

Symmetry Strikes Back: From Single-Image Symmetry Detection to 3D Generation

Xiang Li¹ Zixuan Huang¹ Anh Thai² James M. Rehg¹

¹University of Illinois at Urbana-Champaign

²Georgia Institute of Technology

<https://ryanxli.github.io/reflect3d>



Figure 1. We propose Reflect3D, a zero-shot 3D reflection symmetry detector capable of accurately detecting 3D symmetry from a single RGB image of an arbitrary object. Conditioned on the detected symmetry, we improve single-image 3D generation in both geometry and texture quality.

Abstract

Symmetry is a ubiquitous and fundamental property in the visual world, serving as a critical cue for perception and structure interpretation. This paper investigates the detection of 3D reflection symmetry from a single RGB image, and reveals its significant benefit on single-image 3D generation. We introduce Reflect3D, a scalable, zero-shot symmetry detector capable of robust generalization to diverse and real-world scenarios. Inspired by the success of foundation models, our method scales up symmetry detection with a transformer-based architecture. We also leverage generative priors from multi-view diffusion models to address the inherent ambiguity in single-view symmetry detection. Extensive evaluations on various data sources

demonstrate that Reflect3D establishes a new state-of-the-art in single-image symmetry detection. Furthermore, we show the practical benefit of incorporating detected symmetry into single-image 3D generation pipelines through a symmetry-aware optimization process. The integration of symmetry significantly enhances the structural accuracy, cohesiveness, and visual fidelity of the reconstructed 3D geometry and textures, advancing the capabilities of 3D content creation.

1. Introduction

”Symmetry is what we see at a glance.” — Blaise Pascal.

Symmetry is ubiquitous in both natural and artificial

objects. It serves as a crucial visual cue that helps humans perceive object structure and interpret spatial relationships [35]. In computer vision, symmetry has long been leveraged as a structural constraint to simplify complex visual tasks, such as pose estimation [23, 47], grasp detection [31], and 3D reconstruction [38, 39, 41, 43]. While existing methods achieve impressive symmetry detection results from 3D or depth data [8, 12, 13, 25, 30, 32, 46], detecting symmetry from a single RGB image remains challenging and underexplored. Past approaches have made progress with restricted object categories [20, 48]. However, robust zero-shot symmetry detection from a single image of arbitrary objects remains an open challenge.

Inspired by the transformative impact of foundation models in NLP [4], vision [16, 26], and speech [29], we revisit single-image symmetry detection through a foundation model perspective. The key to the success of foundation models has been the combination of versatile transformer-based architectures and large, diverse datasets. Based on this observation, we design a transformer-based model for symmetry detection with minimal use of explicit 3D inductive priors. We train this model end-to-end on our large-scale 3D symmetry dataset that combines Objaverse [6] and ShapeNet [5].

One of the major challenges in single-image symmetry detection is single-view ambiguity [48], where depth, perspective distortion, and occlusion obscure symmetry cues. We propose to address this problem by leveraging a generative prior. We apply multi-view diffusion models [21] to synthesize surrounding views of the target object, providing a more complete and less ambiguous perspective on symmetry. By aggregating symmetry predictions across multiple views, we achieve comprehensive and accurate symmetry detection.

Specifically, our approach introduces a novel pipeline for single-image symmetry detection. We first generate multiple surrounding views of the object using a multi-view diffusion model. Next, we apply our feed-forward symmetry detector across each view. Our symmetry detector includes three modules: 1) a frozen DINOv2 encoder that extracts geometry-aware image features, 2) a transformer-based symmetry decoder that uses cross-attention to query these features with various symmetry hypotheses, and 3) MLP heads that classify these hypotheses and refine them through a regressive correction stage. Finally, predictions from all views are clustered and aggregated, yielding precise and comprehensive symmetry predictions.

We evaluate our detected symmetry on two challenging real-world scanned-object datasets, GSO [7] and OmniObject3D [40]. Despite the significant domain shifts from synthetic training data, our model demonstrates remarkable generalization ability. Even without multi-view aggregation, our feed-forward symmetry detector already achieves

state-of-the-art performance. With multi-view aggregation, our symmetry detection performance further improves by a significant margin. We also illustrate the robustness of our approach on in-the-wild internet images.

Finally, we demonstrate the practical utility of symmetry detection in improving single-image 3D generation pipelines. By integrating symmetry predictions into the Score Distillation Sampling (SDS) optimization process, we observe improved 2D and 3D fidelity across both real-world datasets and in-the-wild images, highlighting the broader impact of our symmetry detection approach.

2. Related Work

Symmetry Detection. Early works on symmetry detection introduce several foundational algorithms [2, 17, 37, 44]. Most of these efforts focus on detecting 3D symmetry from explicit 3D representations. For instance, [25] introduces methods for detecting partial and approximate 3D symmetry. More recent works [8, 13, 19] apply neural networks to the problem. Specifically, [12] leverages diffusion models to detect partial symmetry in noisy 3D data, achieving strong performance but still relying on 3D input. Several papers [30–32, 46] extend 3D symmetry detection to RGB-D or depth data. They unproject depth into incomplete 3D point clouds and detect symmetry from the 3D point clouds. Despite these efforts, detecting 3D symmetry directly from 2D RGB image remains challenging and underexplored. [48] introduces NeRD, a neural reflection symmetry detector that constructs 3D cost volumes from RGB image and camera parameters, then detects symmetry based on these cost volumes. While NeRD [48] and its successor NeRD++ [20] achieve reasonable results on in-domain categories and images, they struggle to generalize to in-the-wild scenarios. In contrast, our Reflect3D demonstrates a strong zero-shot generalization ability to diverse real-world imagery.

