This ICCV paper is the Open Access version, provided by the Computer Vision Foundation.
Except for this watermark, it is identical to the accepted version;
the final published version of the proceedings is available on IEEE Xplore.

OCSplats: Observation Completeness Quantification and Label Noise Separation in 3DGS

Han Ling[1], Xian Xu[2,3], Yinghui Sun[2], Quansen Sun[1*]
[1]Nanjing University of Science and Technology [2] Southeast University [3]CityU HK
1536233573@qq.com,230248289@seu.edu.cn

Abstract

3D Gaussian Splatting (3DGS) has become one of the most promising 3D reconstruction technologies. However, label noise in real-world scenarios—such as moving objects, non-Lambertian surfaces, and shadows—often leads to reconstruction errors. Existing 3DGS-Bsed anti-noise reconstruction methods either fail to separate noise effectively or require scene-specific fine-tuning of hyperparameters, making them difficult to apply in practice. This paper re-examines the problem of anti-noise reconstruction from the perspective of epistemic uncertainty, proposing a novel framework, OCSplats. By combining key technologies such as hybrid noise assessment and observation-based cognitive correction, the accuracy of noise classification in areas with cognitive differences has been significantly improved. Moreover, to address the issue of varying noise proportions in different scenarios, we have designed a label noise classification pipeline based on dynamic anchor points. This pipeline enables OCSplats to be applied simultaneously to scenarios with vastly different noise proportions without adjusting parameters. Extensive experiments demonstrate that OCSplats always achieve leading reconstruction performance and precise label noise classification in scenes of different complexity levels. Code is available [1].

1. Introduction

The new view synthesis addresses the challenge of previously unseen viewpoints. 3D Gaussian Splatting [13, 28, 41] has recently emerged as a groundbreaking technique in this domain, primarily due to its ability to reconstruct geometrically consistent point cloud scenes rapidly. As a result, it has found broad applications in areas such as depth estimation [20, 38], optical flow training [17, 19], and autonomous driving [7, 9, 16, 18, 42].

However, training a 3DGS model requires static scene images with known camera poses. In real-world scenarios, non-Lambertian [1] surfaces, moving objects, and shadows introduce noise into the labels for 3DGS training. Eliminating such noise is nontrivial. Beyond apparent moving objects [29, 30], more subtle phenomena—like shadows and reflections on non-Lambertian surfaces—demand manual per-pixel annotation. These factors underscore the limitations of current reconstruction methods when applied to complex real-world scenes.

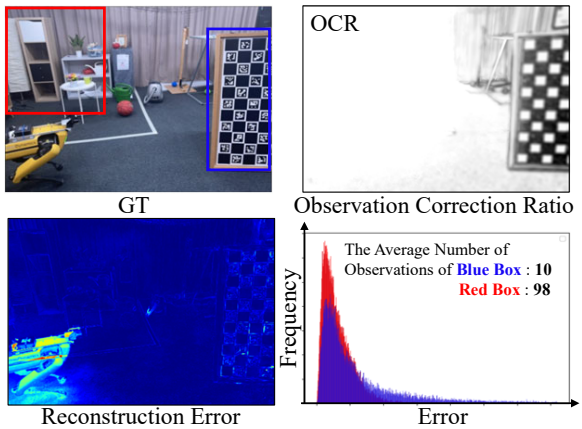

Figure 1. **Observation Cognitive Correction.** The error histogram in the bottom right indicates that even without occlusion, the reconstruction error of the missing observation areas (blue box) in the photo group is often greater than that of the areas with more observations (red box). We proposed the observation cognitive correction technique to measure these difficult areas (blue boxes) and corrected the noise assessment through OCR when classifying label noise, achieving accurate label noise separation.

In previous works [4, 15, 31, 39], anti-noise reconstruction was commonly framed as a metric-classification problem. These methods use reconstructed residuals or uncertainty values to quantify noise and design a metric-based classifier to separate noisy labels from clean data. However, in different scenarios, the types and proportions of noise labels, as well as the limitations of the observation itself, can affect the distribution of noise and clean background. The

*Corresponding author
[1]github.com/HanLingsgjk/OCSplats

previous naive metric classification strategy is insufficient to handle real-world scene variability. To address the above issues, we propose the techniques of **observation cognitive correction** and **dynamic anchor point combination thresholding**.

Observation Cognitive Correction. Consider the question: **Are the captured images truly sufficient to reconstruct every region of the scene perfectly?** In classic datasets [3, 29, 30], areas outside the primary view are often far away and observed fewer times. Even without occlusion, reconstruction errors are still significant in these regions. As shown in Fig. 1, both the red and blue boxed areas are clean, but the blue box, with fewer observations, exhibits more significant reconstruction errors than the red box. This indicates that sufficient observation is a crucial factor for both reconstruction and noise label separation.

Inspired by the principle of triangulation [8, 37], we quantified the camera's observation of Gaussian primitives in 3DGS reconstruction, called observation completeness (OC), which is proportional to the number of effective observations and the diversity of observation camera angles. Furthermore, we propose a noise assessment correction scheme based on OC, which balances measurement errors caused by insufficient observations and significantly improves the generalization of the noise classifier.

Dynamic Anchor Combination Threshold. The proportion of noise and the difficulty of scene reconstruction significantly affect the statistical distribution of background and noisy foreground. Therefore, we propose a dynamic threshold anchor scheme. The noise foreground and background classification threshold for each scene is calculated by two statistical anchor points. They are the maximum inter-class variance position [26] and the mean position of the background class. Unlike the fixed threshold methods [15, 31], our threshold is calculated based on the unique statistical characteristics of each scene, so it can be directly applied to most scenes without adjusting any parameters. In Tab. 1, Tab. 3 and Fig. 9, we proved that our dynamic threshold scheme could achieve much better noise separation and scene reconstruction effects even though scene characteristics are vastly different.

Finally, we developed our method OCSplats based on the 3DGS [13]. The overall pipeline can be summarized as reconstruction noise assessment, observation completeness-based correction, and self-supervised label noise classification based on dynamic thresholds. Moreover, we propose a pruning strategy for 3DGS based on observation completeness. The core idea is to remove Gaussian primitives with severely insufficient effective observations. This strategy not only reduces the total number of Gaussian primitives by approximately 10%, but also significantly stabilizes the quality of 3DGS reconstruction. Through experiments on mainstream datasets with noisy reconstructions, OCSplats achieves the best visual performance in label noise classification and demonstrates significantly superior reconstruction accuracy, especially in complex scenes.

