EG UK Computer Graphics & Visual Computing (2026)
Y. Sheng and M. Elshehaly (Editors)

Information-Guided 3D Gaussian Splatting

Y. Dang[1], C. Amghane[2], V. Stenvers[1] and P. Vangorp[1]

[1]Utrecht University, Netherlands
[2]Royal Netherlands Aerospace Centre

Abstract

Creating 3D Gaussian splats of solely an object of interest (OOI) at the center of a scene is traditionally an ill-posed task, as standard pipelines optimize for global photometric loss across the entire environment. This requires either labor-intensive manual segmentation masks or computationally expensive pre-processing with foundational models to isolate the OOI [JMG24].
*We present an intrinsically-informed pipeline that leverages the underlying distribution and density of Gaussians initialized from Structure from Motion (SfM) to segment the OOI, including a prior-informed segmentation extension. This is to our knowledge the first method that allows 3D Gaussian splatting (3DGS) reconstruction of a specific object without foundational models. We evaluate this pipeline on a subset of the MiP-NeRF 360 [BMV*22] dataset, for which the ground truth segmentation masks have been manually created.*
The results demonstrate promising advances in isolating the OOI from the rest of the scene to reconstruct it in more detail. Despite challenges arising from the sparse nature of the initial SfM point cloud, the intrinsically-informed pipeline can effectively separate an OOI that is the single prominent object in the majority of the input images. The prior-informed approach using feature maps avoids the need for manual annotations but significantly increases computation time.

CCS Concepts
• ***Computing methodologies*** → *Reconstruction; 3D imaging; Image segmentation;*

1. Introduction

Current 3D reconstruction methods that aim to reconstruct solely an object within a scene rely on a fully-reconstructed scene, on which the object extraction is performed [GL24; YSG24]. This leads to inefficient use of computational power and time since the entire scene is reconstructed instead of only the object of interest (OOI). Additionally, reconstructing the entire scene reduces the overall reconstruction quality of the OOI due to the limited allocation of Gaussians for its representation [BRMG25].

We present an intrinsically-informed segmentation pipeline, designed to reconstruct an OOI within a 3D Gaussian splatting (3DGS) scene without requiring external foundational models. Unlike existing methods [FXZ*25] that first reconstruct entire scenes and extract the OOI post-training, this pipeline leverages the inherent spatial distribution and density of point clouds generated from Structure from Motion (SfM) to isolate the OOI early in the process. This work builds on the assumption that datasets focus on single objects, similar to the datasets in Mip-NeRF 360 [BMT*21]. By focusing computational resources exclusively on the OOI, this method reduces the processing time and enhances the fidelity of the reconstructed target.

The **intrinsically-informed** pipeline involves isolating the OOI within a point cloud and relies on the underlying distribution of Gaussians in 3DGS. This distribution inherently reflects the density of points, with higher densities corresponding to regions of greater detail, particularly around the OOI. SfM-generated point clouds, such as the one in Figure 3, naturally exhibit this property [SF16]. Firstly, SfM algorithms reconstruct 3D points by triangulating corresponding features across multiple images. OOIs tend to appear in a larger number of images and from various angles. This increased coverage results in more feature matches and, consequently, a higher density of reconstructed 3D points around the OOI. Secondly, OOIs are typically captured from closer distances, leading to higher image resolution and more detectable features in these areas. The closer proximity also allows for better feature matching across images, further increasing point density. Finally, OOIs often have more distinct textures and features, which are easier for SfM algorithms to detect and match consistently across images. In contrast, background elements or less important areas may be captured less frequently, from greater distances, or with less consistent angles, resulting in sparser point reconstruction.

Additionally, we evaluate a **prior-informed** pipeline which is an extension of the intrinsically-informed pipeline. The prior-informed pipeline does not rely on segmentation masks to extract the OOI. Instead, a single Gaussian is selected, the Reference Gaussian (RG), by comparing the semantic similarity between the Gaus-

© 2026 The Author(s).
Proceedings published by Eurographics - The European Association for Computer Graphics.
This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

sians and a text-encoded vector representing the OOI. The underlying rationale is that Gaussians in close proximity often model the same object, enabling efficient clustering into groups using methods such as k-means clustering. The prior-informed pipeline requires distilling semantic feature-vectors into the 3DGS representation but alleviates the need to segment the OOI in all input views.

We evaluate both pipeline variants using a manually masked dataset based on the Mip-NeRF 360 dataset [BMV*22].

This paper makes the following contributions:

- An information-guided segmentation pipeline, leveraging the variation in point cloud density where objects typically coincide with high density regions and background or noise with sparse regions.
- An optional prior-informed segmentation extension for the aforementioned pipeline, selecting Gaussians belonging to the OOI based on feature maps.
- A Mip-NeRF 360-based dataset of manually masked images, ensuring correct coverage of the ground truth, available at https://doi.org/10.17605/OSF.IO/5ZX8E.

2. Related Work

Techniques such as neural radiance field (NeRF) and 3DGS have proven to be capable of photorealistic 3D scene reconstruction. The information about the viewpoints, i.e. camera poses, of the input images is required to train these models. The camera poses are not always known a priori and sometimes have to be estimated using photogrammetric techniques. Structure from Motion (SfM) [SF16] is a photogrammetric technique to create 3D reconstructions and estimate camera parameters from a set of images. SfM is based on the principle of motion parallax, where objects in a scene appear to move differently based on their distance from the observer. The output of the SfM process is a dense 3D point cloud and the corresponding camera pose for each image, providing a solution when camera poses are unknown. However, even if the camera poses are known, ray-tracing based 3D reconstruction methods such as NeRF are computationally intensive due to the large number of volumetric scene elements that need to be accumulated for each pixel [GKJ*21].

Addressing the computational performance drawback of NeRF, Kerbl et al. proposed 3D Gaussian splatting [KKLD23], which contributes threefold: introduction of 3D Gaussians as a high-quality unstructured representation of radiance fields, optimization method of 3D Gaussian properties with adaptive density control, and a fast differentiable rendering approach for the GPU. Many parameters involved in the optimization method of the 3D gaussians are determined experimentally and based on heuristics. These heuristics are used to determine whether Gaussians should be added or removed. The performance of these heuristics depends on having a high-quality initial point cloud.