Single-image to 3D. Single-image 3D reconstruction or generation aims to infer 3D structures from a single 2D image. This is a highly challenging task due to the inherent ambiguity of inferring 3D from only a single viewpoint. Recent work has made significant progress and can be classified into three main directions: 1) feed-forward 3D reconstruction [10, 11, 33, 36], 2) diffusion-based 3D generation [14, 18], and 3) optimization-based approaches [27, 34]. In our work, we take optimization-based approaches as a case study, where we integrate our detected symmetry into an optimization pipeline. Specifically, most of the recent optimization-based 3D generation methods are based on Score Distillation Sampling (SDS). DreamFusion [27] initially proposes SDS. The high-level idea is to leverage a pre-trained 2D diffusion model as an image prior to regularize the optimization of a Neural Radiance Field (NeRF) [24]. Zero-1-to-3 [21] introduces an image-conditioned novel

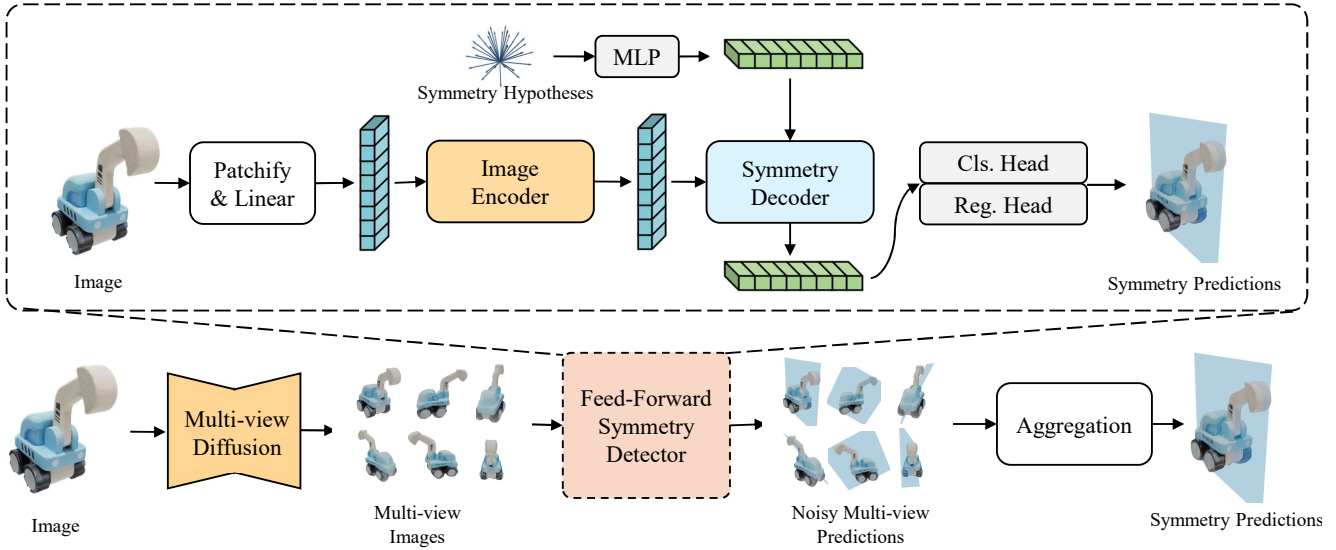


Figure 2. Overview of Reflect3D, our zero-shot single-image symmetry detector. Top: Our transformer-based feed-forward symmetry detector (Section 3.2) predicts symmetry planes from a single RGB image. Bottom: Our multi-view symmetry enhancement pipeline (Section 3.3) leverages multi-view diffusion to resolve the inherent single-view ambiguity in symmetry detection. Aggregating symmetry predictions from multiple synthesized views results in more precise and comprehensive symmetry predictions.

view synthesis diffusion model, which can be used to improve SDS guidance. DreamGaussian [34] incorporates the 3D Gaussian splatting [15] representation to improve the efficiency and quality of SDS-guided 3D generation. Our approach to leveraging symmetry detection for reconstruction builds on DreamGaussian, and incorporates symmetry priors to enhance the geometry and texture fidelity of single-image 3D generation.

Symmetry for 3D Reconstruction. Symmetry has long been recognized as a powerful cue for recovering 3D structures of objects [42]. Recent works [38, 39, 41, 43] study the utility of symmetry in the less constrained single-image 3D reconstruction problem. However, these methods are often constrained to specific object categories, such as faces [38] or rotational artifacts [39], or limited to small datasets covering a few categories [41, 43]. In contrast, our work builds on the insight of leveraging symmetry prior to aid 3D reconstruction in a more generalizable manner. We introduce a zero-shot symmetry detection approach extending beyond category-specific constraints. Unlike [41, 43] that can detect symmetry and reconstruct 3D shapes on about 15 categories, we train our symmetry detector on thousands of categories to enable generalization to in-the-wild imagery. Furthermore, we demonstrate the benefit of integrating symmetry cues in modern single-image 3D generation pipelines [34] guided by diffusion models. Using our accurate symmetry estimation, we achieve superior fidelity in the generated 3D structure and texture.