Our key contributions can be summarized as follows:

1. A novel quantitative measure of observation completeness (OC) method, which allows people to visually determine whether the observation of each region in the scene is sufficient.
2. An assessment correction and pruning method based on observation completeness was proposed, significantly improving the erroneous noise classification caused by lack of observation.
3. A noise classifier self-supervised training pipeline based on dynamic threshold anchor points, achieving robust noise label classification in scenes with different complexities and occlusions.

2. Related Works

2.1. Epistemic Uncertainty and Reconstruction

Generally, epistemic uncertainty [10, 11, 24, 36] mainly arises from insufficient training data. For example, in detection tasks, the detector cannot recognize objects it has never encountered before. In the context of scene reconstruction, epistemic uncertainty typically originates from the limited shooting perspective [2, 23]. Shen et al. [35] quantified epistemic uncertainty using ensemble learning to visualize unobserved regions in NeRF reconstructions. Similarly, Goli et al. [6] quantifies the location uncertainty of NeRF voxels by actively introducing perturbations and constructing an uncertainty field.

In this paper, observation completeness (OC) is similar to epistemic uncertainty, but our OC focuses more on whether the scene appearance has been adequately learned. In contrast to prior work [6, 35], our OC employs an iterative computation method tailored for 3DGS, making it highly efficient without requiring extensive post-processing. Additionally, we apply OC to noise-robust reconstruction, effectively helping 3DGS improve reconstruction quality and reduce label noise classification errors.

2.2. Stochastic Uncertainty and Reconstruction

Stochastic uncertainty [21] arises from the inherent noise in the data itself and does not diminish even as the data size grows. Most existing NeRF-based denoising reconstruction methods utilize the stochastic uncertainty loss [12, 29] to supervise the NeRF reconstruction. In this setting, the stochastic uncertainty loss actively suppresses errors in regions that are difficult to learn—such as moving objects, non-Lambertian surfaces, or areas lacking sufficient observations—thereby achieving a noise-robust learning effect.

However, in the 3DGS method, Gaussian split pruning relies on gradients. The stochastic uncertainty loss signif-

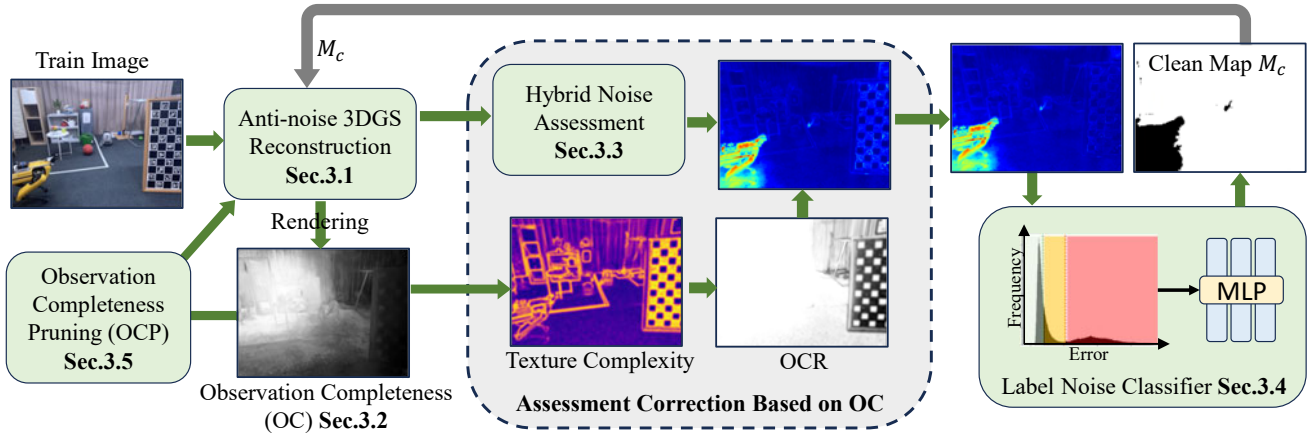

Figure 2. **Pipeline** Our method is mainly divided into two stages. In the first stage, we assess the noise present in the labels. In the second stage, we self-supervised train an MLP network to infer reasonable clean regions from feature embeddings. Specifically, this paper proposes a learning completeness theory for 3DGS to modify noise assessment, which can more accurately evaluate and identify noise in different scenarios.

icantly shifts the loss weighting, which ultimately causes 3DGS to collapse. Consequently, existing approaches [15] block the back-propagation of the uncertainty loss and adopt a metric-threshold classification strategy to identify noisy labels. A similar residual-based method [31] also exists, differing only in that it uses reconstruction residuals rather than stochastic uncertainty as its threshold classification metric. The core problem with these threshold-based methods is their inability to accommodate diverse scenarios: each scenario requires retuning the classification threshold to achieve ideal performance. Our proposed dynamic anchor threshold approach circumvents threshold-based methods' poor generalizability while maintaining high reconstruction quality.

3. Our Approach

This section introduces how to implement our method, and the overall process is shown in Fig. 2. Specifically, Sec.3.1 introduces key background knowledge about 3DGS reconstruction; Sec.3.2 proposes quantification and visualization methods for observation completeness; Sec.3.3 proposes a hybrid noise assessment scheme and how to perform assessment correction based on observation completeness; Sec.3.4 proposes a self-supervised separation scheme for label noise based on hybrid assessment; Sec.3.5 introduces a novel pruning strategy based on observation completeness.

3.1. Background Knowledge of 3DGS

Our method is implemented by 3DGS technology, where the scene is represented as a set of explicit 3DGS primitives $\mathcal{G} = \{g_1, g_2, ..., g_N\}$. Each Gaussian primitive g is defined by its center position $\mathbf{p}_k$, pose parameter $\mathbf{\Sigma}_k$, opacity parameter $\alpha_k \in [0, 1]$, and spherical harmonic parameter $\mathbf{SH}$ with respect to color $\mathbf{c}$. During rendering, the 3D Gaussian primitive is first projected onto the imaging plane, and the projected 2D Gaussian primitive g_k^{2D} can be written as:

$$g_k^{2D}(\boldsymbol{x}) = e^{-\frac{1}{2}(\boldsymbol{x}-\mathbf{p}_k)^T\left(\mathbf{\Sigma}_k^{2D}\right)^{-1}(\boldsymbol{x}-\mathbf{p}_k)} \tag{1}$$

where $\mathbf{\Sigma}_k^{2D}$ is a 2D covariance matrix calculated by $\mathbf{\Sigma}_k$.

