HQGS: HIGH-QUALITY NOVEL VIEW SYNTHESIS WITH GAUSSIAN SPLATTING IN DEGRADED SCENES

Anonymous authors
Paper under double-blind review

ABSTRACT

3D Gaussian Splatting (3DGS) has shown promising results for Novel View Synthesis. However, while it is quite effective when based on high-quality images, its performance declines as image quality degrades, due to lack of resolution, motion blur, noise, compression artifacts, or other factors common in real-world data collection. While some solutions have been proposed for specific types of degradation, general techniques are still missing. To address the problem, we propose a robust HQGS that significantly enhances the 3DGS under various degradation scenarios. We first analyze that 3DGS lacks sufficient attention in some detailed regions in low-quality scenes, leading to the absence of Gaussian primitives in those areas and resulting in loss of detail in the rendered images. To address this issue, we focus on leveraging edge structural information to provide additional guidance for 3DGS, enhancing its robustness. First, we introduce an edge-semantic fusion guidance module that combines rich texture information from high-frequency edge-aware maps with semantic information from images. The fused features serve as prior guidance to capture detailed distribution across different regions, bringing more attention to areas with detailed edge information and allowing for a higher concentration of Gaussian primitives to be assigned to such areas. Additionally, we present a structural cosine similarity loss to complement pixel-level constraints, further improving the quality of the rendered images. Extensive experiments demonstrate that our method offers better robustness and achieves the best results across various degraded scenes. The source code and trained models will be made available to the public.

1 INTRODUCTION

Novel view synthesis has significantly advanced in recent years, with the introduction of Neural Radiance Fields (NeRF) and 3D Gaussian Splatting (3DGS), benefiting applications such as augmented reality (AR) and virtual reality (VR) (Bian et al., 2016; Dawood, 2009; Farshid et al., 2018; Fassi et al., 2016). Nevertheless, existing methods are mainly designed for high-quality images captured with precise camera parameters. When faced with images of low resolution, with motion blur, compression artifacts, noise, or other degradation common in real-world imaging, they often struggle. Some NeRF (Bahat et al., 2022; Ma et al., 2022; Zhou et al., 2023b; Pearl et al., 2022) and 3DGS (Feng et al., 2024) models attempt to address the problem by incorporating various strategies or constraints, such as degradation kernels or super-sampling networks. However, these methods are tailored for specific types of degradation. For example, as shown in Figure 1, NeRFlix and SRGS, designed to address blur and low-resolution scenes, respectively, fail to handle each other's scenarios effectively. Hence, there is a need for methods that can achieve high-quality novel view synthesis across various types of degradation.

We consider the problem in the context of 3D Gaussian Splatting and start by showing that image degradation can severely compromise the two key stages, reconstruction and rendering, of NVS. For reconstruction, our preliminary experiments (Figure 2(b)) show that, under all these degraded conditions, the distribution of recovered Gaussian primitives becomes too sparse to allow the capture of fine scene details, especially for small objects. For rendering, these experiments confirm the findings of (Dong et al., 2023; Lin et al., 2023b), showing that the low-frequency information derived from the global scene structure is essential for image quality enhancement. This motivates us to propose a robust HQGS framework that leverages structural information resilient to low-level

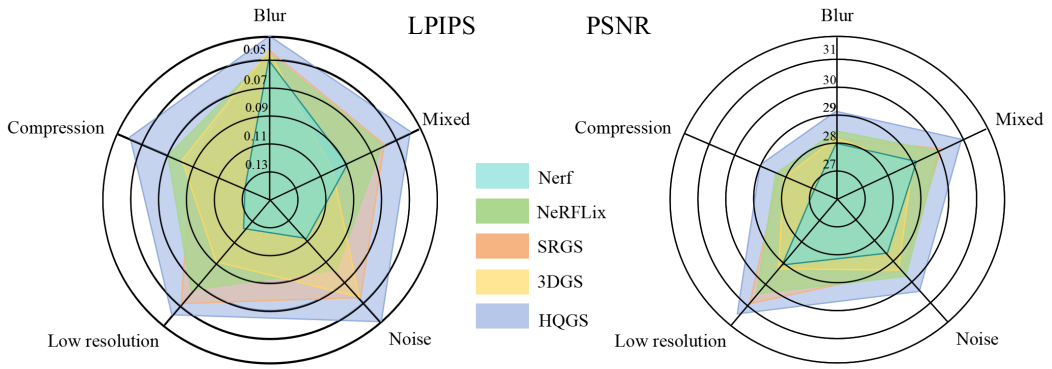


Figure 1: Results of exiting methods (Mildenhall et al., 2021; Kerbl et al., 2023; Zhou et al., 2023b; Feng et al., 2024) and our HQGS on five degradation scenes. HQGS performs well against others in these different degradation types.

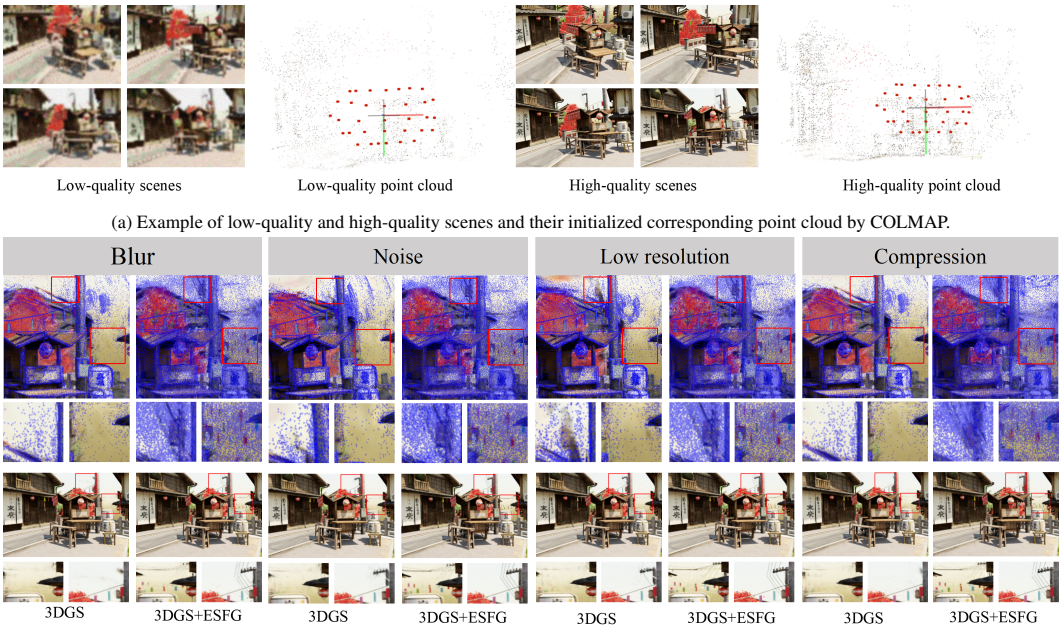
degradation types, allowing for better focus on high-frequency edge areas while aiding in the recovery of low-frequency information during the reconstruction and rendering stages, respectively. First, we introduce an Edge-Semantic Fusion Guidance (ESFG) module that utilizes combined semantic-aware and edge-aware features as a reliable indicator for robust Gaussian Splatting. This ESFG module has two branches: the upper one uses a cross-attention mechanism between high-frequency edge maps and low-quality images to learn the semantic feature with edge information, and the lower one understands high-frequency edge information further. Additionally, we design an efficient global Structural Cosine Similarity Loss ($\mathcal{L}_{SCS}$) to constrain global low-frequency structure information between rendered images and their corresponding target images. It provides the model with additional constraints, encouraging it to improve the quality on these areas. Extensive experimental results in various degradation scenarios (low resolution, JPEG compression, noise, blur, and mixed degradation) have demonstrated the superiority of our method, which has significant advantages over other 3DGS-based methods when trained within the same time. Notably, HQGS performs robustly on more highly degraded images.

The main contributions of our work are:

- We propose a general framework HQGS, designed to render high-quality novel view images under various degradation conditions, such as low resolution, noise, blur, JPEG compression, and mixed degradation.
- We introduce an edge-semantic fusion guidance (ESFG) module, which utilizes combined edge contour prior information and semantic information to guide the model's sensitivity to positional information and improve the generation of fine details. Furthermore, the structural cosine similarity loss ($\mathcal{L}_{SCS}$) is presented to constrain the consistency of global low-frequency structures, thereby enhancing the quality of rendered images.
- Comprehensive experiments demonstrate that the proposed method exhibits better robustness and achieves favorable performance against state-of-the-art approaches.