3. Method

In this section, we describe our single-image 3D symmetry detector, Reflect3D, and our symmetry-aware 3D generation approach. We begin by formalizing the symmetry detection problem in Section 3.1. We then introduce our feed-forward symmetry detector in Section 3.2. We discuss our multi-view symmetry enhancements pipeline in Section 3.3. Finally, in Section 3.4, we describe how we integrate our detected symmetry into SDS for single-image 3D generation.

3.1. Problem Definition

Reflection Symmetry. We follow the definition of reflection symmetry in [48]. Consider the set of points in homogeneous coordinates on the surface of a 3D shape, $S \subset \mathbb{R}^4$. The shape exhibits reflection symmetry with respect to a plane p , if there exists a plane p such that for every point $\mathbf{x} \in S$, its reflected point $\mathbf{M}_p \mathbf{x}$ is also contained within S . Here, \mathbf{M}_p is the reflection transformation matrix across plane p . The surface properties $F(\mathbf{x})$ at point \mathbf{x} are preserved under reflection. Formally:

$$S = M_p(S) \quad \text{and} \quad F(\mathbf{x}) = F(\mathbf{M}_p \mathbf{x}) \quad (1)$$

The reflection matrix $\mathbf{M}_p \in \mathbb{R}^{4 \times 4}$ corresponding to plane p is defined as:

$$\mathbf{M}_p = \begin{bmatrix} \mathbf{I} - 2\mathbf{n}_p \mathbf{n}_p^T & -2d_p \mathbf{n}_p \\ 0 & 1 \end{bmatrix} \quad (2)$$

Where the plane p is parameterized by its normal vector $\mathbf{n}_p \in \mathbb{R}^3$ and its distance d_p from the origin.

In Equation (1), $M_p(S) = \{\mathbf{M}_p \mathbf{x} \mid \mathbf{x} \in S\}$ represents the set of all points in S reflected across the plane p , and $F(\mathbf{x})$ represents the surface properties at point \mathbf{x} , such as texture.

Symmetry Detection. In this work, we address the problem of single-image reflection symmetry detection. Given an object-centric image $I \in \mathbb{R}^{h \times w \times 3}$ with underlying 3D shape $S \subset \mathbb{R}^3$, we aim to find a set of symmetry planes \mathcal{P} , defined as:

$$\mathcal{P} = \{p \mid p = (\mathbf{n}_p, d_p), \mathbf{n}_p \in \mathbb{R}^3, d_p \in \mathbb{R}, \|\mathbf{n}_p\| = 1\} \quad (3)$$

Such that for each $p \in \mathcal{P}$, $S = M_p(S)$ and $F(\mathbf{x}) = F(\mathbf{M}_p \mathbf{x})$. \mathcal{P} could be an empty set for completely asymmetric objects.

Due to single-view scale ambiguity, it is challenging to determine the absolute distance d_p from the symmetry plane to the origin during symmetry detection. We follow [48] to detect $\mathbf{n}_p \in \mathbb{R}^3$ and resolve the relative d_p using other cues, *e.g.*, make it consistent with generated 3D representation in symmetry-aware 3D generation.

3.2. Feed-Forward Symmetry Detector

Inspired by the success of foundation models, we propose our feed-forward single-image symmetry detector Reflect3D, illustrated in the top section of Figure 2.

Image Encoder. Previous works in correspondence learning [3, 45] have identified that DINOv2 [26] is sensitive to 3D spatial features of objects. In contrast to other foundation models, such as CLIP [28], which focuses more on object semantics, DINOv2 aligns better with our goal of detecting geometric cues that facilitate accurate symmetry prediction. We use a frozen DINOv2 encoder to encode our images into a set of features.

Symmetry Decoder. Given the spatial features from our encoder, we design a decoder that transforms these features into symmetry predictions. We discretize the space of possible symmetry plane normals by evenly sampling N fixed unit vectors spanning a hemisphere, which serve as *symmetry hypotheses*. Each hypothesis represents a small neighborhood containing all normal vectors closest to this normal. Hence our hypotheses cover all possible normals. For each hypothesis, we classify whether its neighborhood contains a ground-truth symmetry plane and regress the exact normal vector from the hypothesis. Note that our symmetry hypotheses only span a hemisphere rather than the entire unit sphere because a normal vector \mathbf{n} and its opposite $-\mathbf{n}$ represent the same plane.

Our symmetry hypotheses are N 3-dimensional vectors. We transform the N hypotheses into high-dimensional query features with a shallow MLP. We then apply cross-attention between the query features and image features, followed by a set of self-attention and MLP layers. This yields N feature vectors, each corresponding to a symmetry

hypothesis. For each feature vector, we apply an MLP head to perform binary classification of symmetry, and another MLP head to regress the rotation residual parameterized by a quaternion. Symmetry predictions are obtained by applying the quaternion to the corresponding positive hypothesis. Our detector also provides a confidence score for each prediction derived from the classification probability.