Next, 3DGS calculates the color $\mathbf{C}(\boldsymbol{x})$ of pixels via alpha blending according to the primitive's depth order.

$$\mathbf{C}(\boldsymbol{x}) = \sum_{k=1}^{K} w_k \mathbf{c}_k \tag{2}$$

$$w_k = \alpha_k g_k^{2D}(\boldsymbol{x}) \prod_{j=1}^{k-1} \left(1 - \alpha_j g_j^{2D}(\boldsymbol{x})\right) \tag{3}$$

among them, $\mathbf{c}_k$ is a view-dependent color decoded from spherical harmonic parameter [34] $\mathbf{SH}$.

Finally, 3DGS optimizes the parameters contained in the Gaussian primitive through photometric and SSIM mixed loss:

$$\mathcal{L}_{gs} = \lambda_{\text{s}}\,\text{SSIM}(\mathbf{C}, \mathbf{C}_{\boldsymbol{gt}}) + (1 - \lambda_{\text{s}})\,\|\mathbf{C} - \mathbf{C}_{\boldsymbol{gt}}\|_1 \tag{4}$$

where $\mathbf{C}$ is the rendered color, $\mathbf{C}_{\boldsymbol{gt}}$ is the ground-truth color, and SSIM [40] is the structural similarity function.

Anti-noise Reconstruction. In order to prevent the impact of noise labels on reconstruction, the usual anti-noise reconstruction method [15, 31] estimates a clean mask M_c to filter the reconstruction loss:

$$\mathcal{L}_{final} = M_c \mathcal{L}_{gs} \tag{5}$$

In this paper, we also use M_c to achieve anti-noise reconstruction. Compared to the uncertainty loss [12, 29] used in NeRF, a masking scheme with accurate noise segmentation can achieve more accurate reconstruction results.

3.2. Observation Completeness

This section proposes a calculation and rendering method for observation completeness (OC), which is an indicator constructed based on the principle of triangulation. It describes whether Gaussian primitives have obtained sufficient observations to ensure correct reconstruction. Specifically, the observation completeness metric O_i^n of the Gaussian primitive g_i at the nth iteration is defined as follows:

$$O_i^n = \lambda_{oc} O_i^{n-1} + (1 - \lambda_{oc})\delta_i^n \tag{6}$$

$$\delta_i = \hat{\sigma}^i(n) u_i^n \tag{7}$$

$$u_i^n = \begin{cases} 1 & Gradp(g_i^n) > 1 \times 10^{-7} \\ 0 & \text{otherwise} \end{cases} \tag{8}$$

where $\lambda_{oc} = 0.98$, $\hat{\sigma}^i(n)$ represents the positional variance between cameras observing the Gaussian primitive g_i, which reveals whether the observation perspective of g_i is sufficient. u_i^n determines whether the Gaussian primitive g_i is effectively observed in the current training frame, and we set u_i^n to 1 when the positional gradient $Gradp(g_i^n)$ is greater enough.

In Eq.7, u_i^n can be directly obtained during a single training iteration, but $\hat{\sigma}^i(n)$ requires computing the variance of historical camera positions, which significantly increases computational resource consumption. Therefore, we use an iterative formula for variance to compute $\hat{\sigma}^i(n)$:

$$\hat{\mu}_m^i = \hat{\mu}_{m-1}^i + \frac{1}{m}\left(T_n - \hat{\mu}_{m-1}^i\right) \tag{9}$$

$$\hat{\sigma}_m^i = \left(\frac{m-2}{m-1}\right)\hat{\sigma}_{m-1}^i + \frac{1}{m}\left(T_n - \hat{\mu}_{m-1}^i\right)^2 \tag{10}$$

$$m = m + u_i^n \tag{11}$$

where $\hat{\sigma}^i(n) = \|\hat{\sigma}_m^i\|_2$, Eq.9, Eq.10 and Eq.11 are executed if and only if $u_i^n = 1$, so m is not greater than n. T_n is the spatial position of the camera in the nth iteration.

Rendering of OC. Referring to the color rendering in Eq. 2, the rendering for OC is as follows:

$$\mathbf{O}(\boldsymbol{x}) = \sum_{k=1}^{K} w_k O_k \tag{12}$$

3.3. Hybrid Noise Assessment and Correction

We propose a hybrid noise evaluation scheme consisting of residual $\mathcal{R}$ and uncertainty β. As shown in Fig. 3, $\mathcal{R}$ reflects the significant noise present in the current training stage, while β reflects the persistent interference. However, $\mathcal{R}$ is difficult to detect weaker noise (foot shadows), and β cannot distinguish difficult-to-reconstruct areas (distant trees), all of which interfere with accurate noise separation. Our proposed simple and effective hybrid assessment $\mathcal{A}_h$ can avoid these drawbacks:

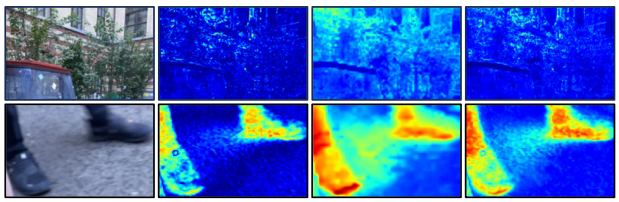

Figure 3. **Hybrid Noise Assessment.** Hybrid assessment can reduce additional noise caused by reconstruction difficulties (up) while focusing on persistent little errors (down).

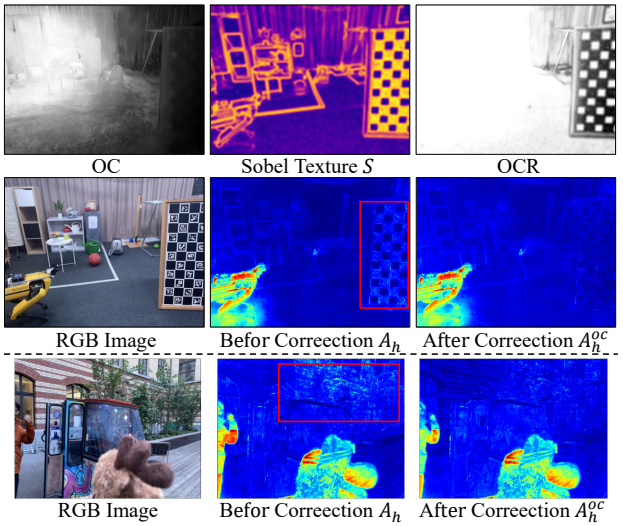

Figure 4. **Observation Correction.** The observation correction ratio OCR is determined by considering the completeness of observations $\mathbf{O}$ in a certain region and its reconstruction difficulty S.

$$\mathcal{A}_h = \lambda_2 \mathcal{R} + (1 - \lambda_2)\beta \tag{13}$$

where $\lambda_2 = 0.5$, $\mathcal{R} = \|\mathbf{C} - \mathbf{C}_{\boldsymbol{gt}}\|_1$. As shown in Fig. 3, the hybrid assessment enhances the error region shared by the residuals $\mathcal{R}$ and β, indirectly suppressing the interference caused by learning difficulties in β. We elaborated on the calculation process of β in detail in Sec 3.4.