2 Related Works

Novel View Synthesis. Mildenhall et al. (Mildenhall et al., 2021) introduce the neural radiance field (NeRF) to implicitly represent static 3D scenes and synthesize novel views from multiple images with known poses. Building on this foundation, numerous NeRF-based models (Xu et al., 2022; Deng et al., 2022; Chen et al., 2022; Fridovich-Keil et al., 2022; Garbin et al., 2021; Reiser et al., 2021; Chen et al., 2023) have been developed. Point-NeRF (Xu et al., 2022) and DS-NeRF (Deng et al., 2022) integrate sparse 3D point cloud and depth information to resolve the geometric ambiguities inherent in NeRFs, leading to more accurate 3D point sampling and improved rendering quality. On the other hand, Plenoxels (Fridovich-Keil et al., 2022), TensoRF (Chen et al., 2022), FastNeRF (Garbin et al., 2021), KiloNeRF (Reiser et al., 2021), and MobileNeRF (Chen et al., 2023) focus on employing various advanced techniques to accelerate the training and inference processes.



(a) Example of low-quality and high-quality scenes and their initialized corresponding point cloud by COLMAP.

(b) Comparison of Gaussian primitives and rendered images of 3DGS with proposed edge-semantic fusion guidance (ESFG) module under various degraded scenes.

Figure 2: (a) Comparison of low-quality and high-quality images and corresponded point clouds obtained from COLMAP, which is used for Gaussian primitives initializing. (b) The visualization of the distribution of Gaussian primitives in a trained 3DGS under multiple degradation scenes, along with the rendered 2D images. Due to our edge-semantic fusion guidance module providing priors for detailed regions, it is able to distribute more Gaussian primitives in some detailed areas. As a result, the rendered novel view images contain richer details, particularly in elements like power lines and colorful flags.

Most recently, 3D-GS (Kerbl et al., 2023), based on point cloud rendering, facilitates real-time novel view synthesis. While these methods have made notable progress in rendering high-quality scenes, they can still produce artifacts with low-quality images and imprecise camera poses.

Novel View Synthesis in Degraded Scenes. Several NeRF- and 3DGS-based approaches (Huang et al., 2022; Bahat et al., 2022; Feng et al., 2024; Pearl et al., 2022; Zhou et al., 2023b) synthesizing high-quality novel views from degraded scenes using paired clean and degraded images. NeRF-SR (Wang et al., 2022) enhances output resolution through sub-pixel sampling, though this approach demands more computational resources and extends training times. On the other hand, NVSR (Bahat et al., 2022) trains a NeRF super-resolution (SR) network with multi-view data, leveraging the triplane structure to perform SR on low-resolution planes, thereby improving the overall NeRF resolution. In (Huang et al., 2022), RefSR-NeRF introduces a specialized SR module that incorporates high-resolution reference images, which can lead to longer training and inference times. To achieve high-quality rendering results in the noisy scene, NaN (Pearl et al., 2022) adapts the feed-forward IBRNet view synthesis method to achieve competitive burst denoising results. Instead of focusing on 3D learning, NeRFLiX (Zhou et al., 2023b) and NeRFLiX++ (Zhou et al., 2023a) learn a general 2D viewpoint mixer via simulated image degradation. However, if the distribution of the rendering artifacts shifts from the simulated data, the performance degrades. Drantal-NeRF (Yang et al., 2024) employs a diffusion-based image quality enhancement model to create higher-quality image pairs for training a NeRF model. Recently, SRGS (Feng et al., 2024) has emerged as a 3DGS-based model capable of achieving super-resolution rendering using paired low- and high-quality images.

Image Quality Enhancement and Restoration. Image restoration aims to enhance the quality of degraded images affected by various types and levels of degradation. This challenging task encompasses denoising (Lin et al., 2023b; Ren et al., 2021; 2022), deblurring (Fang et al., 2023; Sun et al., 2023; Pan et al., 2023), and general restoration (Zamir et al., 2021; 2022a; Lin et al., 2023a; 2024).

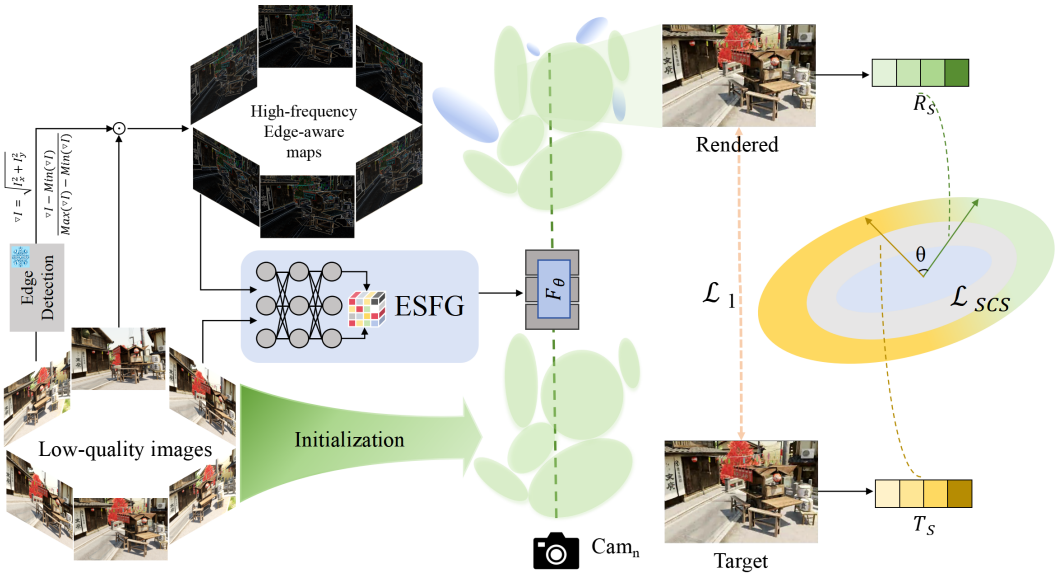


Figure 3: HQGS framewok. The high-frequency edge-aware maps and low-quality images are jointly fed into the learnable ESFG module for feature modulation and fusion. The fused features guide the training process of HQGS, providing reliable detail priors. During training, we employ pixel-level $\mathcal{L}_1$ loss and a global structural cosine similarity loss $\mathcal{L}_{SCS}$ to optimize the model.

Restormer (Zamir et al., 2022a) incorporates transformers for low-level restoration to balance the performance and computational costs, and MIRNetv2 (Zamir et al., 2022b) restores images through a novel feature extraction network. On the other hand, Kong et al. (Kong et al., 2023) present a Transformer-based method for high-quality image deblurring, by utilizing frequency-domain properties to simplify scaled dot product attention and alleviate the need for complex matrix multiplication. Recently, MRLPFNet (Dong et al., 2023) proposes a simple yet efficient low-pass filtering network, and demonstrate the importance of repairing low-frequency regions to improve image quality. SCPGabNet (Lin et al., 2023b) and SCP^{2}GAN (Lin et al., 2024) are non-paired unsupervised restoration methods that exploit the importance of low-frequency constraints and introduce background consistency loss (BGM) specifically for low-frequency areas.

3 Proposed Method

Figure 3 shows the overall pipeline of the proposed HQGS that receives multi-view low-quality images $I \in \mathcal{R}^{N\times H\times W\times 3}$ as the input. The N, H and W are the number, height and width of images. Using these images, we first generate the initial point cloud and corresponding camera views through the classical COLMAP method. Meanwhile, we detect the boundaries of each image to construct the high-frequency edge-aware maps. Next, we train a structure-assisted 3DGS model to leverage structural information in both the reconstruction and rendering stages. On the one hand, we propose an edge-semantic fusion guidance (ESFG) module that utilizes high-frequency edge-aware maps as a condition signal for reconstructing 3DGS. On the other hand, we utilize structural cosine similarity loss ($\mathcal{L}_{SCS}$) to constrain the global structure when rendering Gaussian primitives into an image within a view.

In the following subsections, we first introduce the various degradation types designed in our method. Then, we will elaborate on the ESFG module and $\mathcal{L}_{SCS}$, for robust HQGS.

3.1 Degradation Types

To simulate various adverse real-world conditions, we introduce five types of degradation:

- *Low Resolution:* 4× downsampling is applied to create low-resolution images.
- *JPEG Compression:* Images are compressed using JPEG with setting the quality level 10.

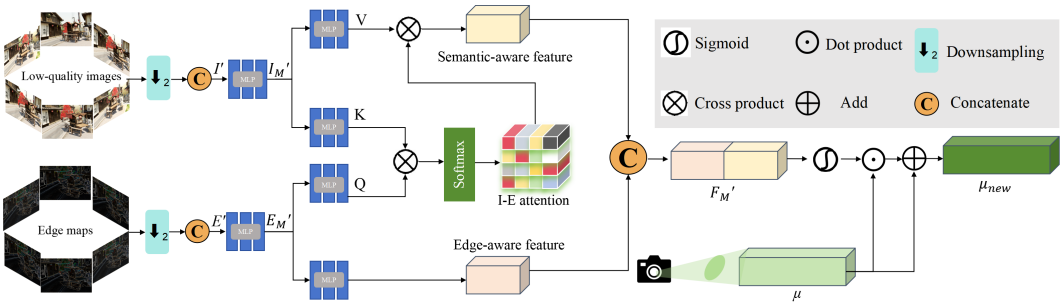


Figure 4: The ESFG module separately learns semantic-aware feature and edge-aware feature, and then jointly guides the training of HQGS.