Training. To train our feed-forward symmetry detector, we need to assign each ground truth symmetry plane normal to a corresponding symmetry hypothesis. We sample a sufficient number of hypotheses to ensure that the geodesic distance between each two ground truths is much larger than the geodesic distance between two hypotheses. Then each ground truth is matched to its nearest symmetry hypothesis. We supervise the symmetry binary classification using binary cross-entropy (BCE) loss, and supervise the quaternion regression using mean squared error (MSE) between the predicted normal and the ground truth.

3.3. Multi-view Symmetry Enhancement

A major difficulty in single-image symmetry prediction is the single-view ambiguity—the high uncertainty of the back surface of objects makes the learning ambiguous, especially under a regression framework. We explore using a multi-view diffusion model to resolve this ambiguity and enhance the robustness of our symmetry prediction (see Figure 2 bottom).

To aggregate symmetry predictions from different views, we first apply the multi-view diffusion model conditioned on the input image, and generate M views surrounding the origin. We then filter out inconsistent output images via CLIP-Similarity [28]. We apply our feed-forward symmetry model to the remaining views to obtain a set of symmetry predictions. After rotating all predictions into the coordinate system of the input image view, we perform a K-Means clustering to aggregate the predictions. Finally, we take the cluster centers as our symmetry normal prediction.

3.4. Single-image 3D Generation

We present a novel pipeline for generating 3D textured meshes from a single image, guided by our detected symmetry. Our method is illustrated in Figure 3.

Preliminary. Our work is built on DreamGaussian. DreamGaussian addresses the single-image 3D generation through two stages: SDS optimization and UV texture refinement. In the SDS optimization stage, the scene is initialized as Gaussian splats distributed on a unit sphere. The optimization leverages two loss functions: an MSE loss to align with the input view image and an SDS loss, guided by a diffusion model, applied to randomly sampled novel views. After this stage, a mesh is extracted from the Gaussian splatting scene using the marching cubes algorithm [22]. The texture is then refined in the UV space. The optimization

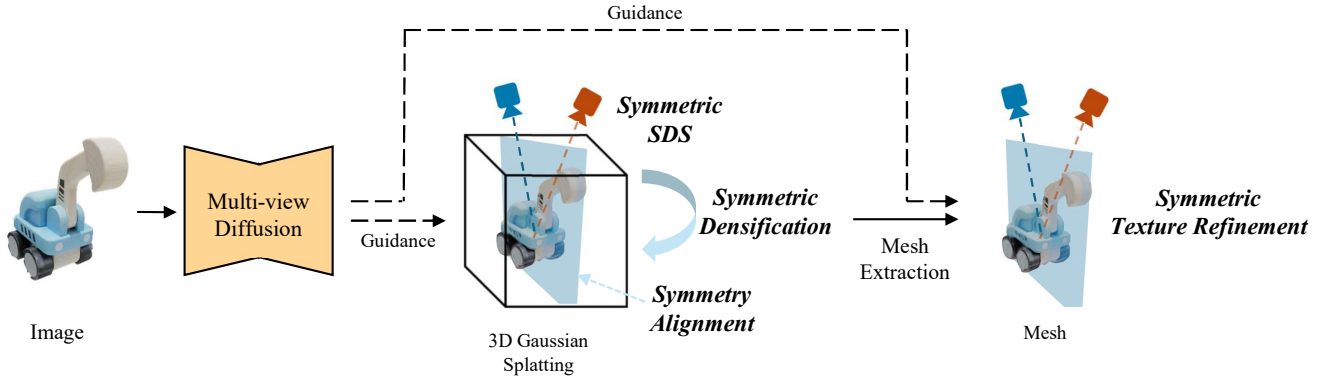


Figure 3. Our symmetry-aware 3D generation pipeline (Section 3.4). Building on DreamGaussian [34], we integrate the detected symmetry through three steps, namely, symmetry alignment, symmetric SDS optimization, and symmetric texture refinement.

	Base Dataset	#Objects	#Categories	#Images	#Symmetry Planes
SN-ShapeNet [30]	ShapeNet	30,000	11	300,000	38,592*
SN-YCB [30]	YCB	18	18	16,189	57*
NeRD-ShapeNet [48]	ShapeNet	35,030	13	175,122	35,030
NeRD-Pix3D [48]	Pix3D	5,285	9	5,285	5,285
Ours	Objaverse + ShapeNet	84,789	1,154	1,095,949	152,019

Table 1. Statistics of our curated dataset. Compared to datasets used in prior works, our curated data enjoys a much higher object diversity and image quantity. * denotes an estimate because the exact training datasets are not released in full.

target of the refinement includes an MSE loss applied to the input view and a texture refinement loss. To compute the texture refinement loss, we first add noise to the rendered images from random viewpoints. A 2D diffusion model is then applied to denoise these images. Finally, we calculate the MSE loss between the rendered and denoised images. We incorporate symmetry into DreamGaussian [34] in three stages: symmetry alignment, symmetric SDS optimization, and symmetric texture refinement.

Symmetry Alignment. As described in Section 3.1, our symmetry detection process identifies the symmetry plane direction $\mathbf{n}_p \in \mathbb{R}^3$. In this stage, we optimize the 3D location and orientation of the symmetry plane to align it with the Gaussian scene. Specifically, we perform the first stage of DreamGaussian optimization without the MSE loss for a limited number of steps. This produces a coarse Gaussian representation with blurred texture and limited detail. From the coarse Gaussian, a point cloud is extracted, and the detected symmetry plane is aligned with the point cloud. Specifically, we reflect the point cloud using the current symmetry estimation, and perform Iterative Closest Points (ICP) between the original and reflected point clouds. We provide detailed description of this alignment process in the appendix.