Observation Cognitive Correction. The example in Fig. 1 shows that in 3DGS reconstruction, there is a high correlation between the number of observations and the reconstruction error. Therefore, we further correct the label noise assessment using the observation completeness proposed in Eq.12:

$$\mathcal{A}_h^{oc} = \mathcal{A}_h \cdot OCR \tag{14}$$

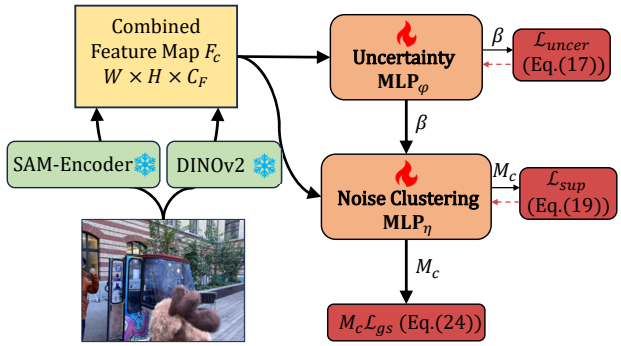

Figure 5. **Label Noise classification Pipeline.** The pre-trained feature extractor (DINOv2,SAM) extracts feature maps from the training images. Then, the uncertainty $\mathbf{MLP}_\psi$ and noise classifier $\mathbf{MLP}_\eta$ predict the uncertainty β and static background mask M_c from the feature map. Three losses (right side) are used to optimize $\mathbf{MLP}_\psi$, $\mathbf{MLP}_\eta$, and 3DGS model.

$$OCR = 1 - \lambda_3(th - \mathbf{O})S \tag{15}$$

where OCR represents the observation correction ratio, with its range constrained between 0 and 1. λ_3 and th are hyperparameters set to 3.0 and 0.3, respectively. $\mathbf{O}$ denotes the observation completeness, whose range is constrained between 0 and th. S represents the texture intensity, calculated by convolving the image with the Sobel [5] operator.

The process of correcting the observation is shown in the upper part of Fig. 4. After correcting based on observation completeness $\mathbf{O}$, the errors caused by the lack of observation (red box in Fig. 4) in $\mathcal{A}_h$ are significantly suppressed. This enables subsequent classifiers to learn the correct static scene mask more robustly from $\mathcal{A}_h^{oc}$.

3.4. Label Noise Classifier

As shown in Fig. 5, we train two MLPs to learn clean mask M_c and uncertainty β from image feature encoding. One significant advantage of using an MLP for learning is that predicting masks from features with clear semantics enables a better understanding of chunked semantic regions, further avoiding fragmented noise.

Feature Extraction. We use SAM[14] and DINOv2[25] with frozen weights to extract features from the training images. Then, we scale the features to the original image size and concatenate them together to obtain F_c. Moreover, we also concatenated position encoding F_{pos} on F_c to assist MLP in learning uncertainty in low feature response areas.

Learning Uncertainty. Unlike the Nerf methods [22, 29], introducing uncertainty β will greatly affect the splitting and pruning of 3D Gaussian primitive, leading to training collapse. Therefore, we blocked the gradient feedback related to Gaussian primitive parameters, and the final uncertainty loss $\mathcal{L}_{uncer}$ is as follows:

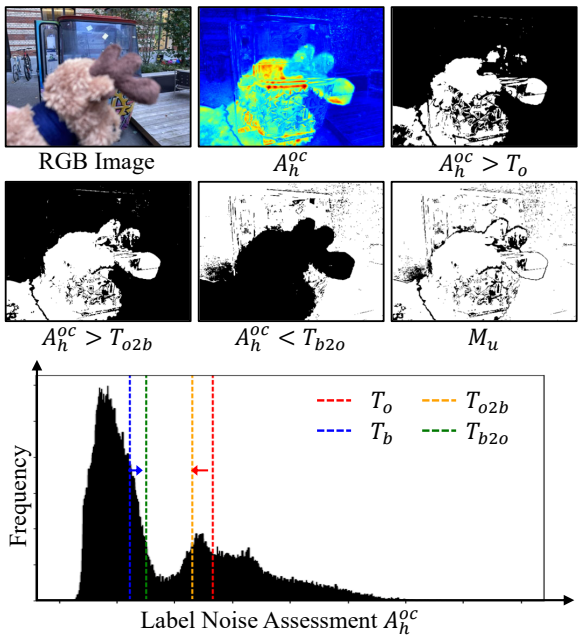

Figure 6. **Dynamic Threshold Clustering.** Using two dynamic anchors, the maximum inter-class variance (red) and the background noise mean (blue), we can stably calculate the threshold intervals for foreground and background.

$$\beta = \mathbf{MLP}_\psi(F_{pos}, F_{id}, F_c) \tag{16}$$

$$\mathcal{L}_{uncer} = \frac{Detach(\mathcal{L}_{gs})}{2\beta^2} + \log\beta \tag{17}$$

where $Detach$ is a gradient blocker (Similar to $.detach()$ in pytorch [27]), F_{id} is the image number encoding, ψ is the network parameter to be learned.

Learning Noise Classifier. We train a noise classifier $\mathbf{MLP}_\eta$ by generating self-supervised labels from $\mathcal{A}_h^{oc}$:

$$M_c = \mathbf{MLP}_\eta(F_{pos}, \beta, F_c) \tag{18}$$

$$\mathcal{L}_{sup} = M_u \|M_c - M_{self}\|_1 \tag{19}$$

among them, M_{self} and M_u are the self-supervised labels and effective training regions, respectively. Next, we introduce how to calculate the labels required for self-supervision from $\mathcal{A}_h^{oc}$.