Kheradmand et al. [KRS*24] propose a new approach where 3DGS is reimagined as Markov Chain Monte Carlo (MCMC) samples drawn from an underlying probabilistic distribution that is proportional to how faithfully the Gaussians reconstruct the scene. Stochastic Gradient Langevin Dynamics (SGLD) is used to explore the scene and update the Gaussians in the densification process. This method renders the heuristics involved in the densification process (densifying, pruning, and resetting opacity) unnecessary. Gaussians serve as samples for MCMC, with their locations explored through SGLD updates. Changes to the set of Gaussians can be reformulated as deterministic state transitions. Kheradmand et al.'s improvements eliminate the dependency on a good initial point cloud, as the assumed underlying distribution is modeled by the Gaussians themselves.

Unlike the original 3DGS implementation [KKLD23], which dynamically adjusts the number of Gaussians in each scene, 3DGS Markov chain Monte Carlo (MCMC) [KRS*24] provides the ability to set a maximum number of Gaussians explicitly. This will make it easier to control and compare our proposed method with existing approaches, as the number of Gaussians used for reconstructing the OOI can be precisely defined.

2.1. Segmentation

Other works focusing on segmenting 3D reconstructions use large neural networks such as CLIP [RKH*21] to extract feature vectors for the images, which are then distilled into the neural radiance field [KKG*23]. In [KWK*24], the Segment Anything Model (SAM) [KMR*23] is used to extract semantic masks for the input views which are then used to learn a scaled affinity-field resulting in multi-view consistency on various scales. In [ZCJ*24], Gaussian Splatting is used as the scene representation. It utilizes feature-maps generated by SAM and distills these features into the gaussian scene. The segmentation relies on feature vectors or segmentation masks generated by models such as CLIP and SAM.

2.2. Meshing

A point cloud of a 3D scene consisting of various objects has a varying degree of information density. Some regions of the point cloud contain a large number of points compared to other regions of the point cloud. Density-Based Spatial Clustering of Applications with Noise (DBSCAN) [EKSX96; SSE*17] operates on the principle of density-based clustering where clusters are defined as dense regions of points separated by areas of lower point density. The separation of points in lower density regions is ideal for removing noise and outliers from a point cloud. In [Par62; DLP11] Kernel Density Estimation (KDE) estimates the probability density function of a random variable by placing a kernel function (typically a Gaussian distribution) at each data point and then summing their contributions in order to create a smooth and continuous estimate of the underlying density. Random Sample Consensus (RANSAC) [FB81] is a robust iterative method for estimating parameters of a mathematical model from a set of observed data that contains outliers. The method iteratively selects a random subset of the data and estimates the underlying model which is then compared to the entire dataset. The RANSAC method is also capable of fitting 2D planes in the dataset, making it highly applicable for wall and floor detection in human-made 3D environments. However, the iterative sampling approach is a limitation in large point clouds.

The Poisson Surface Reconstruction (PSR) [KBH06] method for generating watertight 3D surfaces from point cloud data formulates

© 2026 The Author(s).
Proceedings published by Eurographics - The European Association for Computer Graphics.

the problem as a Poisson equation. Its key strengths lie in its ability to handle noisy data and produce smooth, closed surfaces, efficiently handling large datasets and capturing fine details in point cloud with high density. PSR is effective at filling holes and completing missing data, which makes it robust to incomplete scans or sparse point clouds. However, this hole-filling property can sometimes lead to oversmoothing of sharp features or the creation of spurious surfaces in areas with sparse data.

2.3. Airplane Reconstruction

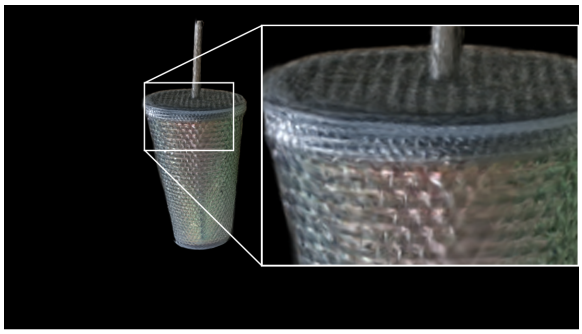
To accelerate aircraft inspections, the industry increasingly relies on semi-autonomous drones that fly around the airframe and capture high resolution images which are then processed into a report. While this method can reduce inspection time by up to 75% [Mai24], there are some limitations:

- **Lack of Depth Information**: Scratches or punctures in the hull have different severity levels, depending on the depth of the damage. Single images do not provide any depth information, making it difficult to assess the extent of the damage.
- **Single Viewpoint Constraint**: The inspection is limited to the angle from which the image was taken, potentially missing issues visible from other perspectives.
- **Human Validation**: Since each image corresponds to just one part of the aircraft, it is challenging for a human inspector to assess the aircraft as a whole. Stitching these images together often lacks coherence and fails to provide a seamless experience.

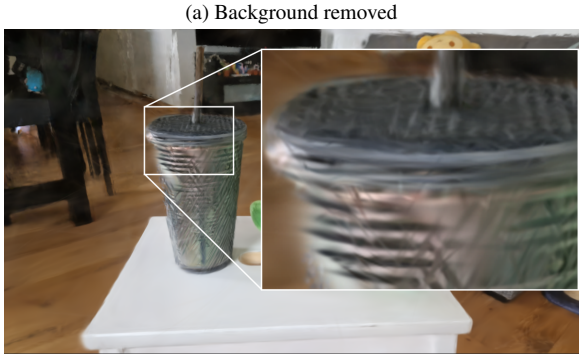
To summarize, NeRF and 3DGS are promising techniques for 3D reconstruction. In order to be useful for aircraft inspections the level of detail of the aircraft should be high enough to identify small cracks, dents and ruptures. One possible way to achieve this is to focus all reconstruction effort on the OOI and ignore the information in the periphery. This requires segmenting the 3D reconstruction to separate the OOI. Current work related to segmenting 3D reconstructions relies on large neural networks to infer information such as feature-vectors and segmentation masks which is used in the training process of the 3D reconstruction methods. This process is time consuming, requires manual input and is sensitive to the performance of neural networks such as CLIP. To overcome this limitation, we propose utilizing density information already present by nature in the point clouds to extract the object of interest and synthesize a high detail 3D reconstruction.