Symmetric SDS Optimization. In this stage, symmetry information is incorporated in two key steps. First, instead of sampling a single camera view for the SDS loss, we com-

pute SDS losses for both the sampled view and its symmetric counterpart. Second, we symmetrically densify the Gaussian splatting periodically. Every 100 iterations, all Gaussians are reflected across the symmetry planes, and 50% of these reflected Gaussians are randomly sampled and appended to the original set. These strategies encourage symmetry while allowing for minor deviations, accommodating the natural asymmetries often observed in real-world objects.

Symmetric Texture Refinement. In this stage, we extract a mesh from the 3D Gaussians and optimize the UV texture conditioned on the symmetry. Specifically, for mesh regions visible from the input view, the texture is refined directly using an MSE loss with the input image. For regions visible only from the mirrored view, we refine the UV texture using an MSE loss with a flipped image. The same texture refinement loss as DreamGaussian is applied to the remaining regions. Overall, this three-stage approach ensures that the generated 3D shape respects its inherent symmetry, enhancing consistency between the object’s front and back sides.

3.5. Implementation Details

We train our feed-forward symmetry detector using the Adam optimizer with a learning rate of $3e^{-5}$. We use a batch size of 120 and train for 15 epochs. We set the number of symmetry hypotheses N to 31. Since reflection symmetry planes are often far apart in direction, we find this number of hypotheses is sufficient to prevent multiple symmetry planes from being assigned to the same query. We empirically discover that varying the number of hypotheses in a reasonable range has minimal impact on performance. In multi-view enhancement, we set the number of novel views M to 8. We found that 8 views are sufficient and the performance saturates with more views. More implementation details are available in the appendix.

Method	GSO					OmniObject				
	F@5°↑	F@15°↑	F@30°↑	F@50°↑	GD↓	F@5°↑	F@15°↑	F@30°↑	F@50°↑	GD↓
Random Guess	0.007	0.043	0.190	0.464	48.4	0.007	0.062	0.187	0.374	52.1
Shape to Symmetry	0.021	0.128	0.283	0.785	34.9	0.023	0.112	0.236	0.587	34.6
NeRD [48]	0.040	0.398	0.461	0.564	36.2	0.055	0.166	0.328	0.590	41.3
Reflect3D-FF (Ours)	<u>0.191</u>	<u>0.452</u>	<u>0.676</u>	<u>0.867</u>	<u>22.7</u>	<u>0.103</u>	<u>0.276</u>	<u>0.478</u>	<u>0.710</u>	<u>31.1</u>
Reflect3D (Ours)	0.390	0.756	0.848	0.889	13.3	0.173	0.474	0.650	0.799	22.8

Table 2. Quantitative results of our symmetry detection method. Best results are in **bold**, and second-best results are underlined. GD represents average geodesic distance. Our feed-forward model Reflect3D-FF, directly applied to the input image, already achieves state-of-the-art performance. Our multi-view enhanced Reflect3D delivers significant additional improvements.

Method	GSO					OmniObject				
	CLIP-Sim↑	CD↓	F@0.1↑	F@0.2↑	F@0.5↑	CLIP-Sim↑	CD↓	F@0.1↑	F@0.2↑	F@0.5↑
DreamGaussian [34]	0.592	0.442	0.158	0.404	0.767	0.704	0.441	0.199	0.427	0.724
Ours	0.629	0.414	0.172	0.451	0.827	0.734	0.421	0.208	0.457	0.765

Table 3. Quantitative results of our single-image 3D generation method for symmetric objects. CD represents Chamfer distance. Our approach improves both 2D and 3D fidelity compared to the baseline.

Method	F@5°↑	F@15°↑	F@30°↑	F@50°↑	GD↓
Reflect3D (Ours)	0.390	0.756	0.848	0.889	13.3
w/o clustering	0.312	0.655	0.752	0.789	16.0
Reflect3D-FF (Ours)	0.191	0.452	0.676	0.867	22.7
DINOv2 → ViT	0.094	0.386	0.616	0.728	24.7
Frozen → Finetune	0.038	0.236	0.435	0.644	34.2

Table 4. Ablation studies for our symmetry detector Reflect3D on the GSO dataset. Both the clustering-based multi-view aggregation and our frozen, geometric feature-aware encoder are crucial to symmetry detection performance.

4. Data Curation

Training Data. To scale up the training of our Reflect3D, we construct a large-scale symmetry detection dataset by combining Objaverse [6] and ShapeNet [5], resulting in a diverse and extensive 3D reflection symmetry dataset. For the Objaverse dataset, we utilize only the high-quality subset corresponding to LVIS [9] categories. Symmetry ground truth is generated using an optimization-based approach. Specifically, we uniformly sample many candidate planes, reflect the 3D shape by all the candidate planes, and use Chamfer distance to verify whether this candidate likely corresponds to a symmetry plane, followed by an ICP-based refinement for the detected symmetry plane. More details are available in the appendix. The generated symmetry ground truth has been verified through manual inspection of a random subset to ensure alignment with human perception. Table 1 highlights the advantages of our dataset compared to previous symmetry datasets. Our dataset offers significantly greater diversity in object categories and more comprehensive coverage of symmetry planes, as reflected in the higher number of symmetry planes per object.