- *Blur:* Blurred images are generated using a Gaussian blur kernel with a radius ranging from 10 to 20, and the blur angle is randomly selected within the range [0, 2π].
- *Noise:* Gaussian noise with a standard deviation 10 is added to the images.
- *Mixed Degradation:* A combination of the degradations above, applied in the following order: low resolution, JPEG compression, blurring, and noise.

Following 3DGS (Kerbl et al., 2023), we input a set of low-quality scene images with the corresponding cameras calibrated by COLMAP, producing a sparse point cloud as a side effect. Figure 2(a) compares low-quality and high-quality scenes with their corresponding point clouds. The low-quality images provide less visual information, creating a sparser point cloud than the high-quality scene.

3.2 Edge-Semantic Fusion Guidance Module

Motivation. Since degradation affects the edge details of some objects in the scene, making them difficult to capture. Figure 2(b) shows Gaussian primitives distributed in these areas, such as power lines and colorful flags, become sparse. This causes these objects to be omitted in the rendered 2D images, reducing the quality of the rendered novel views. To address this problem, we leverage specific edge contour priors as a 'reminder' to inform the model that particular objects in these regions must be generated. This encourages it to cover these areas with more Gaussian primitives during reconstruction.

Solution. We propose the Edge-Semantic Fusion Guidance (ESFG) module to enhance semantics from low-quality images by the high-frequency edge-aware maps. To extract global high-frequency edge-aware maps (E), we use the Sobel operator to calculate the gradient maps $\nabla I \in \mathcal{R}^{N\times H\times W\times 3}$ of the low-quality images I that contain critical edge information. We then normalize ∇I to the gradient masks $\nabla I'$ along with height and width dimensions:

$$\nabla I' = \frac{\nabla I - Min(\nabla I)}{Max(\nabla I) - Min(\nabla I)}, \tag{1}$$

where $Max(\cdot)$ and $Min(\cdot)$ is the maximum and minimum value in ∇I. Lastly, we obtain the high-frequency edge-aware map E by

$$E = \nabla I' \odot I. \tag{2}$$

where $\odot$ represents matrix multiplication. The $E \in \mathcal{R}^{N\times H\times W\times 3}$ highlights rich high-frequency image textures and details crucial for image quality enhancement and novel view synthesis. The difference between $\nabla I'$ and E is presented in the appendix.

In the ESFG module, we down-sample them to I' and E' by $2\times$ as the input for computational efficiency where I' and $E' \in \mathcal{R}^{N\times\frac{H}{2}\times\frac{W}{2}\times 3}$. Then, we concatenate the images of I' and E' according to the camera order, respectively. To match the position parameters $\boldsymbol{\mu} \in \mathcal{R}^{M\times 3}$ of the Gaussian primitives, we perform a scale transformation to obtain I'_M and $E'_M \in \mathcal{R}^{\frac{M}{2}\times 3}$ using MLPs, where M represents the number of Gaussian primitives.

The ESFG module contains two branches in Figure 4. In the upper branch, we employ a cross-attention mechanism where E'_M acts as the query, and the I'_M serves as the key and value. This

cross-attention between the low-quality images and the high-frequency edge-aware maps connects local high-frequency details with global semantics, producing more comprehensive guided features, i.e., semantic-aware features. In the lower branch, a learnable MLP layer is used to better interpret edge information, resulting in the edge-aware feature. The output features from both branches are then concatenated to form the final fused feature F'_M, followed by a sigmoid non-linear operation. Finally, F'_M is used to modulate the original position parameters $\boldsymbol{\mu} \in \mathbb{R}^3$ of the Gaussian primitives to obtain new position parameters μ_{new}, and then HQGS models it as $G(x)$:

$$\mu_{new} = Sigmoid(F'_M) \odot \mu + \mu, \tag{3}$$

$$G(x) = e^{(-\frac{1}{2}(x-\mu_{new})^T \Sigma^{-1}(x-\mu_{new}))}. \tag{4}$$

where $\odot$ represents matrix multiplication, and $\boldsymbol{\Sigma} \in \mathbb{R}^{3\times3}$ is an anisotropic covariance matrix, which is factorized into a scaling matrix $\boldsymbol{S}$ and a rotation matrix $\boldsymbol{R}$ as $\boldsymbol{\Sigma} = \boldsymbol{R}\boldsymbol{S}\boldsymbol{S}^\top\boldsymbol{R}^\top$.

3.3 Structural Cosine Similarity Loss

While high-frequency information is essential for enhancing image quality, low-frequency information should not be overlooked, as it corresponds to smooth areas and colors closely related to global structural information (Dong et al., 2023; Lin et al., 2023b).

We propose a Structural Cosine Similarity Loss ($\mathcal{L}_{SCS}$) to enhance the global structure of low-frequency regions and improve overall rendering quality. This loss emphasizes directional consistency in the low-frequency feature space, allowing it to better capture the image's global structure and thereby achieve a closer match to its overall characteristics.

In the optimization stage, besides the original $\mathcal{L}_1$ between the rendered image (R) and its corresponding target image (T). We add our $\mathcal{L}_{SCS}$ as a low-frequency-aware loss. Specifically, we first obtain the low-frequency structure-aware map (S) of the R and T:

$$R_S = (1 - \nabla I') \odot R, \quad T_S = (1 - \nabla I') \odot T. \tag{5}$$

We then compute structural cosine similarity loss $\mathcal{L}_{SCS}$:

$$\mathcal{L}_{SCS} = 1 - \frac{\sum_{i=1}^{N} R_S^i \cdot T_S^i}{\|R_S\|_2 \cdot \|T_S\|_2}, \tag{6}$$

where N is the total number of pixels, and i represents the i-th pixel. The overall loss function is:

$$\mathcal{L} = \lambda_1 \mathcal{L}_1 + \lambda_2 \mathcal{L}_{SCS}, \tag{7}$$

where λ_1 and λ_2 denote weight parameters.

4 Experiments

We first describe the utilized datasets and present the implementation details. Next, we provide a comprehensive analysis of the experimental results, qualitatively and quantitatively. Finally, we conduct ablation studies to validate the effectiveness of the proposed modules and the robustness of our HQGS compared to other approaches.

4.1 EXPERIMENTAL SETUP

Datasets. We evaluate the proposed HQGS pipeline on two datasets: (1) The LLFF dataset (Mildenhall et al., 2019), which contains real-world images from eight distinct scenes, with each scene comprising 20 to 62 images. Of these, 1/4 are reserved for testing, while the remaining are used for training. (2) A synthetic dataset derived from the Blender scenes used in DeblurNeRF (Ma et al., 2022), where 1/8 of the data is used for testing and the other 7/8 for training.

Implementation Details. Our implementation is based on the 3DGS (Kerbl et al., 2023) framework. The total number of training iterations is 50,000. The learning rate for the learnable parameters of 3D Gaussians follows the official settings, while the learning rate for the ESFG module is set to 1e-6. We evaluate our method using various metrics, including PSNR, SSIM, and LPIPS, following previous work (Zhou et al., 2023c; Kerbl et al., 2023; Feng et al., 2024). All experiments are conducted on a single Nvidia GeForce RTX 3090 GPU. The λ_1 and λ_2 in equation 7 are set to 1 and 5, respectively.

	Methods	Low resolution		Compression		Motion Blurry		Gaussian Noisy		Mixed		Ren time (s)
		PSNR↑	LPIPS↓	PSNR↑	LPIPS↓	PSNR↑	LPIPS↓	PSNR↑	LPIPS↓	PSNR↑	LPIPS↓	
NeRF based	NeRF	29.12	0.121	26.83	0.129	28.00	0.054	28.66	0.112	29.16	0.090	∼4.7
	Nan	25.63	0.171	20.43	0.353	23.70	0.215	20.33	0.535	20.32	0.688	∼10.7
	NVSR	30.00	0.075	27.25	0.088	28.48	0.041	29.00	0.073	29.69	0.067	∼4.3
	NeRFLiX	30.63	0.064	28.40	0.069	28.52	0.039	29.84	0.083	30.17	0.059	∼7.5
3DGS based	3DGS	29.25	0.092	27.95	0.076	28.28	0.039	29.49	0.052	28.99	0.109	∼0.2
	SRGS	31.12	0.048	28.22	0.087	28.14	0.041	29.69	0.047	30.33	0.056	∼0.2
	HQGS	31.70	0.036	28.92	0.044	29.03	0.029	30.41	0.029	30.92	0.037	∼0.2

Table 1: Results of both NeRF-based and 3DGS-based methods on LLFF dataset (Ma et al., 2022). Ren time denotes the rendering time for each frame. The red color indicates the best results, and the blue color indicates the second-best results.