3. Naive Segmentation

In order to evaluate the efficacy of isolating an object of interest for 3D Gaussian splatting, we first compared a naive approach to object segmentation for 3DGS with a full scene reconstruction. The naive approach to creating a 3DGS reconstruction of solely the OOI is to create masks for the OOI in the input images and set the loss for all regions outside the OOI to zero. This approach was tested on a custom dataset [Stu22] to serve as a proof of concept and trained for 30,000 iterations. Figure 1 illustrates the difference in detail between the reconstruction with segmentation masks (a), where the background loss is set to zero, and the reconstruction without any modifications (b). It is evident that the detail on the reconstructed cup is visibly higher using the naive segmentation approach compared to the original method without modifications.



(a) Background removed



(b) Full scene reconstructed

Figure 1: 3D reconstruction of a custom scene with a cup [Stu22]. (a) Improved reconstruction quality with clear details on the cup when the background is not reconstructed. (b) Blurred or missing details on the cup when the full scene is reconstructed.

4. Intrinsically-Informed Segmentation Pipeline

This work builds on the assumption that datasets focus on single objects, similar to the datasets in Mip-NeRF 360.

A more adaptable method to isolate the OOI in 3D scene reconstruction involves clustering techniques. A 3D scene from SfM is represented as a point cloud, where higher detail corresponds to denser point regions. By using this property, clusters of high density can be identified, with the OOI expected to reside in one of these clusters.

This method serves as a preprocessing method for both the input images and the initial point cloud that is fed into any 3DGS method. It also offers advantages over feature-based 3DGS approaches such as Feature 3DGS [ZCJ*24] and SAGA [CFY*25], as it avoids reliance on complex foundational models. Since the underlying probability distribution is proportional to the faithfulness of the Gaussians reconstructing the scene and does not need to be modeled explicitly [KRS*24], this method can also work on 3DGS methods that do not rely on the initial SfM point cloud initialization.

An overview of the segmentation pipeline is shown in figure 2. The first three stages are responsible for extracting only the points belonging to the OOI. They were designed to gradually decrease the size of the point cloud, first removing larger chunks that are irrelevant to the OOI and, lastly, going to the finer points around the OOI, which are noisy points due to inaccuracies from the SfM pro-

© 2026 The Author(s).
Proceedings published by Eurographics - The European Association for Computer Graphics.

cess. The following meshing stage creates a surface reconstruction of the resulting OOI point cloud, essentially a 3D model of the OOI. This stage then allows projecting the mesh onto the input images to create masks of the OOI. The resulting masks and point cloud can subsequently be used as input images for any 3DGS method, following a similar approach to the naive approach outlined in Section 3.

The pipeline is intentionally designed in a modular fashion, with no strong dependencies between successive stages. Each stage exposes a well-defined input and output (a point cloud, a mesh, or a set of binary masks), so any individual stage can be substituted with an alternative technique without affecting the rest of the pipeline.

4.1. Density Clustering

Building on the inherent density-based distribution of points in SfM-generated point clouds, this initial stage of the pipeline aims to drastically downsample the scene. This is done by maximizing point removal while ensuring that all points belonging to the OOI are preserved, therefore maintaining the object's representation in the point cloud.

A Monte Carlo Kernel Density Estimation (KDE) approach is employed to quantify point density and enable OOI extraction. Instead of applying density estimation to the entire point cloud, which can be computationally expensive for large datasets, this method uses random subsampling. Multiple subsets of the point cloud are sampled, and a KDE model is fitted to each subset using a Gaussian kernel. The densities for all points in the original dataset are then estimated for each sampled subset, and the final density is computed as the average of these estimates.

Background objects, although having low-density values, often comprise a large number of points due to their extensive presence in the background of the scene. Therefore, filtering them out provides a solid base for a peak finding algorithm. The estimated density values are filtered to remove low-density regions, allowing a peak-finding algorithm to isolate the OOI correctly. The lowest 30% of density values are discarded, a threshold determined through experimental evaluation.

Through further experimentation, we found that filtering out the points before the first peak leaves the OOI intact and filters out the non-OOI points that are further away from the high-density region (the OOI and its immediate surroundings).

4.2. Floor Detection

The previous step does not completely remove the area around the OOI, leaving parts of the surrounding surface, such as the floor or a table, still present. These surfaces are not essential for reconstructing the OOI, but removing them can be challenging. If not done carefully, parts of the OOI close to the surface might also be removed, affecting the reconstruction's completeness. To address this challenge, the process is divided into three substages: detecting the floor, analyzing the model's orientation, and removing the floor.

The process of floor detection employs a Random Sample Consensus (RANSAC)-based algorithm to fit a plane model to the point cloud. Initially, a random subset of three points is selected to approximate the plane of the floor. The distance threshold, which defines the maximum allowable distance a point can be from the model to be considered an inlier (part of the floor), is then determined dynamically. The algorithm chooses the ideal distance threshold by trying 50 different thresholds and choosing the optimal threshold by the minimum value of the second derivative of the ratio between inliers and the total points. This threshold represents the transition where adding points provides minimal new information about the floor while minimizing the inclusion of extraneous points.

Orientation analysis of the detected floor plane is needed to ensure proper separation. The normal vector of the floor plane is computed to identify its orientation in the 3D space. Points above and below the plane are then segregated. Given that no part of the OOI should exist beneath the floor, the region with the fewest points is discarded. This ensures that only the points corresponding to the OOI remain, free of any extraneous points from the floor or the surrounding area.

Figure 5 shows a view of the OOI before and after removing the floor. Note that due to a non-optimal floor thickness, part of the OOI was removed as well. The resulting point cloud consists of the OOI, with the majority of the floor points removed.

4.3. Noise Removal

Feature extraction and matching in SfM pipelines are inherently imperfect and often introduce noise into the generated point cloud. As a result, noisy points in proximity to the OOI may remain in the dataset, complicating surface reconstruction on the point cloud and potentially degrading the quality of the reconstructed model.