Evaluation Data. Previous works on symmetry detection, such as [20, 30, 48], evaluate their methods on a validation split of their training datasets. In contrast, we evaluate our

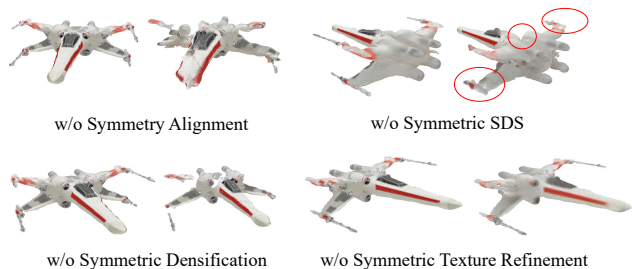


Figure 4. Ablation studies for our single-image 3D generation pipeline. Removing each component adversely affects geometry quality, texture quality, or both.

symmetry detection model, Reflect3D, in a more challenging approach to assess generalization to novel, unseen objects. Specifically, we evaluate on two scanned real-world 3D object datasets: Google Scanned Object (GSO) [7] and OmniObject [40]. The ground truth for these test datasets is generated and validated using the same methodology employed for the training data. We use all symmetric objects from GSO (572 objects) and a randomly sampled subset of OmniObject3D (100 objects). We also test our single-image 3D generation methods on the same test sets. These test sets provide a comprehensive assessment of models’ generalization ability across diverse semantic categories and domains.

5. Symmetry Detection Results

Evaluations Metrics. Given a set of predicted symmetry plane normals $U = \{\mathbf{u}_1, \mathbf{u}_2, \dots\}$ and ground truth normals $V = \{\mathbf{v}_1, \mathbf{v}_2, \dots\}$, we evaluate predictions U against ground truth V with two metrics based on geodesic distance. The geodesic distance between any two normals \mathbf{u} and \mathbf{v} , defined as $\theta(\mathbf{u}, \mathbf{v}) = \cos^{-1}(\mathbf{u} \cdot \mathbf{v})$, measures the angular difference between them.

We calculate F-scores at various geodesic distance thresholds. Given a threshold ϕ in degrees, a predicted and ground truth normal pair is considered a match if their geodesic distance is below ϕ . The $F@{\phi}$ metric is the harmonic mean of $\text{precision}@{\phi}$ and $\text{recall}@{\phi}$. Intuitively, $F@{\phi}$ represents the percentage of correct predictions within a specified threshold.

We also calculate an average geodesic distance (GD). This metric evaluates both the *exactness* and *completeness* of predictions, similar to the precision-recall trade-off in F-scores. For exactness, each predicted normal \mathbf{u} is matched to its closest ground truth normal \mathbf{v}^* , and their average geodesic distance θ_p is calculated. For completeness, each ground truth normal \mathbf{v} is matched to its closest predicted normal \mathbf{u}^* , yielding an average geodesic distance θ_r . The final metric is computed as $\theta_{avg} = (\theta_p + \theta_r)/2$.

Baseline: Shape to Symmetry. A straightforward baseline for single-image 3D symmetry detection is to first generate a 3D shape conditioned on the image, and then detect symmetry planes from the 3D shape. We use our single-image 3D generation baseline [34] to generate 3D shapes from input images. Then we apply our method for generating symmetry ground truth to these 3D shapes. The details for this method are discussed in Section 4 and the appendix.

Baseline: NeRD [48]. NeRD represents the state-of-the-art in single-image 3D reflection symmetry detection. NeRD iteratively samples candidate symmetry plane normals following a coarse-to-fine strategy. Each candidate normal, together with known camera intrinsics, is used to warp 2D image features and construct a 3D cost volume. The cost volume is processed by 3D convolutions to compute a symmetry score. The candidate with the highest score is selected as the final prediction. We use the official implementation of NeRD and report the performance of the best-performing checkpoint on our test sets.

Symmetry Results. The results of our symmetry detection experiments are summarized in Table 2. Even without any multi-view information, our feed-forward model already significantly outperforms all baselines. Specifically, our feed-forward method is both 2 to 4 times more precise than prior state-of-the-art methods reflected in $F@5^\circ$, and also more comprehensive than previous methods as shown in $F@50^\circ$. Incorporating multi-view information further enhances performance, leading to a 1.7 to 2 times improvement in $F@5^\circ$ and reducing the average geodesic distance by 9.4° and 8.3° on the two datasets respectively. These results validate the effectiveness of scaling up our feed-forward transformer model and highlight the substantial gains from leveraging multi-view information to mitigate single-view ambiguity.

Qualitative results in Figure 5 further demonstrate the robustness of our approach. NeRD struggles to generalize to unseen novel objects, often incorrectly predicting a

false front-to-back symmetry plane. In contrast, Reflect3D achieves accurate symmetry predictions for these complex, unseen objects.