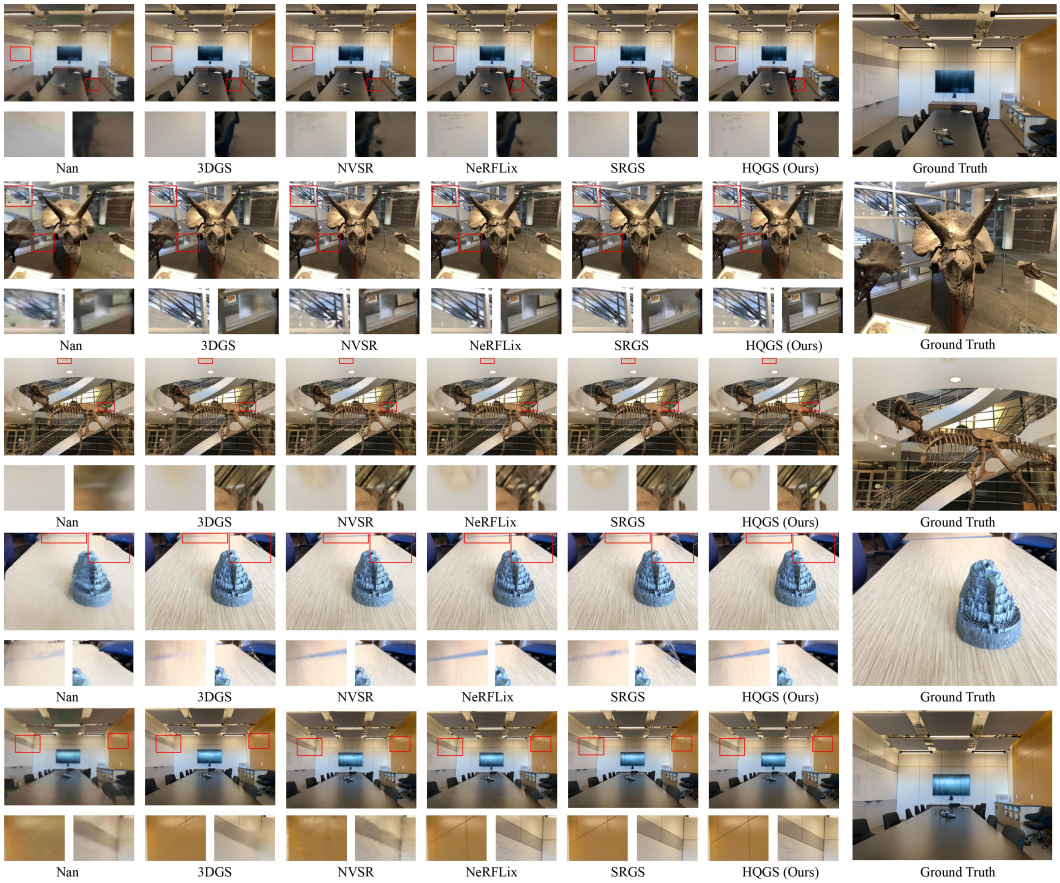
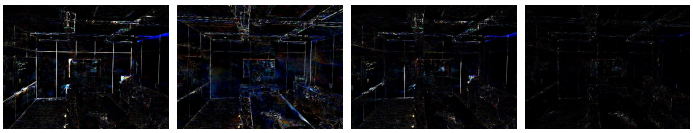


Figure 5: Visualization of the low-resolution, compression, blurry, noisy and mixed degradation scenes on LLFF (Mildenhall et al., 2019).

Evaluated Methods. We evaluate methods that provide code and pre-trained models for fair and comprehensive comparisons. For NeRF-based methods, we choose NeRF (Mildenhall et al., 2021), Nan (Pearl et al., 2022), NVSR (Bahat et al., 2022), and NeRFLiX (Zhou et al., 2023b). In terms of 3DGS-based approaches, 3DGS (Kerbl et al., 2023) and SRGS (Feng et al., 2024) are used for comparison. We retrain all methods using low-high-quality pairs. The pre-trained IVM is used in NeRFLiX. To comprehensively compare the efficiency and performance of the networks, all 3DGS-based models are trained for 50,000 iterations. Meanwhile, we also provide a comparison of the training time in the following section.

4.2 RESULTS AND COMPARISONS.

Novel View Reconstruction on LLFF (Mildenhall et al., 2019). Table 1 compares HQGS to all NeRF (Mildenhall et al., 2021; Pearl et al., 2022; Bahat et al., 2022; Zhou et al., 2023b) and 3DGS (Kerbl et al., 2023; Feng et al., 2024) baselines on the LLFF dataset, showing that HQGS achieves the best results across all degradation conditions. In the low-resolution setting, it achieves gains of 2.49 dB/0.083, 0.98 dB/0.026, 2.36 dB/0.054, and 0.49 dB/0.01 in PSNR/LPIPS over NeRF, NeRFLiX, 3DGS, and SRGS, respectively. For JPEG compression conditions, it has gained 0.42 dB/0.025 and 0.7 dB/0.043 over NeRFLiX and SRGS. It can also be seen that NeRF-based methods, such as NeRFLiX, require 7.5 seconds for rendering vs. only 200 milliseconds for HQGS. Figure 5 compares images synthesized by the different methods under low resolution, compression, blurry, noisy, and mixed degradation conditions. The images generated by HQGS contain richer high-frequency details and more precise visual effects. More results are presented in the appendix. We also compare the difference map between the rendered image from existing methods and its corresponding clean one in Figure 6. A lot of high-frequency details and low-frequency areas involve some structures that are bright, which means these areas are not learned well. In contrast, HQGS achieves better performance in both low-frequency and high-frequency regions.



(a) 3DGS (b) NeRFLix (c) SRGS (d) HQGS

Figure 6: Comparison of the difference map between the rendered image and clean version contains high- and low-frequency information.

Novel View Reconstruction on DeblurNeRF (Ma et al., 2022). Table 2 presents the average PSNR and LPIPS of different methods under five degradation conditions. HQGS achieves the best performance under all metrics. Compared to the NeRFLix (Zhou et al., 2023b), it has a gain of 1.17 dB/0.014/0.014 on PSNR/SSIM/LPIPS. HQGS also delivers substantial improvements across all metrics over SRGS (Feng et al., 2024), which is designed for resolution enhancement.

Methods	PSNR↑	SSIM ↑	LPIPS↓
NeRF	25.25	0.874	0.079
Nan	20.13	0.735	0.093
NVSR	25.67	0.885	0.073
NeRFLiX	25.93	0.888	0.057
3DGS	25.50	0.882	0.070
SRGS	26.03	0.887	0.065
HQGS (Ours)	27.10	0.902	0.043

Table 2: Results on DeblurNeRF dataset (Ma et al., 2022).

4.3 ABLATION STUDIES ON THE PROPOSED MODULES.

In this section, we conduct ablation studies on the proposed ESFG module and $\mathcal{L}_{SCS}$. Additional analyses can be found in the appendix.

Effectiveness of the ESFG Module. In Table 3, we validate the effectiveness of the proposed ESFG module. V1 represents the original 3DGS. V2 extends V1 by incorporating a semantic-aware feature (SAF) from low-quality images, while V3 uses edge-aware features (EAF) extracted from high-frequency edge-aware maps to guide 3DGS. V4 enhances the framework by applying cross-attention (CA) between SAF and EAF. V5 fuses SAF and EAF by concatenating, and V6 represents our complete framework with all four components. The specific structures of these variants are shown in detail in the appendix.

Using SAF slightly improves PSNR from 26.84 dB to 27.19 dB. Replacing SAF with EAF (V3) improves PSNR by 0.69 dB compared to V2. However, V4 with the cross-attention mechanism (CA) further leads to a 0.51 dB improvement over V2. Combining both SAF and EAF results in a 0.16 dB/0.005/0.003 and 0.85 dB/0.013/0.015 gain in PSNR/SSIM/LPIPS over using SAF or

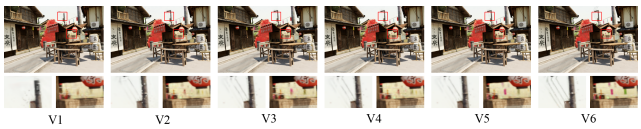


Figure 7: Visual comparison of different variants in Table 3.

Methods	SAF	EAF	Concat	CA	PSNR (dB)↑	SSIM↑	LPIPS↓
V1					26.84	0.902	0.064
V2	✓				$27.19^{\uparrow 0.35}$	$0.904^{\uparrow 0.002}$	$0.062^{\downarrow 0.002}$
V3		✓			$27.88^{\uparrow 1.04}$	$0.912^{\uparrow 0.010}$	$0.050^{\downarrow 0.014}$
V4	✓			✓	$27.70^{\uparrow 0.86}$	$0.910^{\uparrow 0.008}$	$0.054^{\downarrow 0.010}$
V5	✓	✓	✓		$28.04^{\uparrow 1.20}$	$0.917^{\uparrow 0.015}$	$0.047^{\downarrow 0.017}$
V6	✓	✓	✓	✓	$28.22^{\uparrow 1.38}$	$0.919^{\uparrow 0.017}$	$0.045^{\downarrow 0.019}$

Table 3: Effectiveness of the proposed ESFG module. The experiments are conducted on the 'Wine' scene with blurry degradation from the DeblurNeRF dataset.

