolesnikov, A.; Weissenborn, D.; Zhai, X.; Unterthiner, T.; Dehghani, M.; Minderer, M.; Heigold, G.; Gelly, S.; et al. An image is worth 16x16 words: Transformers for image recognition at scale. *arXiv* **2020**, arXiv:2010.11929.
147. Cai, Y.; Bian, H.; Lin, J.; Wang, H.; Timofte, R.; Zhang, Y. Retinexformer: One-stage retinex-based transformer for low-light image enhancement. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Paris, France, 2–3 October 2023; pp. 12504–12513.
148. Wang, T.; Zhang, K.; Shen, T.; Luo, W.; Stenger, B.; Lu, T. Ultra-high-definition low-light image enhancement: A benchmark and transformer-based method. In Proceedings of the AAAI Conference on Artificial Intelligence, Washington, DC, USA, 7–14 February 2023; Volume 37, pp. 2654–2662.

149. Chang, H.; Zhang, H.; Jiang, L.; Liu, C.; Freeman, W.T. Maskgit: Masked generative image transformer. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, New Orleans, LA, USA, 18–24 June 2022; pp. 11315–11325.
150. Hudson, D.A.; Zitnick, L. Generative adversarial transformers. In *International Conference on Machine Learning*; PMLR: New York, NY, USA, 2021; pp. 4487–4499.
151. Chen, X.; Wang, X.; Zhou, J.; Qiao, Y.; Dong, C. Activating more pixels in image super-resolution transformer. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Vancouver, BC, Canada, 17–24 June 2023; pp. 22367–22377.
152. Han, K.; Xiao, A.; Wu, E.; Guo, J.; Xu, C.; Wang, Y. Transformer in transformer. *Adv. Neural Inf. Process. Syst.* **2021**, *34*, 15908–15919.
153. Wu, B.; Xu, C.; Dai, X.; Wan, A.; Zhang, P.; Yan, Z.; Tomizuka, M.; Gonzalez, J.; Keutzer, K.; Vajda, P. Visual transformers: Token-based image representation and processing for computer vision. *arXiv* **2020**, arXiv:2006.03677.
154. Cheng, B.; Misra, I.; Schwing, A.G.; Kirillov, A.; Girdhar, R. Masked-attention mask transformer for universal image segmentation. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, New Orleans, LA, USA, 18–24 June 2022; pp. 1290–1299.
155. Baktash, J.A.; Dawodi, M. Gpt-4: A review on advancements and opportunities in natural language processing. *arXiv* **2023**, arXiv:2305.03195.
156. Chen, R.; Zheng, B.; Zhang, H.; Chen, Q.; Yan, C.; Slabaugh, G.; Yuan, S. Improving dynamic hdr imaging with fusion transformer. In Proceedings of the AAAI Conference on Artificial Intelligence, Washington, DC, USA, 7–14 February 2023; Volume 37, pp. 340–349.
157. Kim, J.; Kim, M.H. Joint demosaicing and deghosting of time-varying exposures for single-shot hdr imaging. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Paris, France, 2–6 October 2023; pp. 12292–12301.
158. Suda, T.; Tanaka, M.; Monno, Y.; Okutomi, M. Deep snapshot hdr imaging using multi-exposure color filter array. In Proceedings of the Asian Conference on Computer Vision, Kyoto, Japan, 30 November–4 December 2020.
159. Yan, Q.; Chen, W.; Zhang, S.; Zhu, Y.; Sun, J.; Zhang, Y. A unified HDR imaging method with pixel and patch level. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Vancouver, BC, Canada, 17–24 June 2023; pp. 22211–22220.
160. Liang, J.; Cao, J.; Sun, G.; Zhang, K.; Van Gool, L.; Timofte, R. Swinir: Image restoration using swin transformer. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Montreal, BC, Canada, 11–17 October 2021; pp. 1833–1844.
161. Zhou, F.; Fu, Z.; Zhang, D. High dynamic range imaging with context-aware transformer. In Proceedings of the 2023 International Joint Conference on Neural Networks (IJCNN), Gold Coast, Australia, 18–23 June 2023; IEEE: New York, NY, USA, 2023; pp. 1–8.
162. Chi, Y.; Zhang, X.; Chan, S.H. Hdr imaging with spatially varying signal-to-noise ratios. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Vancouver, BC, Canada, 17–24 June 2023; pp. 5724–5734.
163. Liu, Z.; Lin, Y.; Cao, Y.; Hu, H.; Wei, Y.; Zhang, Z.; Lin, S.; Guo, B. Swin transformer: Hierarchical vision transformer using shifted windows. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Montreal, BC, Canada, 11–17 October 2021; pp. 10012–10022.