Symmetry Ablations. We perform ablation studies on the GSO dataset to validate our design decisions in Reflect3D. In our multi-view enhancement pipeline, clustering the predictions from different views is crucial, since this eliminates redundant predictions that are close together and correspond to the same underlying symmetry plane. In our feed-forward model, the choice of feature representation is crucial. Replacing DINOv2 with ImageNet-trained ViT leads to less precise and less comprehensive predictions. Freezing DINOv2 and only training the decoder is also essential, as fine-tuning drastically degrades the performance. We speculate the pre-trained DINOv2 feature is key to the generalizability of our approach.

6. Single-Image 3D Generation Results

Dataset and Metric. We evaluate single-image 3D generation on the same datasets used for symmetry evaluation, *i.e.*, GSO and OmniObject3D. The evaluation employs both 2D and 3D metrics. For 2D appearance quality, we calculate CLIP similarity following [34]. For 3D geometric quality, we calculate Chamfer distance and F-score following [11]. Specifically, for 3D evaluation, we first leverage ambient occlusion to remove the internal surface. Then we uniformly sample a large number of points from both the predicted and ground truth meshes to calculate Chamfer distance and F-score at various distance thresholds.

Baseline: DreamGaussian [34]. DreamGaussian introduces an efficient and high-quality framework for image-conditioned 3D content generation based on Score Distillation Sampling (SDS). Details of the DreamGaussian pipeline are introduced in Section 3.4. We select the best-performing image-conditioned variant of DreamGaussian, which uses Stable-Zero123 [1] as guidance in the SDS optimization.

3D Generation Results. We summarize our symmetry-conditioned single-image 3D generation results in Table 3. We show that by integrating the symmetry prior, we effectively improve 3D generation in both 2D and 3D metrics. Qualitative comparisons in Figure 6 highlight the benefit of incorporating the symmetry prior. Symmetry helps to avoid missing details (rows 1, 2, and 5) and structural errors. For example, in the baseline results, the left leg of the glasses (row 3) is misinterpreted as part of the circular frame, and the tank barrel (row 4) is incorrectly connected to the base rather than the turret. Our method rectifies these issues. Additionally, the unobserved back side of the baseline-generated objects often lacks geometric detail and texture, whereas our approach leverages symmetry to produce significantly improved geometry and textures for the back side (rows 6, 7, and 8).



Figure 5. Qualitative results for our symmetry detection pipeline. Our Reflect3D achieves better generalization and precision than NeRD [48]. Please refer to our project webpage for video results.

Single-Image 3D Generation Ablations. Ablation studies for our method are shown in Figure 4. Each of the four components in our pipeline is critical to achieving high-quality geometry and texture. Removing symmetry alignment results in inaccurate symmetry, causing distorted appearances. Removing symmetrically sampled SDS leads to lower overall quality and worse symmetry in details, as highlighted in red. Omitting symmetric densification produces sparse outputs that fail to fully respect symmetry. Finally, removing symmetric texture refinement results in blurry textures on the unobserved back side. The synergy of these components is essential for the superior geometry and texture quality achieved by our method.

7. Limitations

A majority of man-made objects and a significant proportion of natural objects are symmetrical. However, there exist objects that are completely asymmetric or highly deformable. Our approach is not designed to handle these objects. However, as mentioned in Section 3.4, our method



Figure 6. Qualitative results for our symmetry-conditioned single-image 3D method. Leveraging detected symmetry, our method avoids missing details and corrects geometric errors (top 4 rows). Meanwhile, we greatly improve the texture and geometric quality of the unobserved object backside (bottom 3 rows). Please refer to our project webpage for video results.

does have tolerance for asymmetric details. Addressing the challenges for these objects remains an avenue for future work.

8. Conclusion

We propose Reflect3D, a scalable zero-shot approach for single-image reflection symmetry detection. Through improvements in architecture, data, and our multi-view aggregation to tackle single-view ambiguity, we significantly improve the generalization performance of single-image symmetry detection. Furthermore, we empirically demonstrate that our detected symmetry can enhance SDS-based single-image 3D generation pipelines, resulting in substantial improvements in the structural and textural quality of the generated objects.