Self-supervised Labels. We propose an dynamic threshold anchor method to handle various scenarios. Specifically, we first convert $\mathcal{A}_h^{oc}$ into a histogram to analyze the occurrence frequencies of noise with different magnitudes statistically. Then, we compute the maximum inter-class variance threshold T_o and the mean of the background noise

Table 1. **Ablation Study.** We conducted ablation in different scenes, including indoor scenes with moderate (*Statue*) and high (*Spot*) occlusion, and outdoor scene (*Fountain*) with less occlusion but containing reflections.

		Statue		*Fountain*		*Spot*		*Mean*	
	ID	PSNR↑	SSIM↑	PSNR↑	SSIM↑	PSNR↑	SSIM↑	PSNR↑	SSIM↑
Ours	**A**	23.13	0.863	**23.05**	**0.794**	**26.32**	**0.902**	**24.17**	**0.853**
only uncertainty β	**B**	22.85	0.862	19.97	0.691	26.31	0.901	23.04	0.818
only residuals $\mathcal{R}$	**C**	**23.16**	**0.867**	23.01	0.782	25.21	0.899	23.79	0.849
w/o OCP	**D**	23.14	0.863	22.78	0.792	26.24	0.901	24.05	0.852
w/o OCC	**E**	23.03	0.865	22.85	0.794	25.72	0.901	23.86	0.853
fixed threshold	**F**	22.90	0.862	22.87	0.791	25.60	0.899	23.78	0.850

T_b as dynamic classification anchors. Finally, the self-supervised labels required for training are calculated using a weighted combination of these dynamic anchors:

$$T_{b2o} = (1 - \lambda_4)T_b + \lambda_4 T_o \tag{20}$$

$$T_{o2b} = \lambda_5 T_b + (1 - \lambda_5)T_o \tag{21}$$

$$M_{self} = \mathcal{A}_h^{oc} < T_{b2o} \tag{22}$$

$$M_u = (\mathcal{A}_h^{oc} < T_{b2o}) \vee (\mathcal{A}_h^{oc} > T_{o2b}) \tag{23}$$

where T_{b2o} and T_{o2b} are the fine-tuned foreground and background thresholds, M_{self} is the generated self-supervised label, and M_u is the effective training area. Please refer to the **supplementary materials** for the specific calculation process of T_b and T_o.

Fig. 6 shows the complete process of label generation. By leveraging M_u, the MLP focuses on learning the more prominent errors while minimizing the interference caused by ambiguous intermediate regions during training. Moreover, the maximum inter-class variance threshold demonstrates stable performance in the noise classification task proposed in this paper. Through experiments in Tab. 1 and Tab. 3, we show that, compared to fixed-threshold classification, our dynamic thresholding approach can seamlessly adapt to most scenarios without parameter adjustments.

Parameter Optimization. Finally, we optimize our OCSplats through the following loss:

$$\mathcal{L}_{all} = M_c \mathcal{L}_{gs} + \mathcal{L}_{uncer} + \mathcal{L}_{sup} \tag{24}$$

3.5. Observation Completeness Pruning (OCP)

After visualizing the observation completeness, we also found an interesting phenomenon. As shown in Fig. 7, the observation completeness of floating Gaussians (floaters) was extremely low. It is easy to understand. The fundamental reason why the floating Gaussian is not optimized during training is that it is only represented in a small number of viewing angles and invisible in the vast majority of viewing angles. Based on the above observations, we propose a pruning strategy based on observation completeness. Specifically, during the training process, we directly eliminate those Gaussians whose observation completeness is less than 0.03 ($O_i^n < 0.03$) and whose number of observations is less than 3 (during a full round of training).

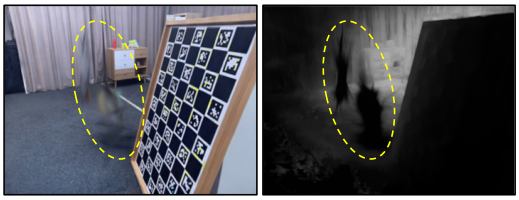

Figure 7. **Observation Completeness Pruning (OCP).** The OC values of floating objects are often very low, making it easy to distinguish them from normal pixels.

4. Experiment

In this section, we first conduct ablation experiments on the key modules proposed in this paper to verify their effectiveness. Next, we compare OCSplats with SOTA methods on interference datasets.

Datasets. We mainly evaluated our method on two datasets, RobustNeRF [30] and On-the-go [29]. The RobustNeRF consists of four scenes of toys-on-the-tables, with low occlusion rates for all scenes except the *Statue*. The On-the-go dataset consists of outdoor and indoor scenes. Referring to previous work [29], we selected 6 sequences in descending order of occlusion rate. In addition, compared to the RobustNeRF dataset, the On-the-go dataset typically has lower observational completeness.

Implementation Details. We stacked the key modules proposed in this paper on the standard 3DGS method. The difference is that our OCSplats run OCP and Gaussian pruning clone every z steps (z is the length of the scene).

Table 2. **Evaluation on RobustNeRF Dataset.** We presented a quantitative comparison between our method and the baseline method, with the upper half being based on the NeRF method and the lower half being based on the 3DGS method. The best method between classes is displayed in bold.

	Android			*Crab*			*Yoda*			*Statue*		
	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓
Mip-NeRF360 [3]	21.81	0.695	0.176	29.25	0.918	0.086	23.75	0.770	0.216	19.86	0.690	0.233
NeRF-W [22]	20.62	0.664	0.258	26.91	0.866	0.157	28.64	0.752	0.260	18.91	0.616	0.369
RobustNeRF [30]	23.28	0.755	0.130	32.22	0.945	**0.060**	29.78	0.821	0.150	20.60	0.758	0.150
NeRF-HuGS [4]	23.32	0.763	0.200	34.16	0.956	0.070	30.70	0.834	0.220	21.00	0.774	0.180
SpotLessSplats [31]	24.83	0.826	0.085	34.43	0.953	0.079	35.11	0.956	0.073	22.24	0.828	0.136
3DGS [13]	22.65	0.810	0.130	31.13	0.936	0.102	26.94	0.910	0.140	21.02	0.810	0.160
OCSplats (**Ours**)	**25.17**	**0.843**	**0.071**	**35.21**	**0.958**	0.068	**35.37**	**0.961**	**0.069**	**23.13**	**0.863**	**0.101**

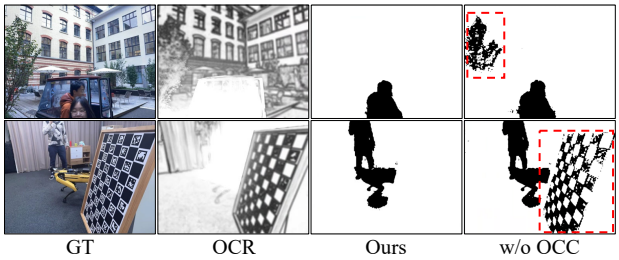

Figure 8. **Ablation Study about OCC.** We visualized and ablated the observation completeness correction (OCC) module. After using OCC, the noise classification performance in areas with fewer observations was significantly improved.