To address this issue, the DBSCAN algorithm is employed to refine the filtered point cloud, grouping points into clusters based on their spatial proximity. DBSCAN also identifies outliers as noise, effectively removing isolated points that do not belong to any cluster. This dual capability makes it particularly suitable for noisy datasets. Following clustering, the resulting clusters are evaluated to determine the one most representative of the OOI. Typically, the largest cluster defined by the highest number of points is assumed to correspond to the OOI, as it reflects the region of greatest spatial coherence within the filtered point cloud.

Another noise removal method, statistical outlier removal, was also tested for this purpose. While this method mostly removes the noisy points around the OOI, similar to DBSCAN, some smaller clusters not part of the OOI still remain. A comparison with DBSCAN is shown in Figure 6.

The resulting point cloud can be used instead of the original SfM point cloud for initializing the Gaussians.

4.4. Meshing

To generate segmentation masks, the OOI point cloud must first be meshed into a continuous surface. The mesh is only used to produce the per-view masks consumed by 3DGS and is never exposed as a final output, so any closed-surface mesher works here and the

© 2026 The Author(s).
Proceedings published by Eurographics - The European Association for Computer Graphics.

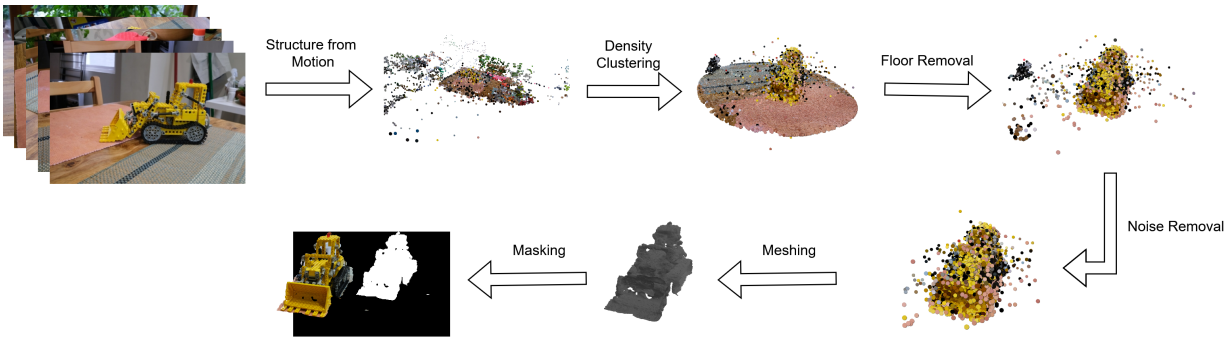


Figure 2: Pipeline for intrinsically-informed segmentation.

(a) (b)

Figure 3: (a) Point cloud generated from SfM is used as the input for the pipeline. (b) Zoomed in on the area around the OOI.

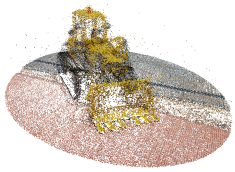


Figure 4: Extracted points from the OOI based on density, where low-density regions have been omitted.

(a) (b)

Figure 5: (a) Point cloud of OOI with floor, where the detected floor is colored red. (b) Same point cloud after floor removal. Note that the bottom of the tracks on the bulldozer has also been mostly removed.

(a) (b)

Figure 6: Noise removal from the OOI after floor separation, using (a) DBSCAN or (b) statistical outlier removal.

choice reduces to robustness and convenience. We therefore explored simple, widely available methods rather than the most recent ones: ball-pivot algorithm (BPA), Delaunay triangulation, and Poisson Surface Reconstruction (PSR). Of these, PSR proved the most suitable due to its robustness in handling noisy data.

Both BPA and Delaunay triangulation are sensitive to noise and require dense or highly uniform points [BTS*17; KA16], making them unsuitable for this task. By integrating smoothing into the reconstruction process, PSR effectively mitigates noise, although this comes at the cost of potentially losing finer geometric details.

However, a key limitation of PSR is its reliance on point normals. SfM point clouds often lack precomputed normals; meaning an additional preprocessing step to estimate them is necessary, which, especially for sparse point clouds, can yield incorrect normals.

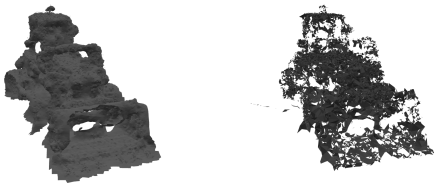
The normals are estimated using Open3D's [ZPK18] `estimate_normals` function. This function uses a k-d tree to identify the set of neighboring points for each point in the point cloud. The `KDTreeSearchParamHybrid` parameter controls how this neighborhood is determined: it considers all points within a given radius and ensures that at most `max_nn` neighbors are selected. A smaller radius captures fine local details but can be more sensitive to noise, while a larger radius smooths the normals over a broader area, which is beneficial for noisy or sparse data. Through experimentation, we found that the parameter choice did not affect the resulting normals for the scenes evaluated in this work. This is likely due to the point cloud of the OOI being dense enough.

For PSR, the depth parameter indirectly controls the level of detail in the final mesh. A higher depth allows for more detailed surfaces, but it also increases noise and computational demands. To

© 2026 The Author(s).
Proceedings published by Eurographics - The European Association for Computer Graphics.

balance detail preservation and noise suppression, a default depth value of 9 has been experimentally selected for all scenes. This depth provides sufficient detail to capture the geometry of the OOI while maintaining a degree of smoothing to address noise commonly present in SfM-generated point clouds.

As illustrated in figure 7, the mesh generated by the PSR method shows its robustness against noise. Specifically, the reconstructed surface effectively disregards the residual noisy points. Contrasting this, the BPA method produces a less consistent and noisy surface.



(a) (b)

Figure 7: Resulting meshes of the denoised OOI point cloud with (a) PSR and (b) BPA.

4.5. Masking

By applying meshing, a 3D segmentation of the OOI is effectively achieved. The meshed representation of the point cloud ensures consistency across all viewpoints, which can then be used to generate segmentation masks for the input images.