Methods	$\mathcal{L}_1$	$\mathcal{L}_{BGM}$	$\mathcal{L}_{SP}$	$\mathcal{L}_{SCS}$	PSNR (dB)↑	SSIM↑	LPIPS↓
V1	✓				28.22	0.919	0.045
V2	✓	✓			$28.31^{\uparrow 0.09}$	$0.921^{\uparrow 0.002}$	$0.040^{\downarrow 0.005}$
V3	✓		✓		$28.33^{\uparrow 0.11}$	$0.923^{\uparrow 0.004}$	$0.041^{\downarrow 0.004}$
V4	✓			✓	$29.09^{\uparrow 0.78}$	$0.930^{\uparrow 0.011}$	$0.031^{\downarrow 0.014}$

Table 4: Ablation studies on the proposed $\mathcal{L}_{SCS}$. The experiments are conducted on the 'Wine' scene with blurry degradation from the DeblurNeRF dataset.

Method	3DGS	3DGS	3DGS	3DGS+ESFG	3DGS+ESFG	3DGS+ESFG
Sobel	✓			✓		
Gaussian		✓			✓	
Laplace			✓			✓
PSNR(dB)↑	27.88	27.48	27.62	28.22	27.89	27.92
SSIM↑	0.912	0.908	0.910	0.919	0.913	0.914
LPIPS↓	0.050	0.056	0.053	0.045	0.052	0.051

Table 5: Ablation studies on different guidance. The experiments are conducted on 'Wine' scenes with blurry degradation from the DeblurNeRF dataset.

EAF individually. Finally, incorporating all components (V6) achieves the best results, with a 1.38 dB/0.017/0.019 improvement in PSNR/SSIM/LPIPS over V1. Figure 7 shows visual results in blurry scenes, where V1 loses some details. While additional information helps, the improvements remain limited. V2 renders only three power lines, with the flag still blurry. V3 and V4 correctly capture four antennas and slightly improve flag clarity, but colors remain uniform. V6 accurately captures the power lines and introduces noticeable color variations in the flag.

Effectiveness of $\mathcal{L}_{\mathbf{SCS}}$. We assess the effectiveness of the $\mathcal{L}_{SCS}$ in Table 4. V1 denotes the 3DGS with the ESFG module trained using only $\mathcal{L}_1$. V2 builds upon V1 by incorporating a structure-aware pixel-level loss ($\mathcal{L}_{SP}$), which calculates the $\mathcal{L}_1$ between structural elements in the rendered and target images. V3 introduces the $\mathcal{L}_{BGM}$ (Lin et al., 2023b), and V4 replaces the $\mathcal{L}_{SP}$ with our structural cosine similarity loss ($\mathcal{L}_{SCS}$). We observe that adding the $\mathcal{L}_{BGM}$ to the $\mathcal{L}_1$ provides a slight improvement (approximately 0.1 dB in PSNR). Similarly, $\mathcal{L}_{SP}$ results in a marginal improvement. However, substituting local structural constraints with the global $\mathcal{L}_{SCS}$ yields a significant 0.87 dB improvement over the original L1 loss, indicating that global low-frequency structural information is more effective at capturing overall image consistency than pixel-level constraints.

Comparison of Other Methods for Obtaining High-Frequency Maps. We compare the results of different operators in both the 3DGS framework (V3 in Table 3) and the 3DGS with our ESFG module (V6 in Table 3). The visualization of these maps is shown in the appendix. Table 5 shows that high-frequency information from the Gaussian filter and Laplace operator can improve performance. However, it is still lower than our high-frequency edge-aware feature in both V3 and V6 frameworks.

Table 6: Robustness validation on two progressive degradation test sets on (Ma et al., 2022). The red color indicates the best results, and the blue color indicates the second-best results.

Methods	Gaussian noise								Low resolution							
	0		10		25		50		1 ×		2 ×		4 ×		8 ×	
	PSNR	LPIPS	PSNR	LPIPS	PSNR	LPIPS	PSNR	LPIPS	PSNR	LPIPS	PSNR	LPIPS	PSNR	LPIPS	PSNR	LPIPS
NeRF	30.42	0.072	29.11	0.079	26.78	0.091	23.04	0.143	29.83	0.113	29.71	0.128	28.83	0.138	28.16	0.148
3DGS	30.21	0.043	29.19	0.052	27.46	0.077	22.78	0.109	30.25	0.091	30.18	0.094	29.25	0.092	28.63	0.111
NeRFLix	30.86	0.054	29.47	0.063	27.04	0.071	23.32	0.112	31.42	0.069	30.87	0.069	30.12	0.076	29.66	0.096
SRGS	30.71	0.036	29.17	0.045	27.63	0.061	23.21	0.126	31.36	0.078	30.93	0.065	30.54	0.061	30.11	0.087
HQGS	31.32	0.018	30.41	0.029	28.63	0.043	26.31	0.067	32.08	0.031	31.85	0.033	31.61	0.038	31.37	0.051

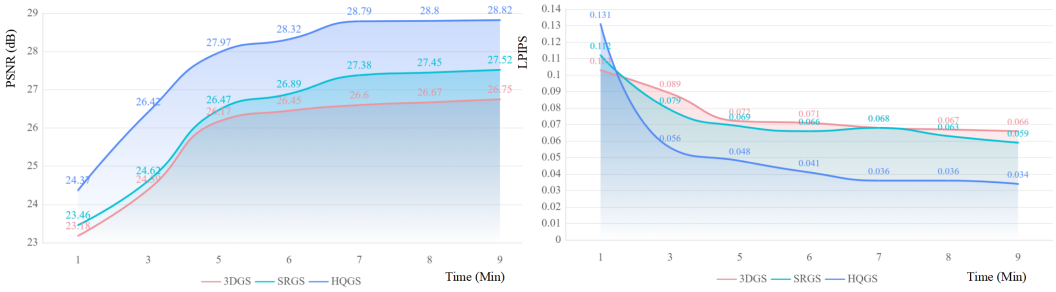


Figure 8: Training time vs quality for blurry scenes. HQGS performs well against 3GDS and SRGS under both the PSNR and LPIPS metrics.

4.4 Analysis on Robustness of our HQGS.

So far, we have considered mild image degradations. In this section, we investigate the robustness of HQGS across degradation levels. We construct two types of progressive degradation test sets: (1) Noise: Gaussian noise with variances of 0 (clean), 10, 25, and 50, and (2) Low resolution: images downsampled to 1× (clean), 2×, 4×, and 8× scales. Table 6 shows that the gains of HQGS over existing methods increases with the degradation strength and are quite large for the strongest degradations for both types of degradation. Especially as noise becomes severe (variance increasing from 25 to 50), all methods exhibit a sharp drop, for example, SRGS from 27.63 dB to 23.21 dB and NeRFLix from 27.04 dB to 23.32 dB, making high-quality rendering challenging. In contrast, HQGS has a much smaller decline and maintains a robust performance of 26.31 dB, which achieves a 3.10 dB and 2.99 dB improvement compared to SRGS and NeRFLix. A similar trend is observed in low-resolution scenarios, where our method consistently demonstrates more robust performance than others. For instance, under the 8 × setting, our method achieves a 1.26 dB/0.036 and 1.71 dB/0.045 increase over SRGS and NeRFLix on PSNR and LPIPS.

4.5 Training Time vs Quality.

In general, there is a trade-off between rendering quality and time, as higher quality rendering can be achieved by considering more views of the scene. Figure 8 compares the performance of different 3DGS-based approaches as a function of training time. It is clear that HQGS achieves a better trade-off than 3DGS and SRGS. For instance, a 5-minute reconstruction by HQGS has a 1.22 dB (0.018) gain in PSNR (LPIPS) over the 3DGS results for 9 minutes.

5 Conclusion

In this paper, we propose a robust Gaussian variant HQGS, which performs favorably under various degradation conditions and exhibits strong robustness as the degradation level increases. Building upon the limitations of the existing 3DGS in rendering scene details due to degradation, we propose an edge-semantic fusion guidance module to guide the positional parameters of 3DGS. Furthermore, during optimization, we introduce a global structural cosine similarity loss ($\mathcal{L}_{SCS}$) to complement the pixel-level $\mathcal{L}_1$ loss. Extensive experimental results demonstrate the effectiveness and robustness of the proposed HQGS.