4.1. Ablation Study

Hybrid Assessment. To validate the effectiveness of the hybrid assessment, we separately tested the performance of residual $\mathcal{R}$ (**C**) and uncertainty β (**B**). As shown in Tab. 1, when only residual $\mathcal{R}$ was used, the model performed slightly worse in the shadow-heavy *Spot* scenario. In parallel, when only uncertainty β was applied, the model experienced a performance collapse in the *Fountain* (have many non-Lambertian surfaces) scene (23.05 to 19.97). On the contrary, our hybrid assessment (**A**) achieved the optimal average performance, which proves that our hybrid strategy can preserve the advantages of each indicator as much as possible while reducing the impact of its defects on overall performance.

Observation Completeness Pruning (OCP). Unlike SplotLessSplats [31] that crudely removed low-gradient Gaussian elements (which resulted in performance degradation), OCP removes Gaussian primitives that cannot be reconstructed correctly from the triangulation perspective, which can further improve overall performance while pruning. As shown in Tab. 1 (**D**), the average performance decreased from 24.17 to 24.05 after removing OCP pruning.

Observation Completeness Correction (OCC). As shown in Tab. 1 (**E**), the average performance significantly decreased after further ablation of OCC. Moreover, OCC is of great significance for separating label noise. As shown in Fig. 8 and Fig. 10, using OCC can significantly improve the noise separation effect and rendering performance in areas with missing observations.

Dynamic Threshold. To verify the effectiveness of the dynamic threshold in Sec.3.4, we attempted a fixed threshold version based on Goli et al.'s work [31]. Specifically, we consider the pixels with the last 50% error as static background and those with the first 10% as noise foreground. As shown in Tab. 1 (**F**), the average PSNR decreased from 24.17 to 23.78, and the performance degradation was more pronounced in high-occlusion scenes (*Spot*). This demonstrates the importance of our dynamic thresholding strategy for anti-noise reconstruction in complex scenes.

4.2. Evaluation on RobustNeRF Dataset

As shown in Tab. 2, our method achieves leading performance in almost all scenarios. It is worth noting that, since *Android*, *Crab*, and *Yoda* belong to small-scale scenarios with higher observation frequencies, the advantages of our method are less pronounced. However, in large-scale and heavily occluded scenarios such as *Statue*, our method demonstrates significantly greater advantages. As shown in the bottom of Fig. 9, thanks to the observation cognitive correction and the automatic thresholding, our method handles regions with incomplete observations (red circle) and high occlusion more effectively compared to previous SOTA methods.

4.3. Evaluation on On-the-go Dataset

As shown in Tab. 3, our method maintains leading performance in both high occlusion and low occlusion scenarios. The performance improvement is more significant in high occlusion scenes, which proves that our method can better handle complex, noisy scenes. As shown in the top of Fig. 9, when only using the default fixed thresh-

Table 3. **Evaluation on On-the-go Dataset.** The **SpotLessSplasts** in the table are the results we reproduced; We used default fixed thresholds in all scenarios.

	Low Occlusion				Medium Occlusion				High Occlusion			
	Mountain		*Fountain*		*Corner*		*Patio*		*Spot*		*Patio high*	
	PSNR↑	SSIM↑	PSNR↑	SSIM↑	PSNR↑	SSIM↑	PSNR↑	SSIM↑	PSNR↑	SSIM↑	PSNR↑	SSIM↑
Mip-NeRF360 [3]	19.64	0.601	13.91	0.290	20.41	0.660	15.48	0.503	17.82	0.306	15.73	0.432
NeRF-W [22]	18.07	0.492	17.20	0.410	20.21	0.708	17.55	0.532	16.40	0.384	12.99	0.349
NeRFOn-the-go [29]	20.15	0.644	20.11	0.609	24.22	0.806	20.78	0.754	23.33	0.787	21.41	0.718
RobustNeRF [30]	17.54	0.496	15.65	0.318	23.04	0.764	20.39	0.718	20.65	0.625	20.54	0.578
Wild-GS [15]	20.92	0.673	20.94	0.672	23.69	0.815	21.23	0.805	23.93	0.781	22.11	0.737
3DGS [13]	20.98	0.709	21.46	0.728	22.54	0.800	17.13	0.671	18.24	0.657	15.74	0.515
SpotLessSplats [31]	21.39	0.708	22.62	0.761	26.09	0.877	22.28	0.823	24.15	0.819	22.27	0.771
OCSplats(**Ours**)	**21.45**	**0.738**	**23.05**	**0.794**	**26.74**	**0.904**	**22.45**	**0.841**	**26.32**	**0.902**	**23.61**	**0.846**

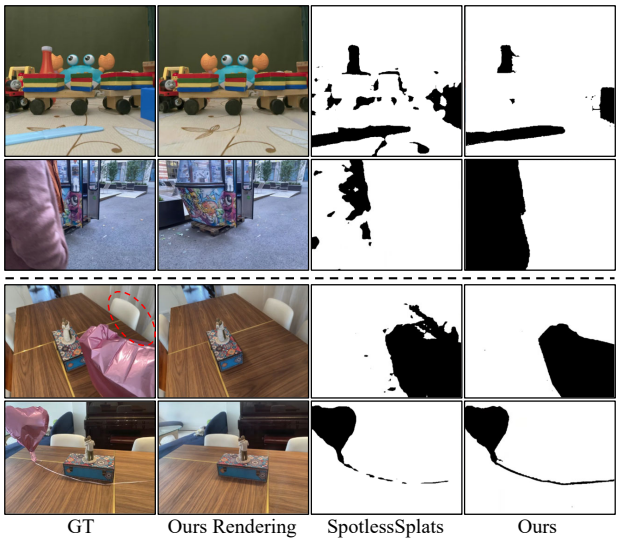

Figure 9. **Noise Label Segmentation.** Compared to the previous method of using a fixed threshold, our approach can robustly segment the correct static background in various complex scenarios.