To achieve accurate mask generation, the positions and orientations of the original cameras must be replicated. Using the camera extrinsics and intrinsics derived from the SfM process, virtual cameras are placed in 3D space to precisely emulate the locations and orientations of the real-world cameras that captured the dataset. The meshed OOI is then rendered from the perspective of each emulated virtual camera, producing binary masks that delineate the OOI's silhouette as viewed from those angles.

During the meshing process, small gaps may arise in the reconstructed surface due to the inherent characteristics of PSR, such as smoothing or insufficient point density in certain areas. These gaps can result in incomplete coverage of the OOI in the generated masks. To address this, a morphological closing operation with a fixed kernel size, applied uniformly across all images and scenes, fills small isolated holes inside the silhouette without altering the mask boundary. This step is a minor convenience rather than a critical component of the pipeline.

The result of the masking process is shown in figure 8. The generated masks can subsequently be used as input images for 3DGS.

4.6. 3D Gaussian Splatting

To reconstruct solely the OOI, we integrate the generated masks into the 3DGS optimization pipeline. Specifically, we apply a pixel-wise binary mask to the photometric loss (comprising $\mathcal{L}_1$ and $\mathcal{L}_{SSIM}$), effectively zeroing the loss contribution from background

Figure 8: Projected mask on the input viewpoint to create a segmentation mask of the OOI.

regions. This prevents the model from allocating Gaussians to the environment and ensures the optimization is strictly confined to the OOI's geometry.

Additionally, the point clouds produced in Section 4.3 can be directly used for the Gaussian initialization and replace the SfM point cloud. This approach provides the advantage that the initialized Gaussians are constrained exclusively to the OOI and absent from the background.

5. Prior-informed Segmentation

Manually segmenting the OOI for all the input views is a time consuming process. We therefore opt for an approach utilizing semantic feature-vectors distilled into the Gaussians. The masking process is then reduced to selecting the OOI from only one viewpoint for each scene since the semantic feature-vectors are lifted from 2D to 3D.

However, this approach requires selecting a single Gaussian, the reference Gaussian (RG), that represents the OOI within a 3D Gaussian Scene. The RG is defined as the Gaussian with the highest semantic similarity to a text-encoded vector that represents the OOI. The RG acts as a reference point identifying all Gaussians that belong to the OOI. This method allows segmentation of the OOI using user inputs such as text descriptions or point prompts (similar to approaches like SAM [KMR*23]), standardizing the selection process and ensuring consistency across scenes.

Additionally, even if the similarity between the text vector and the most similar Gaussian is low, normalizing all comparisons relative to the RG allows the use of a fixed threshold for selecting OOI-related Gaussians. This normalization step ensures the method's robustness regardless of the scene's variability or the number of training iterations.

We leverage these advantages by expanding our existing pipeline with feature maps produced using the semantic similarity between the Gaussians and the text vector of the desired OOI. First, the text description of the OOI is encoded into a feature vector using CLIP [RKH*21], a model designed to connect visual and textual features by embedding both images and text into a shared semantic space. Then, since the Gaussians' semantic features are also based on CLIP, it is possible to compare the Gaussian's features to the text feature. The cosine similarity between the Gaussian and the text feature is calculated for each Gaussian. The Gaussian with the highest cosine similarity to the text feature is chosen as the RG, ensuring that the RG is the Gaussian most semantically aligned with the OOI. Finally, the Gaussians corresponding to the OOI are identified. This is done by comparing the Gaussians' features to the RG using the cosine similarity. This allows us to use a fixed threshold

© 2026 The Author(s).
Proceedings published by Eurographics - The European Association for Computer Graphics.

to identify all Gaussians belonging to the OOI, as cosine similarity is not transitive, and all comparisons are normalized relative to the RG.

Once the RG is identified, the labeled Gaussians segment the OOI. We adjusted the 3DGS Gaussian pruning function to remove not only low-opacity Gaussians but also those not labeled as part of the OOI. After a pruning step, the 3DGS model only consists of Gaussians belonging to the OOI.

Although semantic segmentation qualitatively isolates the OOI, optimizing against the original ground-truth images introduces boundary artifacts, such as dilated Gaussians at the edge of the OOI. To mitigate this, pixels in the ground-truth image not covered by the OOI mask are replaced with their corresponding values from the current render. By aligning the ground truth with the rendered background, the resulting gradients for background regions are zeroed, preventing them from influencing backpropagation.

However, this approach is limited by the inherent spatial extent of Gaussian primitives. Since each Gaussian represents a continuous distribution, its influence exceeds its mean position. When a Gaussian is classified as part of the OOI based on its center but its volume overlaps the background, the resulting boundary artifacts degrade the quality of the reconstruction along the edges of the OOI. This issue can be resolved by relying on segmentation masks derived directly from the input images rather than using rendered images. However, this approach ties the quality of the reconstruction directly to the accuracy of the input segmentation masks.

6. Manually Masked Dataset

We utilized the Mip-NeRF 360 [BMV*22] dataset to evaluate our segmentation pipeline. This dataset was chosen due to its compliance with a key requirement of the method under consideration: it must capture a complete 360-degree rotation around the OOI. While other datasets, such as Tanks and Temples [KPZK17], also provide 360-degree scenes, they are primarily designed for larger-scale environments.

However, to evaluate both OOI extraction from the point cloud and mask generation, ground truth data for the point cloud and segmentation masks are essential. While Liu et al. [LHTT24] evaluated their method using Mip-NeRF 360 by randomly selecting approximately 10 views and refining the generated masks with CascadePSP [CCTT20], this approach is insufficient for comprehensive evaluation. The generated masks, derived from semi-supervised algorithms, do not constitute true ground truth data, as they are still algorithmically created and may lack the accuracy needed for reliable benchmarking.

To address this limitation, we applied a more rigorous approach to the Mip-NeRF 360 dataset. Specifically, the kitchen scene has been fully manually masked for all 279 images to provide high-quality ground truth segmentation. For other scenes, 20 representative views have been selected and similarly manually masked. This ensures that the evaluation of mask generation is based on true ground truth data, significantly enhancing the reliability of the assessment.