REFERENCES

Yuval Bahat, Yuxuan Zhang, Hendrik Sommerhoff, Andreas Kolb, and Felix Heide. Neural volume super-resolution. *arXiv preprint arXiv:2212.04666*, 2022. 1, 3, 7, 8

Yulong Bian, Chenglei Yang, Fengqiang Gao, Hanchao Li, Xiaowen Sun, Xiangxu Meng, and Yu Wang. A framework for physiological indicators of flow in vr games: construction and preliminary evaluation. *Personal and Ubiquitous Computing*, pp. 821–832, 2016. 1

Anpei Chen, Zexiang Xu, Andreas Geiger, Jingyi Yu, and Hao Su. Tensorf: Tensorial radiance fields. In *ECCV*, pp. 333–350, 2022. 2

Zhiqin Chen, Thomas Funkhouser, Peter Hedman, and Andrea Tagliasacchi. Mobilenerf: Exploiting the polygon rasterization pipeline for efficient neural field rendering on mobile architectures. In *CVPR*, pp. 16569–16578, 2023. 2

Nashwan Dawood. Vr–roadmap: A vision for 2030 in the built environment. *Virtual Futures for Design, Construction and Procuremen*, pp. 261, 2009. 1

Kangle Deng, Andrew Liu, Jun-Yan Zhu, and Deva Ramanan. Depth-supervised nerf: Fewer views and faster training for free. In *CVPR*, pp. 12882–12891, 2022. 2

Jiangxin Dong, Jinshan Pan, Zhongbao Yang, and Jinhui Tang. Multi-scale residual low-pass filter network for image deblurring. In *CVPR*, pp. 12345–12354, 2023. 1, 4, 6

Zhenxuan Fang, Fangfang Wu, Weisheng Dong, Xin Li, Jinjian Wu, and Guangming Shi. Self-supervised non-uniform kernel estimation with flow-based motion prior for blind image deblurring. In *CVPR*, pp. 18105–18114, 2023. 3

Mana Farshid, Jeannette Paschen, Theresa Eriksson, and Jan Kietzmann. Go boldly!: Explore augmented reality (ar), virtual reality (vr), and mixed reality (mr) for business. *Business horizons*, pp. 657–663, 2018. 1

Francesco Fassi, Alessandro Mandelli, Simone Teruggi, Fabrizio Rechichi, Fausta Fiorillo, and Cristiana Achille. Vr for cultural heritage. In *In International Conference on Augmented Reality, Virtual Reality and Computer Graphics*, pp. 6445–6454, 2016. 1

Xiang Feng, Yongbo He, Yubo Wang, Yan Yang, Wen Li, Yifei Chen, Zhenzhong Kuang, Jiajun Ding, Jianping Fan, and Jun Yu. Srgs: Super-resolution 3d gaussian splatting. *arXiv preprint arXiv:2404.10318*, 2024. 1, 2, 3, 6, 7, 8

Sara Fridovich-Keil, Alex Yu, Matthew Tancik, Qinhong Chen, Benjamin Recht, and Angjoo Kanazawa. Plenoxels: Radiance fields without neural networks. In *CVPR*, pp. 5501–5510, 2022. 2

Stephan J. Garbin, Marek Kowalski, Matthew Johnson, Jamie Shotton, and Julien Valentin. Fastnerf: High-fidelity neural rendering at 200fps. In *ICCV*, pp. 14346–14355, 2021. 2

Xudong Huang, Wei Li, Jie Hu, Hanting Chen, and Yunhe Wang. Refsr-nerf: Towards high fidelity and super resolution view synthesis. In *CVPR*, pp. 8244–8253, 2022. 3

Bernhard Kerbl, Georgios Kopanas, Thomas Leimkühler, and George Drettakis. 3d gaussian splatting for real-time radiance field rendering. *ACM Trans. Graph*, pp. 1–14, 2023. 2, 3, 5, 6, 7, 8

Lingshun Kong, Jiangxin Dong, Jianjun Ge, Mingqiang Li, and Jinshan Pan. Efficient frequency domain-based transformers for high-quality image deblurring. In *CVPR*, pp. 5886–5895, 2023. 4

Xin Lin, Chao Ren, Kelvin CK Chan, Lu Qi, Jinshan Pan, and Ming-Hsuan Yang. Multi-task image restoration guided by robust dino features. *arXiv preprint arXiv:2312.01677*, 2023a. 3

Xin Lin, Chao Ren, and Xiao Liu. Unsupervised image denoising in real-world scenarios via self-collaboration parallel generative adversarial branches. In *ICCV*, pp. 12642–12652, 2023b. 1, 3, 4, 6, 9

Xin Lin, Yuyan Zhou, Jingtong Yue, Chao Ren, Kelvin CK Chan, Lu Qi, and Ming-Hsuan Yang. Re-boosting self-collaboration parallel prompt gan for unsupervised image restoration. *arXiv preprint arXiv:2408.09241*, 2024. 3, 4

Li Ma, Xiaoyu Li, Jing Liao, Qi Zhang, Xuan Wang, Jue Wang, and Pedro V. Sander. Deblur-nerf: Neural radiance fields from blurry images. In *CVPR*, pp. 12861–12870, 2022. 1, 6, 7, 8, 10

Ben Mildenhall, Pratul P. Srinvasan, Rodrigo Ortiz-Cayon, Nima Khademi Kalantari, Ravi Ramamoorthi, Ren NG, and Abhishek Kar. Local light field fusion: Practical view synthesis with prescriptive sampling guidelines. *ToG*, pp. 1–14, 2019. 6, 7, 8

Ben Mildenhall, Pratul P. Srinivasan, Matthew Tancik, Jonathan T. Barron, Ravi Ramamoorthi, and Ren Ng. Nerf: Representing scenes as neural radiance fields for view synthesis. *Communications of the ACM*, 2021. 2, 7, 8

Jinshan Pan, Boming Xu, Jiangxin Dong, Jianjun Ge, and Jinhui Tang. Deep discriminative spatial and temporal network for efficient video deblurring. In *CVPR*, pp. 22191–22200, 2023. 3

Naama Pearl, Tali Treibitz, and Simon Korman. Nan: Noise-aware nerfs for burst-denoising. In *CVPR*, pp. 12672–12681, 2022. 1, 3, 7, 8

Christian Reiser, Songyou Peng, Yiyi Liao, and Andreas Geiger. Kilonerf: Speeding up neural radiance fields with thousands of tiny mlps. In *ICCV*, pp. 14335–14345, 2021. 2

Chao Ren, Xiaohai He, and Chuncheng Wang. Adaptive consistency prior based deep network for image denoising. In *CVPR*, pp. 8596–8606, 2021. 3

Chao Ren, Yizhong Pan, and Jie Huang. Enhanced latent space blind model for real image denoising via alternative optimization. In *NeurIPS*, pp. 38386–38399, 2022. 3

Lei Sun, Christos Sakaridis, Jingyun Liang, Peng Sun, Jiezhang Cao, Kai Zhang, Qi Jiang, Kaiwei Wang, and Luc Van Gool. Event-based frame interpolation with ad-hoc deblurring. In *CVPR*, pp. 18043–18052, 2023. 3

Lindeberg T. Scale space. *Encyclopedia of Computer Science and Engineering.*, pp. 2495–2504, 2009. 16

Chen Wang, Xian Wu, Yuan-Chen Guo, Song-Hai Zhang, Yu-Wing Tai, and Shi-Min Hu. Nerf-sr: High quality neural radiance fields using supersamplin. In *ACM MM*, pp. 6445–6454, 2022. 3

Qiangeng Xu, Zexiang Xu, Julien Philip, Sai Bi, and Zhixin Shu. Point-nerf: Point-based neural radiance fields. In *CVPR*, pp. 5438–5448, 2022. 2

Ganlin Yang, Kaidong Zhang, Jingjing Fu, and Dong Liu. Drantal-nerf: Diffusion-based restoration for anti-aliasing neural radiance field. *arXiv preprint arXiv:2407.07461*, 2024. 3

Syed Waqas Zamir, Aditya Arora, Salman Khan, Munawar Hayat, Fahad Shahbaz Khan, Ming-Hsuan Yang, and Ling Shao. Multi-stage progressive image restoration. In *CVPR*, pp. 14821–14831, 2021. 3

Syed Waqas Zamir, Aditya Arora, Salman Khan, Munawar Hayat, Fahad Shahbaz Khan, and Ming-Hsuan Yang. Restormer: Efficient transformer for high-resolution image restoration. In *CVPR*, pp. 5728–5739, 2022a. 3, 4

Syed Waqas Zamir, Aditya Arora, Salman Khan, Munawar Hayat, Fahad Shahbaz Khan, Ming-Hsuan Yang, and Ling Shao. Learning enriched features for fast image restoration and enhancement. *IEEE TPAMI*, 2022b. 4

Kun Zhou, Wenbo Li, Nianjuan Jiang, Xiaoguang Han, and Jiangbo Lu. From nerflix to nerflix++: A general nerf-agnostic restorer paradigm. *IEEE TPAMI*, 2023a. 3

Kun Zhou, Wenbo Li, Yi Wang, Tao Hu, Nianjuan Jiang, Xiaoguang Han, and Jiangbo Lu. Nerflix: High-quality neural view synthesis by learning a degradation-driven inter-viewpoint mixer. In *CVPR*, pp. 12363–12374, 2023b. 1, 2, 3, 7, 8

Xiaoyu Zhou, Zhiwei Lin, Xiaojun Shan, Yongtao Wang, Deqing Sun, and Ming-Hsuan Yang. Drivinggaussian: Composite gaussian splatting for surrounding dynamic autonomous driving scenes. In *CVPR*, pp. 12345–12354, 2023c. 6

A SPECIFIC STRUCTURE OF VARIANTS IN SECTION 4.3 OF THE MAIN PAPER

Five comparative variants are tested to verify the influence of the proposed Edge Fusion Guidance (ESFG) module. Since the experimental results have been reported in sub-section 4.3, we mainly describe their specific structures in this section. As the V1 in Table 3 of the main paper is the 3DGS without learnable modules, we only provide details for the remaining variants. As shown in Figures. 9, 10, 11 and 12.