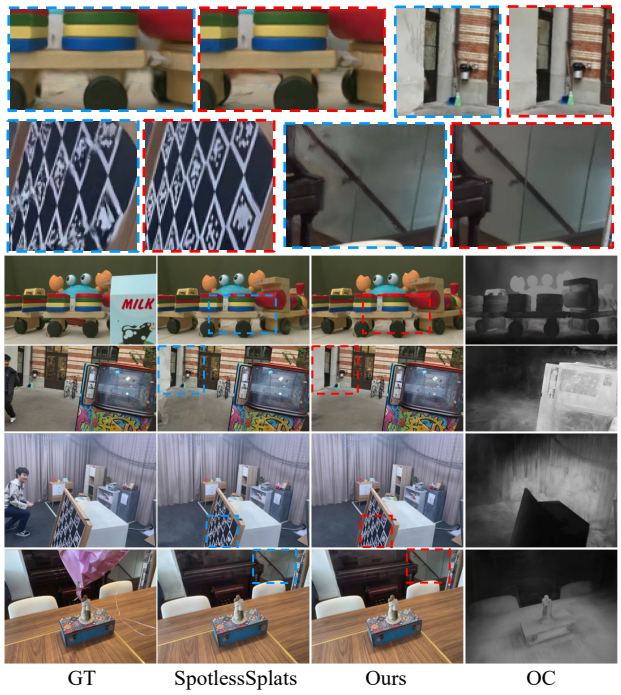

Figure 10. **Visual Comparison of Rendering Results.** Compared with the state-of-the-art method (Blue box), our approach (Red box) achieved better rendering results in areas lacking observation.

old, SplotLessSplats removes too many pixels in low-occlusion scene (*Crab*) and appears too conservative in high-occlusion scene (*Patio high*). On the contrary, our dynamic threshold strategy can effectively handle scenes with different occlusion rates.

5. Conclusion

This paper proposes a robust reconstruction framework OCSplasts based on observation completeness and 3DGS, which is used to separate noise elements in scenes containing complex noise and reconstruct clean 3DGS scenes. Our method overcomes the limitation of poor generalization of existing methods. It proposes for the first time a label noise measurement correction theory based on observation completeness, as well as a dynamic anchor threshold calculation method, achieving robust reconstruction of multiple complex scenes without adjusting any hyperparameters. Moreover, the observed completeness obtained through quantification can also be used to determine whether the observations obtained for each region in the scene are sufficient to guide subsequent shooting.

Acknowledgements National Natural Science Foundation of China (Grant nos.62372235,62406069); China Postdoctoral Science Foundation(2024M750425).

References

[1] Sanjay Bakshi and Yee-Hong Yang. Shape from shading for non-lambertian surfaces. In *Proceedings of 1st International Conference on Image Processing*, pages 130–134. IEEE, 1994. 1

[2] Riccardo Barbano, Željko Kereta, Chen Zhang, Andreas Hauptmann, Simon Arridge, and Bangti Jin. Quantifying sources of uncertainty in deep learning-based image reconstruction. *arXiv preprint arXiv:2011.08413*, 2020. 2

[3] Jonathan T Barron, Ben Mildenhall, Dor Verbin, Pratul P Srinivasan, and Peter Hedman. Mip-nerf 360: Unbounded anti-aliased neural radiance fields. In *CVPR*, pages 5470–5479, 2022. 2, 7, 8

[4] Jiahao Chen, Yipeng Qin, Lingjie Liu, Jiangbo Lu, and Guanbin Li. Nerf-hugs: Improved neural radiance fields in non-static scenes using heuristics-guided segmentation. In *CVPR*, pages 19436–19446, 2024. 1, 7

[5] Wenshuo Gao, Xiaoguang Zhang, Lei Yang, and Huizhong Liu. An improved sobel edge detection. In *2010 3rd International conference on computer science and information technology*, pages 67–71. IEEE, 2010. 5

[6] Lily Goli, Cody Reading, Silvia Sellán, Alec Jacobson, and Andrea Tagliasacchi. Bayes' rays: Uncertainty quantification for neural radiance fields. In *CVPR*, pages 20061–20070, 2024. 2

[7] Huasong Han, Kaixuan Zhou, Xiaoxiao Long, Yusen Wang, and Chunxia Xiao. Ggs: Generalizable gaussian splatting for lane switching in autonomous driving. *arXiv preprint arXiv:2409.02382*, 2024. 1

[8] Richard I Hartley and Peter Sturm. Triangulation. *Computer vision and image understanding*, 68(2):146–157, 1997. 2

[9] Georg Hess, Carl Lindström, Maryam Fatemi, Christoffer Petersson, and Lennart Svensson. Splatad: Real-time lidar and camera rendering with 3d gaussian splatting for autonomous driving. *arXiv preprint arXiv:2411.16816*, 2024. 1

[10] Eyke Hüllermeier and Willem Waegeman. Aleatoric and epistemic uncertainty in machine learning: An introduction to concepts and methods. *Machine learning*, 110(3):457–506, 2021. 2

[11] John Jakeman, Michael Eldred, and Dongbin Xiu. Numerical approach for quantification of epistemic uncertainty. *Journal of Computational Physics*, 229(12):4648–4663, 2010. 2

[12] Alex Kendall and Yarin Gal. What uncertainties do we need in bayesian deep learning for computer vision? *NIPS*, 30, 2017. 2, 3

[13] Bernhard Kerbl, Georgios Kopanas, Thomas Leimkühler, and George Drettakis. 3d gaussian splatting for real-time radiance field rendering. *ACM Trans. Graph.*, 42(4):139–1, 2023. 1, 2, 7, 8

[14] Alexander Kirillov, Eric Mintun, Nikhila Ravi, Hanzi Mao, Chloe Rolland, Laura Gustafson, Tete Xiao, Spencer Whitehead, Alexander C Berg, Wan-Yen Lo, et al. Segment anything. In *ICCV*, pages 4015–4026, 2023. 5

[15] Jonas Kulhanek, Songyou Peng, Zuzana Kukelova, Marc Pollefeys, and Torsten Sattler. Wildgaussians: 3d gaussian splatting in the wild. *arXiv preprint arXiv:2407.08447*, 2024. 1, 2, 3, 8