To achieve accurate OOI representation by the point clouds, we manually removed extraneous or noisy points from the SfM-derived point clouds and refined the point clouds to only include the points belonging to the OOI. The resulting point clouds provide a robust baseline for evaluating OOI extraction and point cloud quality.

7. Results

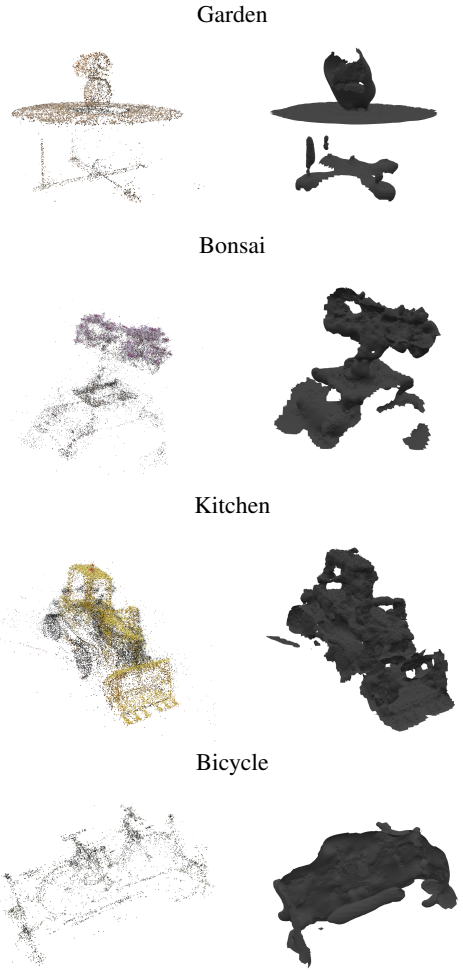


Figure 9: Point cloud results of each scene with the corresponding mesh results.

Our evaluation shows that relying on segmentation masks in conjunction with feature maps does not improve reconstruction quality. In fact, it introduces inefficiencies. While the reported quality scores for both methods are virtually the same, as shown in Table 1, the time it took for the feature map to finish its 30k iterations was significantly longer. The feature map-based approach was almost 3 times as slow as using only the segmentation mask.

Additionally, the quality of the reconstruction is ultimately constrained by the accuracy of the segmentation masks to select all Gaussians belonging to the OOI. Incorporating (threshold-based) clustering of the feature maps introduces even more uncertainty and room for error.

© 2026 The Author(s).
Proceedings published by Eurographics - The European Association for Computer Graphics.

A comparison of both methods is shown in figure 10. The metrics of the resulting masks from the intrinsically-informed segmentation pipeline in Table 2 suggest that while the intrinsically-informed approach is viable, its reliability remains tied to the geometric complexity and isolation of the target object. The sparse nature of the initial SfM point cloud makes it easy for background points to get trapped within the silhouette of objects with complex shapes.

	PSNR ↑	LPIPS ↓	SSIM ↑	Time (h:mm) ↓
Feature Map	23.05	0.102	0.75	1:42
Segmentation Mask	**23.12**	**0.091**	**0.81**	**0:35**

Table 1: Comparison of peak signal-to-noise ratio (PSNR), LPIPS, structural similarity index measure (SSIM), and runtime metrics for the prior-informed/feature map and intrinsically-informed/segmentation mask approaches.

Scene	IoU ↑	Precision ↑	Recall ↑	Dice Coeff ↑
Bicycle (20 images)	0.5549	0.5696	0.9592	0.7123
Bonsai (20 images)	0.6932	0.7205	0.9484	0.8175
Garden (20 images)	0.7706	0.7865	0.9728	0.8694
Kitchen (all images)	0.8646	0.9059	0.9501	0.9273
Average	0.7208	0.7456	0.9576	0.8316

Table 2: Binary segmentation mask metrics for Bicycle, Bonsai, Garden, and Kitchen scenes.

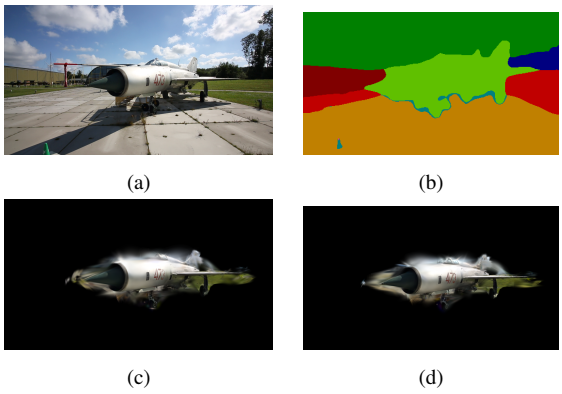


Figure 10: (a) Input image. (b) Segmentation mask. Rendered views of the 3DGS scene (c) with only selecting Gaussians belonging to the OOI via feature maps or (d) using the segmentation mask directly, without feature maps. Prior-informed segmentation using feature maps (c) does not improve over using masks directly (d).

8. Conclusion and Future Work

We presented an approach to segmentation within 3D Gaussian splatting, focusing on an intrinsically-informed method to perform object-specific reconstructions without reliance on pre-segmented data. Despite challenges arising from the sparse nature of the initial SfM point cloud, the results are promising: the object of interest was effectively separated from the rest of the scene and used to create a 3DGS reconstruction. Additionally, we investigated a prior-informed approach using feature maps. This approach significantly increased computation time, proving to be a slower proxy for directly using segmentation masks.

While the results demonstrate promising advances in isolating OOIs using density-based clustering, challenges remain in achieving full reconstruction coverage of the OOI.

One downside of our approach is its reliance on the OOI being the single prominent object in the majority of the input images. Datasets that focus on an OOI but do not capture it from all angles, such as forward-facing scenes in [MSO*19], are likely unsuitable for this method. A possible direction for future work could be to leverage the symmetrical properties of objects to mitigate these challenges.

The use of primarily sparse point clouds generated from SfM resulted in certain details of the OOI not being captured, as they were absent in the initial SfM point cloud. For instance, as shown in Figure 7b, the surface reconstruction using the BPA method results in a mesh with significant gaps. A denser point cloud would allow for a more complete surface reconstruction of the OOI while avoiding over-construction.