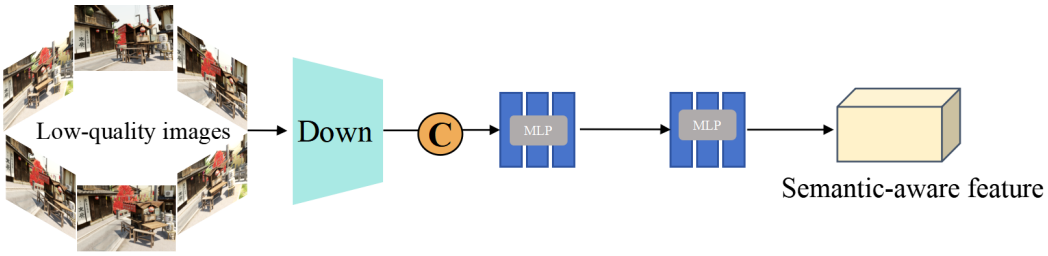


Figure 9: The structure of V2 in Table 3 of the main paper.

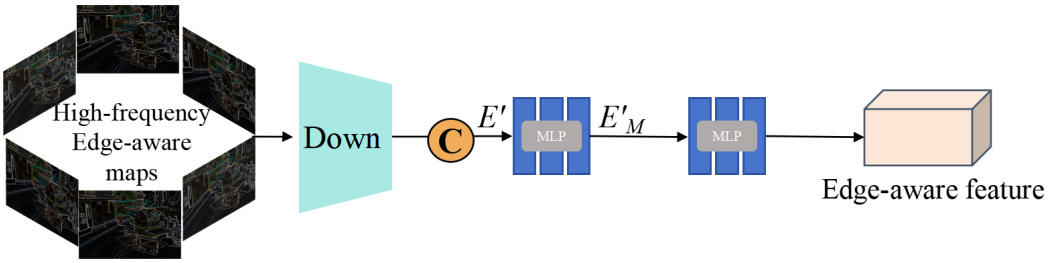


Figure 10: The structure of V3 in Table 3 of the main paper.

B WHOLE RESULTS ON LLFF

In Table 7, we present all PSNR, SSIM, and LPIPS metrics on the comparison results of the LLFF dataset. HQGS achieve the best results across both NeRF-based and 3DGS-based methods.

Methods	Low resolution			Compression			Motion Blurry			Gaussian Noisy			Mixed		
	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓
NeRF	29.12	0.917	0.121	26.83	0.831	0.129	28.00	0.923	0.054	28.66	0.868	0.112	29.16	0.916	0.090
Nan	25.63	0.869	0.171	20.43	0.742	0.353	23.70	0.721	0.215	20.33	0.716	0.535	20.32	0.747	0.688
NVSR	30.00	0.924	0.075	27.25	0.876	0.088	28.48	0.931	0.041	29.00	0.879	0.073	29.69	0.921	0.067
NeRFLiX	30.63	0.932	0.064	28.50	0.899	0.069	28.52	0.931	0.039	29.84	0.890	0.083	30.17	0.932	0.059
3DGS	29.25	0.919	0.092	27.95	0.888	0.076	28.28	0.930	0.039	29.49	0.888	0.052	28.99	0.904	0.109
SRGS	31.12	0.941	0.048	28.22	0.891	0.087	28.14	0.926	0.041	29.79	0.891	0.047	30.33	0.931	0.056
HQGS	31.61	0.947	0.038	28.92	0.912	0.044	29.03	0.938	0.029	30.42	0.898	0.029	30.92	0.940	0.037

Table 7: Results of all NeRF-based and 3DGS-based methods on LLFF dataset. The red color indicates the best results, and the blue color indicates the second-best results.

C ABLATION STUDY ON THE λ_2 IN EQUATION (9) OF THE MAIN PAPER.

In this section, we conduct an ablation study on the λ_2 in equation (9) of the main paper: keeping the L1 loss parameter λ_1 in equation (9) of the main paper fixed at 1, we select values of 0, 0.5, 1,

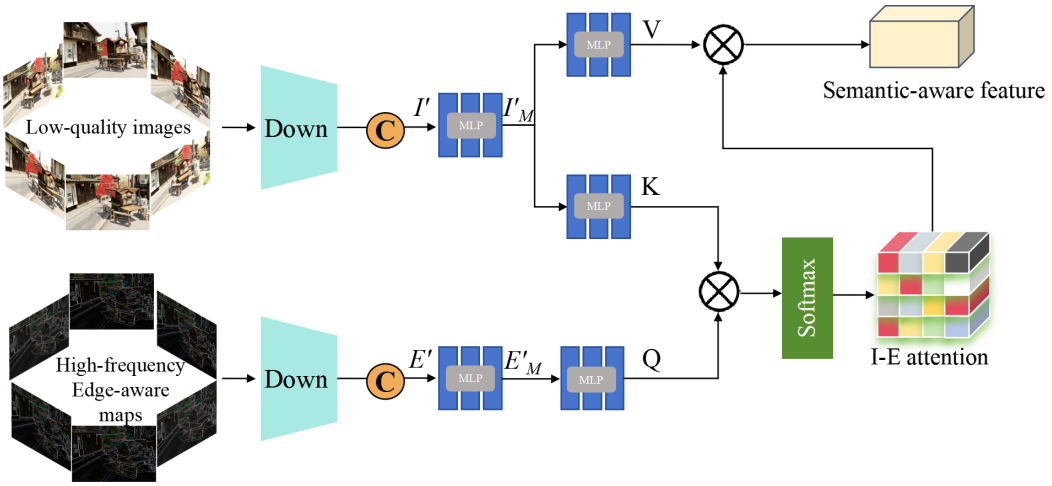


Figure 11: The structure of V4 in Table 3 of the main paper.

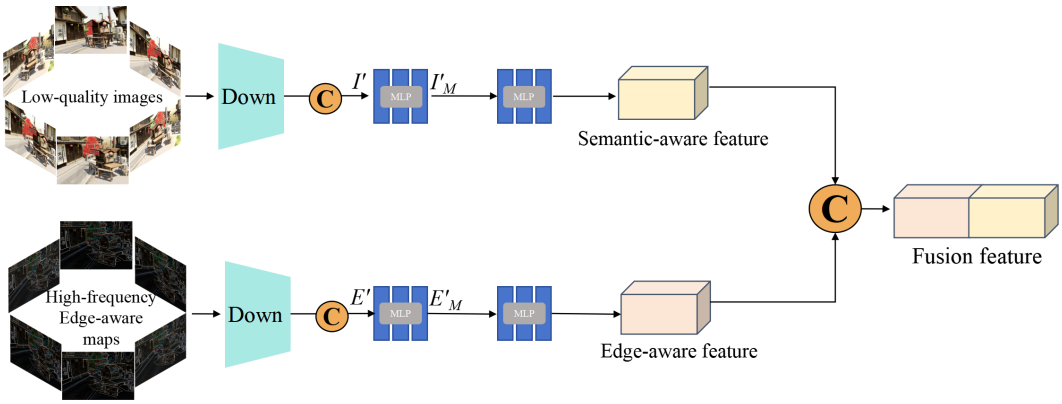


Figure 12: The structure of V5 in Table 3 of the main paper.

2, 3, 4, 5, and 6 for evaluation. Using the low-resolution scenes from the LLFF dataset as input, we test the performance of the generated novel views. As shown in Table 8, the $\lambda_2 = 5$ has the best performance, so we choose this one.

D Visualization Results of Other Methods for Obtaining High-Frequency Maps.

Our goal is to capture contour regions representing the image's high-frequency content. We first show the $\nabla I'$ and E in Section 3.2 in the main paper in Figure 13. After that, as shown in Figure 14, we explore other methods for extracting high-frequency content, such as using a Gaussian high-pass filter and Laplace operator to obtain high-frequency responses.

Method	0	0.5	1	2	3	4	5	6
PSNR(dB)↑	29.25	29.70	30.62	31.15	31.37	31.53	31.61	31.56
SSIM↑	0.919	0.919	0.932	0.941	0.944	0.945	0.947	0.947
LPIPS↓	0.121	0.071	0.051	0.047	0.046	0.042	0.038	0.039

Table 8: Ablation studies on the λ_2 in equation (9) of the main paper.