[16] Hao Li, Jingfeng Li, Dingwen Zhang, Chenming Wu, Jieqi Shi, Chen Zhao, Haocheng Feng, Errui Ding, Jingdong Wang, and Junwei Han. Vdg: vision-only dynamic gaussian for driving simulation. *arXiv preprint arXiv:2406.18198*, 2024. 1

[17] Han Ling, Quansen Sun, Zhenwen Ren, Yazhou Liu, Hongyuan Wang, and Zichen Wang. Scale-flow: Estimating 3d motion from video. In *ACMMM*, pages 6530–6538, 2022. 1

[18] Han Ling, Yinghui Sun, Quansen Sun, and Zhenwen Ren. Learning optical expansion from scale matching. In *CVPR*, pages 5445–5454, 2023. 1

[19] Han Ling, Quansen Sun, Yinghui Sun, Xian Xu, and Xinfeng Li. Adfactory: An effective framework for generalizing optical flow with nerf. In *CVPR*, pages 20591–20600, 2024. 1

[20] Han Ling, Yinghui Sun, Quansen Sun, Ivor Tsang, and Yuhui Zheng. Self-assessed generation: Trustworthy label generation for optical flow and stereo matching in real-world. *arXiv preprint arXiv:2410.10453*, 2024. 1

[21] Wesley J Maddox, Pavel Izmailov, Timur Garipov, Dmitry P Vetrov, and Andrew Gordon Wilson. A simple baseline for bayesian uncertainty in deep learning. *NIPS*, 32, 2019. 2

[22] Ricardo Martin-Brualla, Noha Radwan, Mehdi SM Sajjadi, Jonathan T Barron, Alexey Dosovitskiy, and Daniel Duckworth. Nerf in the wild: Neural radiance fields for unconstrained photo collections. In *CVPR*, pages 7210–7219, 2021. 5, 7, 8

[23] Dominik Narnhofer, Alexander Effland, Erich Kobler, Kerstin Hammernik, Florian Knoll, and Thomas Pock. Bayesian uncertainty estimation of learned variational mri reconstruction. *IEEE transactions on medical imaging*, 41(2):279–291, 2021. 2

[24] Vu-Linh Nguyen, Sébastien Destercke, and Eyke Hüllermeier. Epistemic uncertainty sampling. In *Discovery Science: 22nd International Conference, DS 2019, Split, Croatia, October 28–30, 2019, Proceedings 22*, pages 72–86. Springer, 2019. 2

[25] Maxime Oquab, Timothée Darcet, Théo Moutakanni, Huy Vo, Marc Szafraniec, Vasil Khalidov, Pierre Fernandez, Daniel Haziza, Francisco Massa, Alaaeldin El-Nouby, et al. Dinov2: Learning robust visual features without supervision. *arXiv preprint arXiv:2304.07193*, 2023. 5

[26] Nobuyuki Otsu et al. A threshold selection method from gray-level histograms. *Automatica*, 11(285-296):23–27, 1975. 2

[27] Adam Paszke, Sam Gross, Francisco Massa, Adam Lerer, James Bradbury, Gregory Chanan, Trevor Killeen, Zeming Lin, Natalia Gimelshein, Luca Antiga, et al. Pytorch: An imperative style, high-performance deep learning library. *NIPS*, 32, 2019. 5

[28] Zhiyin Qian, Shaofei Wang, Marko Mihajlovic, Andreas Geiger, and Siyu Tang. 3dgs-avatar: Animatable avatars via deformable 3d gaussian splatting. In *CVPR*, pages 5020–5030, 2024. 1

[29] Weining Ren, Zihan Zhu, Boyang Sun, Jiaqi Chen, Marc Pollefeys, and Songyou Peng. Nerf on-the-go: Exploiting uncertainty for distractor-free nerfs in the wild. In *CVPR*, pages 8931–8940, 2024. 1, 2, 3, 5, 6, 8

[30] Sara Sabour, Suhani Vora, Daniel Duckworth, Ivan Krasin, David J Fleet, and Andrea Tagliasacchi. Robustnerf: Ignoring distractors with robust losses. In *CVPR*, pages 20626–20636, 2023. 1, 2, 6, 7, 8

[31] Sara Sabour, Lily Goli, George Kopanas, Mark Matthews, Dmitry Lagun, Leonidas Guibas, Alec Jacobson, David J Fleet, and Andrea Tagliasacchi. Spotlesssplats: Ignoring distractors in 3d gaussian splatting. *arXiv preprint arXiv:2406.20055*, 2024. 1, 2, 3, 7, 8

[32] Johannes Lutz Schönberger and Jan-Michael Frahm. Structure-from-motion revisited. In *CVPR*, 2016.

[33] Johannes Lutz Schönberger, Enliang Zheng, Marc Pollefeys, and Jan-Michael Frahm. Pixelwise view selection for unstructured multi-view stereo. In *ECCV*, 2016.

[34] Robert T Seeley. Spherical harmonics. *The American Mathematical Monthly*, 73(4P2):115–121, 1966. 3

[35] Jianxiong Shen, Ruijie Ren, Adria Ruiz, and Francesc Moreno-Noguer. Estimating 3d uncertainty field: Quantifying uncertainty for neural radiance fields. In *ICRA*, pages 2375–2381. IEEE, 2024. 2

[36] Laura P Swiler, Thomas L Paez, and Randall L Mayes. Epistemic uncertainty quantification tutorial. In *Proceedings of the 27th International Modal Analysis Conference*, 2009. 2

[37] Veronica A Thurmond. The point of triangulation. *Journal of nursing scholarship*, 33(3):253–258, 2001. 2

[38] Fabio Tosi, Alessio Tonioni, Daniele De Gregorio, and Matteo Poggi. Nerf-supervised deep stereo. In *CVPR*, pages 855–866, 2023. 1

[39] Yuze Wang, Junyi Wang, and Yue Qi. We-gs: An in-the-wild efficient 3d gaussian representation for unconstrained photo collections. *arXiv preprint arXiv:2406.02407*, 2024. 1

[40] Zhou Wang, Alan C Bovik, Hamid R Sheikh, and Eero P Simoncelli. Image quality assessment: from error visibility to structural similarity. *TIP*, 13(4):600–612, 2004. 3

[41] Mulin Yu, Tao Lu, Linning Xu, Lihan Jiang, Yuanbo Xiangli, and Bo Dai. Gsdf: 3dgs meets sdf for improved rendering and reconstruction. *arXiv preprint arXiv:2403.16964*, 2024. 1

[42] 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*, pages 21634–21643, 2024. 1