To address this limitation, future work could focus on incorporating the proposed segmentation and reconstruction methods directly into the Gaussian Splatting training loop. For example, a warm-up phase could be introduced, during which additional Gaussians are iteratively added to the scene, resulting in a denser point cloud. This adaptive approach could enhance the detail and fidelity of the reconstruction, ensuring that finer features of the OOI are adequately represented.

The current floor detection approach, based on RANSAC, may fail to align the detected plane with the actual floor. This misalignment can cause parts of the OOI to be incorrectly classified as floor inliers or can cause sections of the floor to be treated as outliers. A more robust floor detection method could mitigate these challenges. Techniques such as multi-plane RANSAC [LYLZ24], adaptive thresholding, or context-aware segmentation [WXYJ22] could improve the alignment and accuracy of the detected plane. Ensuring proper floor separation would significantly reduce the risk of erroneously removing parts of the OOI near the floor.

The current implementation of noise removal using DBSCAN is applied only once, which can leave residual noise that obstructs subsequent stages, such as meshing. Future improvements could involve adopting an iterative noise removal process with a dynamic stopping criterion or statistical outlier removal. By refining the noise removal step through successive iterations, the process could adapt to the dataset's characteristics, gradually eliminating more noise without compromising the integrity of the OOI. This iterative approach would likely result in cleaner point clouds, improving the quality of the final surface reconstruction and overall segmentation performance.

Acknowledgements This work was conducted during an internship at the Netherlands Aerospace Centre. We thank Thomas Bellucci and Mathijs Henquet for their support during this internship, and Frank van der Stappen for initiating this collaboration.

© 2026 The Author(s).
Proceedings published by Eurographics - The European Association for Computer Graphics.

References

[BMT*21] BARRON, JONATHAN T., MILDENHALL, BEN, TANCIK, MATTHEW, et al. "Mip-NeRF: A Multiscale Representation for Anti-Aliasing Neural Radiance Fields". *IEEE/CVF International Conference on Computer Vision*. 2021, 5835–5844. DOI: 10.1109/ICCV48922.2021.00580 1.

[BMV*22] BARRON, JONATHAN T., MILDENHALL, BEN, VERBIN, DOR, et al. "Mip-NeRF 360: Unbounded Anti-Aliased Neural Radiance Fields". *IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2022, 5470–5479. DOI: 10.1109/CVPR52688.2022.00539 1, 2, 7.

[BRMG25] BUI, QUOC-ANH, ROUGERON, GILLES, MORIN, GÉRALDINE, and GASPARINI, SIMONE. "ROI-GS: Interest-based Local Quality 3D Gaussian Splatting". *International Conference on Visual Communications and Image Processing*. 2025. DOI: 10.48550/arXiv.2510.01978 1.

[BTS*17] BERGER, MATTHEW, TAGLIASACCHI, ANDREA, SEVERSKY, LEE M., et al. "A Survey of Surface Reconstruction from Point Clouds". *Computer Graphics Forum* 36.1 (2017), 301–329. DOI: 10.1111/cgf.12802 5.

[CCTT20] CHENG, HO KEI, CHUNG, JIHOON, TAI, YU-WING, and TANG, CHI-KEUNG. "CascadePSP: Toward Class-Agnostic and Very High-Resolution Segmentation via Global and Local Refinement". *IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2020, 8890–8899. DOI: 10.1109/CVPR42600.2020.00891 7.

[CFY*25] CEN, JIAZHONG, FANG, JIEMIN, YANG, CHEN, et al. "Segment Any 3D Gaussians". *AAAI Conference on Artificial Intelligence*. Vol. 39. 2. 2025, 1971–1979. DOI: 10.1609/aaai.v39i2.32193 3.

[DLP11] DAVIS, RICHARD A., LII, KEH-SHIN, and POLITIS, DIMITRIS N. "Remarks on Some Nonparametric Estimates of a Density Function". *Selected Works of Murray Rosenblatt*. Springer, New York, NY, 2011, 95–100. DOI: 10.1007/978-1-4419-8339-8_13 2.

[EKSX96] ESTER, MARTIN, KRIEGEL, HANS-PETER, SANDER, JÖRG, and XU, XIAOWEI. "A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise". *International Conference on Knowledge Discovery and Data Mining*. 1996, 226–231 2.

[FB81] FISCHLER, MARTIN A. and BOLLES, ROBERT C. "Random Sample Consensus: A Paradigm for Model Fitting with Applications to Image Analysis and Automated Cartography". *Communications of the ACM* 24.6 (1981), 381–395. DOI: 10.1145/358669.358692 2.

[FXZ*25] FEI, BEN, XU, JINGYI, ZHANG, RUI, et al. "3D Gaussian Splatting as New Era: A Survey". *IEEE Transactions on Visualization and Computer Graphics* 31.8 (2025), 4429–4449. DOI: 10.1109/TVCG.2024.3397828 1.

[GKJ*21] GARBIN, STEPHAN J., KOWALSKI, MAREK, JOHNSON, MATTHEW, et al. "FastNeRF: High-Fidelity Neural Rendering at 200FPS". *IEEE/CVF International Conference on Computer Vision*. 2021, 14346–14355. DOI: 10.1109/ICCV48922.2021.01408 2.

[GL24] GUÉDON, ANTOINE and LEPETIT, VINCENT. "SuGaR: Surface-Aligned Gaussian Splatting for Efficient 3D Mesh Reconstruction and High-Quality Mesh Rendering". *IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2024, 5354–5363. DOI: 10.48550/arXiv.2311.12775 1.

[JMG24] JAIN, UMANGI, MIRZAEI, ASHKAN, and GILITSCHENSKI, IGOR. "GaussianCut: Interactive segmentation via graph cut for 3D Gaussian Splatting". *Advances in Neural Information Processing Systems*. Vol. 37. 2024, 89184–89212. DOI: 10.52202/079017-2830 1.

[KA16] KHATAMIAN, ALIREZA and ARABNIA, HAMID R. "Survey on 3D Surface Reconstruction". *Journal of Information Processing Systems* 12.3 (2016), 338–357. DOI: 10.3745/JIPS.01.0010 5.