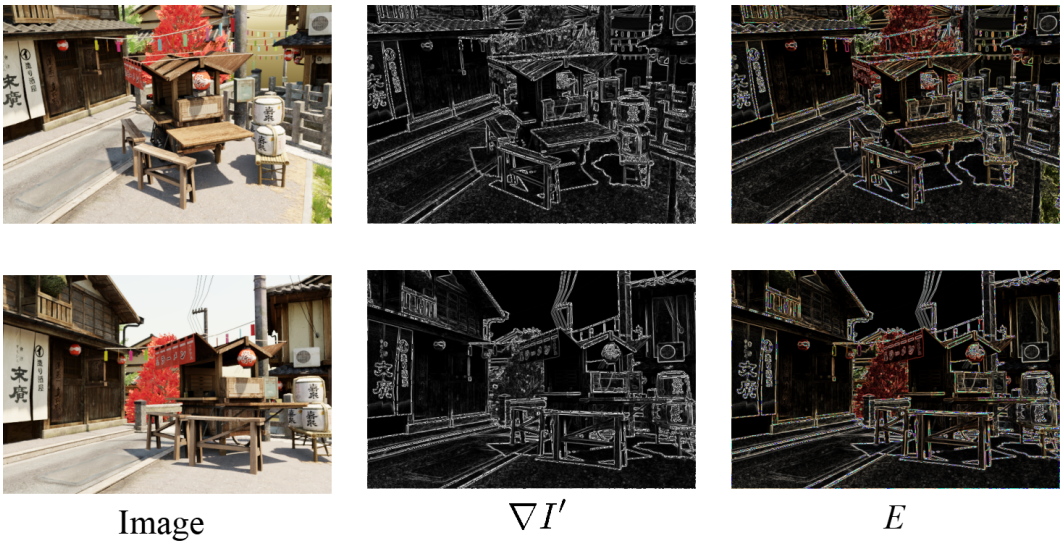


Figure 13: Visualization the $\nabla I'$ and E.

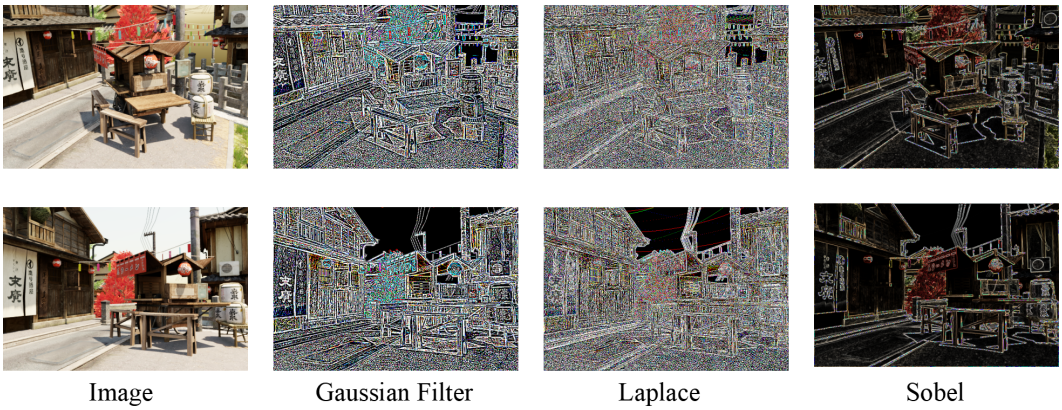


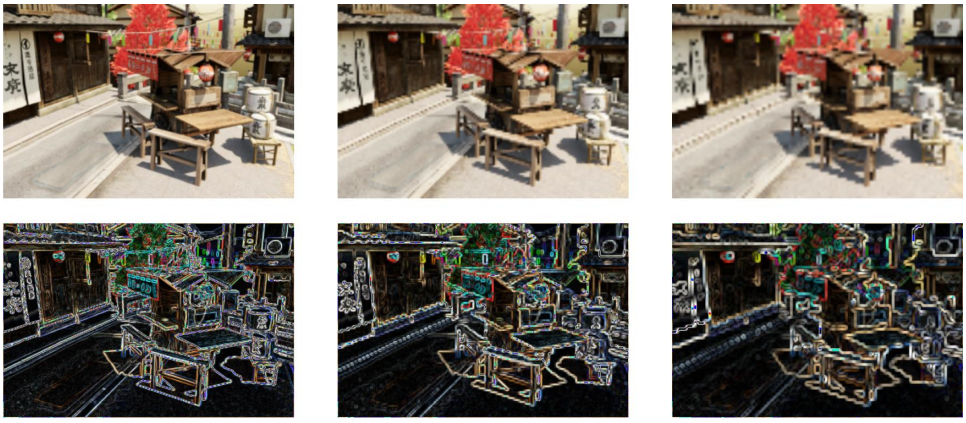
Figure 14: Visualization the map from different operators.

E Robustness Validation on Edge Maps.

We leverage the well-known fact (T., 2009) that edges have sufficient frequency information and can be obtained by an edge detection operator even from degraded images. We provide visualizations of edge detection results obtained using the Sobel operator under progressively challenging conditions, as shown in Figure 15 in the submitted Supplementary Material. Additionally, we quantified the results by calculating the PSNR metrics of the images and edge maps under various conditions compared to the clean ones. Then, we presented the PSNR variance of the images and edge maps: 10.18/0.285. This further demonstrates that the edge maps are more robust under progressively degraded conditions. Thus, edge detection performs well in extracting edge information to a considerable extent, even under relatively challenging conditions.

F Ablation Studies on the Downsampling Parameters in the ESFG Module.

We introduce downsampling in the ESFG module to reduce dimensionality and save computational resources. To further investigate its impact, we supplement additional experiments under two con-



2 × 4 × 8 ×

Figure 15: Visualization of edge maps on progressive degradation scenes.

Methods	1 ×	2 ×	4 ×
PSNR(dB)↑	31.79	31.70	31.42
Time(Min)↓	16.4	13.3	11.2

Table 9: Ablation Studies on the Downsampling Parameters in the ESFG Module.

figurations: Without Downsampling and 4 × Downsampling in Table 9. Therefore, we choose the 2 × to balance the training time and the performance.

G ADDITIONAL QUALITATIVE RESULTS.

This section presents additional results to underscore the advanced capabilities of our HQGS over other leading NeRF-based and 3DGS-based techniques. Illustrated in Figures. 16, 17, 18, our methods consistently deliver superior quality of rendered image.

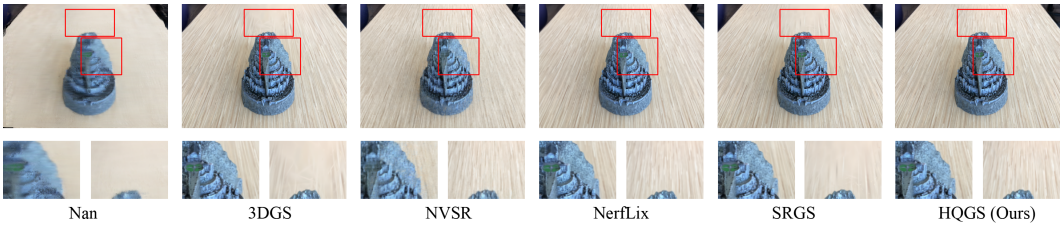


Figure 16: Visualization of the noisy scenes on the LLFF dataset.

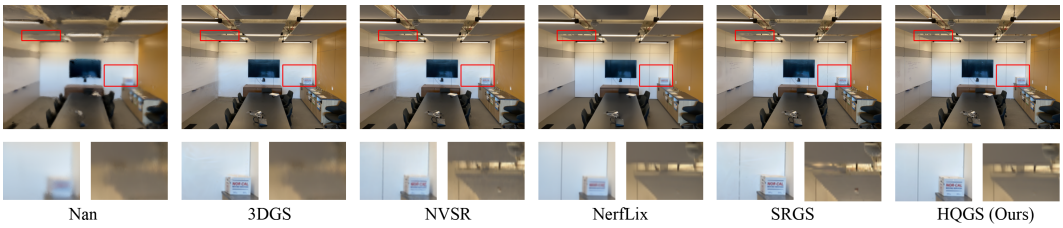


Figure 17: Visualization of the low-resolution scenes on the LLFF dataset.

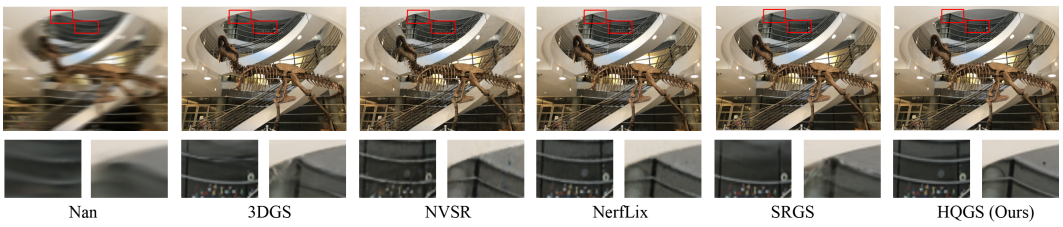


Figure 18: Visualization of the blurry scenes on the LLFF dataset.