164. Prince, S.J. *Understanding Deep Learning*; MIT Press: Cambridge, MA, USA, 2023.
165. Nichol, A.Q.; Dhariwal, P. Improved denoising diffusion probabilistic models. In *International Conference on Machine Learning*; PMLR: New York, NY, USA, 2021; pp. 8162–8171.
166. Song, Y.; Dhariwal, P.; Chen, M.; Sutskever, I. Consistency models. *arXiv* **2023**, arXiv:2303.01469.
167. Bemana, M.; Leimkühler, T.; Myszkowski, K.; Seidel, H.P.; Ritschel, T. Exposure Diffusion: HDR Image Generation by Consistent LDR denoising. *arXiv* **2024**, arXiv:2405.14304.
168. Goswami, A.; Singh, A.R.; Banterle, F.; Debattista, K.; Bashford-Rogers, T. Semantic Aware Diffusion Inverse Tone Mapping. *arXiv* **2024**, arXiv:2405.15468.
169. Hu, T.; Yan, Q.; Qi, Y.; Zhang, Y. Generating content for hdr deghosting from frequency view. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 16–22 June 2024; pp. 25732–25741.
170. Li, B.; Ma, S.; Zeng, Y.; Xu, X.; Fang, Y.; Zhang, Z.; Wang, J.; Chen, K. Sagiri: Low Dynamic Range Image Enhancement with Generative Diffusion Prior. *arXiv* **2024**, arXiv:2406.09389. [CrossRef]
171. Zhang, L.; Rao, A.; Agrawala, M. Adding conditional control to text-to-image diffusion models. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Paris, France, 2–6 October 2023; pp. 3836–3847. [CrossRef]
172. Lugmayr, A.; Danelljan, M.; Romero, A.; Yu, F.; Timofte, R.; Van Gool, L. Repaint: Inpainting using denoising diffusion probabilistic models. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, New Orleans, LA, USA, 18–24 June 2022; pp. 11461–11471. [CrossRef]
173. Wang, Y.; Yu, Y.; Yang, W.; Guo, L.; Chau, L.P.; Kot, A.C.; Wen, B. Exposurediffusion: Learning to expose for low-light image enhancement. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Paris, France, 2–6 October 2023; pp. 12438–12448. [CrossRef]
174. Heide, F.; Steinberger, M.; Tsai, Y.T.; Rouf, M.; Pajak, D.; Reddy, D.; Gallo, O.; Liu, J.; Heidrich, W.; Egiazarian, K.; et al. FlexISP: A flexible camera image processing framework. *ACM Trans. Graph. (TOG)* **2014**, *33*, 1–13. [CrossRef]

175. Kang, S.B.; Uyttendaele, M.; Winder, S.; Szeliski, R. High dynamic range video. *ACM Trans. Graph. (TOG)* **2003**, *22*, 319–325. [CrossRef]
176. Mangiat, S.; Gibson, J. High dynamic range video with ghost removal. In Proceedings of the Applications of Digital Image Processing XXXIII, San Diego, CA, USA, 7–10 August 2010. [CrossRef]
177. Kalantari, N.K.; Shechtman, E.; Barnes, C.; Darabi, S.; Goldman, D.B.; Sen, P. Patch-based high dynamic range video. *ACM Trans. Graph. (TOG)* **2013**, *32*, 202:1–202:8. [CrossRef]
178. Mann, S.; Lo, R.C.H.; Ovtcharov, K.; Gu, S.; Dai, D.; Ngan, C.; Ai, T. Realtime HDR (High Dynamic Range) video for eyetap wearable computers, FPGA-based seeing aids, and glasseyes (EyeTaps). In Proceedings of the 2012 25th IEEE Canadian Conference on Electrical and Computer Engineering (CCECE), Montreal, QC, Canada, 29 April–2 May 2012; pp. 1–6. [CrossRef]
179. Barua, H.; Stefanov, K.; Che, L.; Dhall, A.; Wong, K.; Krishnasamy, G. LLM-HDR: Bridging LLM-based Perception and Self-Supervision for Unpaired LDR-to-HDR Image Reconstruction. *arXiv* **2025**, arXiv:2410.15068.
180. Zhu, L.; Liao, B.; Zhang, Q.; Wang, X.; Liu, W.; Wang, X. Vision Mamba: Efficient Visual Representation Learning with Bidirectional State Space Model. *arXiv* **2024**, arXiv:2401.09417.
181. Guo, H.; Guo, Y.; Zha, Y.; Zhang, Y.; Li, W.; Dai, T.; Xia, S.T.; Li, Y. MambaIRv2: Attentive State Space Restoration. *arXiv* **2024**, arXiv:2411.15269.