[KBH06] KAZHDAN, MICHAEL, BOLITHO, MATTHEW, and HOPPE, HUGUES. "Poisson Surface Reconstruction". *Eurographics Symposium on Geometry Processing*. Vol. 7. 4. 2006, 61–70. DOI: 10.2312/SGP/SGP06/061-070 2.

[KKG*23] KERR, JUSTIN, KIM, CHUNG MIN, GOLDBERG, KEN, et al. "LERF: Language Embedded Radiance Fields". *IEEE/CVF International Conference on Computer Vision*. 2023, 19729–19739. DOI: 10.48550/arXiv.2303.09553 2.

[KKLD23] KERBL, BERNHARD, KOPANAS, GEORGIOS, LEIMKUEHLER, THOMAS, and DRETTAKIS, GEORGE. "3D Gaussian Splatting for Real-Time Radiance Field Rendering". *ACM Transactions on Graphics* 42.4 (2023), 139:1–14. DOI: 10.1145/3592433 2.

[KMR*23] KIRILLOV, ALEXANDER, MINTUN, ERIC, RAVI, NIKHILA, et al. "Segment Anything". *IEEE/CVF International Conference on Computer Vision*. 2023, 4015–4026. DOI: 10.1109/ICCV51070.2023.00371 2, 6.

[KPZK17] KNAPITSCH, ARNO, PARK, JAESIK, ZHOU, QIAN-YI, and KOLTUN, VLADLEN. "Tanks and Temples: Benchmarking Large-Scale Scene Reconstruction". *ACM Transactions on Graphics* 36.4 (2017), 78:1–13. DOI: 10.1145/3072959.3073599 7.

[KRS*24] KHERADMAND, SHAKIBA, REBAIN, DANIEL, SHARMA, GOPAL, et al. "3D Gaussian Splatting as Markov Chain Monte Carlo". *Advances in Neural Information Processing Systems*. Vol. 38. 2024. DOI: 10.48550/arXiv.2404.09591 2, 3.

[KWK*24] KIM, CHUNG MIN, WU, MINGXUAN, KERR, JUSTIN, et al. "GARField: Group Anything with Radiance Fields". *IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2024, 21530–21539. DOI: 10.1109/CVPR52733.2024.02034 2.

[LHTT24] LIU, YICHEN, HU, BENRAN, TANG, CHI-KEUNG, and TAI, YU-WING. "SANeRF-HQ: Segment Anything for NeRF in High Quality". *IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2024, 3216–3226. DOI: 10.1109/CVPR52733.2024.00310 7.

[LYLZ24] LIM, SEUNGHYEON, YOO, YOUNGJAE, LEE, JUN KI, and ZHANG, BYOUNG-TAK. "Multi-Object RANSAC: Efficient Plane Clustering Method in a Clutter". *IEEE International Conference on Robotics and Automation*. 2024. DOI: 10.1109/ICRA57147.2024.10611029 8.

[Mai24] MAINBLADES. *Wing Inspection: 75% faster, 5X more accurate*. 2024. URL: https://www.mainblades.com/case-studies/wing-inspection 3.

[MSO*19] MILDENHALL, BEN, SRINIVASAN, PRATUL P., ORTIZ-CAYON, RODRIGO, et al. "Local Light Field Fusion: Practical View Synthesis with Prescriptive Sampling Guidelines". *ACM Transactions on Graphics* 38.4 (2019), 29:1–14. DOI: 10.1145/3306346.3322980 8.

[Par62] PARZEN, EMANUEL. "On Estimation of a Probability Density Function and Mode". *Annals of Mathematical Statistics* 33.3 (1962), 1065–1076. DOI: 10.1214/aoms/1177704472 2.

[RKH*21] RADFORD, ALEC, KIM, JONG WOOK, HALLACY, CHRIS, et al. "Learning Transferable Visual Models From Natural Language Supervision". *International Conference on Machine Learning*. International Conference on Machine Learning. PMLR 139, 2021, 8748–8763. DOI: 10.48550/arXiv.2103.00020 2, 6.

[SF16] SCHONBERGER, JOHANNES L. and FRAHM, JAN-MICHAEL. "Structure-From-Motion Revisited". *IEEE Conference on Computer Vision and Pattern Recognition*. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2016, 4104–4113. DOI: 10.1109/CVPR.2016.445 1, 2.

[SSE*17] SCHUBERT, ERICH, SANDER, JÖRG, ESTER, MARTIN, et al. "DBSCAN Revisited, Revisited: Why and How You Should (Still) Use DBSCAN". *ACM Transactions on Database Systems* 42.3 (2017), 19:1–21. DOI: 10.1145/3068335 2.

[Stu22] STUURMAN, RENATE. *Sparkly Cup*. 2022 3.

© 2026 The Author(s).
Proceedings published by Eurographics - The European Association for Computer Graphics.

[WXYJ22] WENG, TINGYU, XIAO, JUN, YAN, FEILONG, and JIANG, HAIYONG. "Context-Aware 3D Point Cloud Semantic Segmentation With Plane Guidance". *IEEE Transactions on Multimedia* 25 (2022), 6653–6664. DOI: 10.1109/TMM.2022.3212914 8.

[YSG24] YU, ZEHAO, SATTLER, TORSTEN, and GEIGER, ANDREAS. "Gaussian Opacity Fields: Efficient Adaptive Surface Reconstruction in Unbounded Scenes". *ACM Transactions on Graphics* 43.6 (2024). DOI: 10.1145/3687937 1.

[ZCJ*24] ZHOU, SHIJIE, CHANG, HAORAN, JIANG, SICHENG, et al. "Feature 3DGS: Supercharging 3D Gaussian Splatting to Enable Distilled Feature Fields". *IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2024, 21676–21685. DOI: 10.1109/CVPR52733.2024.02048 2, 3.

[ZPK18] ZHOU, QIAN-YI, PARK, JAESIK, and KOLTUN, VLADLEN. *Open3D: A Modern Library for 3D Data Processing*. 2018. DOI: 10.48550/arXiv.1801.09847 5.

© 2026 The Author(s).
Proceedings published by Eurographics - The European Association for Computer Graphics.