References

- [1] Stability AI. Stable Zero123, 2023. 7
- [2] Mikhail J. Atallah. On symmetry detection. *IEEE Transactions on Computers*, 1985. 2
- [3] Axel Barroso-Laguna, Sowmya Munukutla, Victor Prisacariu, and Eric Brachmann. Matching 2d images in 3d: Metric relative pose from metric correspondences. In *CVPR*, 2024. 4
- [4] Tom Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared D Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, Sandhini Agarwal, Ariel Herbert-Voss, Gretchen Krueger, Tom Henighan, Rewon Child, Aditya Ramesh, Daniel Ziegler, Jeffrey Wu, Clemens Winter, Chris Hesse, Mark Chen, Eric Sigler, Mateusz Litwin, Scott Gray, Benjamin Chess, Jack Clark, Christopher Berner, Sam McCandlish, Alec Radford, Ilya Sutskever, and Dario Amodei. Language models are few-shot learners. In *NeurIPS*, 2020. 2
- [5] Angel X Chang, Thomas Funkhouser, Leonidas Guibas, Pat Hanrahan, Qixing Huang, Zimo Li, Silvio Savarese, Manolis Savva, Shuran Song, Hao Su, et al. Shapenet: An information-rich 3d model repository. *ArXiv*, 2015. 2, 6
- [6] Matt Deitke, Dustin Schwenk, Jordi Salvador, Luca Weihs, Oscar Michel, Eli VanderBilt, Ludwig Schmidt, Kiana Ehsani, Aniruddha Kembhavi, and Ali Farhadi. Objaverse: A universe of annotated 3d objects. In *CVPR*, 2023. 2, 6
- [7] Laura Downs, Anthony Francis, Nate Koenig, Brandon Kinman, Ryan Hickman, Krista Reymann, Thomas B McHugh, and Vincent Vanhoucke. Google scanned objects: A high-quality dataset of 3d scanned household items. In *ICRA*, 2022. 2, 6
- [8] Lin Gao, Ling-Xiao Zhang, Hsien-Yu Meng, Yi-Hui Ren, Yu-Kun Lai, and Leif Kobbelt. Prs-net: Planar reflective symmetry detection net for 3d models. *TVCG*, 2020. 2
- [9] Agrim Gupta, Piotr Dollar, and Ross Girshick. Lvis: A dataset for large vocabulary instance segmentation. In *CVPR*, 2019. 6
- [10] Yicong Hong, Kai Zhang, Jiuxiang Gu, Sai Bi, Yang Zhou, Difan Liu, Feng Liu, Kalyan Sunkavalli, Trung Bui, and Hao Tan. Lrm: Large reconstruction model for single image to 3d. In *ICLR*, 2024. 2
- [11] Zixuan Huang, Stefan Stojanov, Anh Thai, Varun Jampani, and James M Rehg. Zeroshape: Regression-based zero-shot shape reconstruction. In *CVPR*, 2024. 2, 7
- [12] Jiheon Je, Jiayi Liu, Guandao Yang, Boyang Deng, Shengqu Cai, Gordon Wetzstein, Or Litany, and Leonidas Guibas. Robust symmetry detection via riemannian langevin dynamics. In *SIGGRAPH Asia*, 2024. 2
- [13] Penglei Ji and Xinguo Liu. A fast and efficient 3d reflection symmetry detector based on neural networks. *Multimedia Tools and Applications*, 2019. 2
- [14] Heewoo Jun and Alex Nichol. Shap-e: Generating conditional 3d implicit functions. *ArXiv*, 2023. 2
- [15] Bernhard Kerbl, Georgios Kopanas, Thomas Leimkühler, and George Drettakis. 3d gaussian splatting for real-time radiance field rendering. In *SIGGRAPH*, 2023. 3
- [16] 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*, 2023. 2
- [17] Nahum Kiryati and Yossi Gofman. Detecting symmetry in grey level images: The global optimization approach. *IJCV*, 1998. 2
- [18] Yushi Lan, Fangzhou Hong, Shuai Yang, Shangchen Zhou, Xuyi Meng, Bo Dai, Xingang Pan, and Chen Change Loy. Ln3diff: Scalable latent neural fields diffusion for speedy 3d generation. In *ECCV*, 2024. 2
- [19] Ren-Wu Li, Ling-Xiao Zhang, Chun-Peng Li, Yu-Kun Lai, and Lin Gao. E3sym: Leveraging e(3) invariance for unsupervised 3d planar reflective symmetry detection. In *ICCV*, 2023. 2
- [20] Yancong Lin, Silvia-Laura Pintea, and Jan van Gemert. Nerd++: Improved 3d-mirror symmetry learning from a single image. In *BMVC*, 2022. 2, 6
- [21] Ruoshi Liu, Rundi Wu, Basile Van Hoorick, Pavel Tokmakov, Sergey Zakharov, and Carl Vondrick. Zero-1-to-3: Zero-shot one image to 3d object. In *ICCV*, 2023. 2
- [22] William E. Lorensen and Harvey E. Cline. Marching cubes: a high resolution 3d surface construction algorithm. In *SIGGRAPH*, 1998. 4
- [23] Nathaniel Merrill, Yuliang Guo, Xingxing Zuo, Xinyu Huang, Stefan Leutenegger, Xi Peng, Liu Ren, and Guoquan Huang. Symmetry and uncertainty-aware object slam for 6dof object pose estimation. In *CVPR*, 2022. 2
- [24] Ben Mildenhall, Pratul P Srinivasan, Matthew Tancik, Jonathan T Barron, Ravi Ramamoorthi, and Ren Ng. Nerf: Representing scenes as neural radiance fields for view synthesis. In *ECCV*, 2020. 2
- [25] Niloy Jyoti Mitra, Leonidas J. Guibas, and Mark Pauly. Partial and approximate symmetry detection for 3d geometry. In *SIGGRAPH*, 2006. 2
- [26] 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. *TMLR*, 2024. 2, 4
- [27] Ben Poole, Ajay Jain, Jonathan T. Barron, and Ben Mildenhall. Dreamfusion: Text-to-3d using 2d diffusion. In *ICLR*, 2023. 2
- [28] Alec Radford, Jong Wook Kim, Chris Hallacy, Aditya Ramesh, Gabriel Goh, Sandhini Agarwal, Girish Sastry, Amanda Askell, Pamela Mishkin, Jack Clark, et al. Learning transferable visual models from natural language supervision. In *ICML*, 2021. 4
- [29] Alec Radford, Jong Wook Kim, Tao Xu, Greg Brockman, Christine McLeavey, and Ilya Sutskever. Robust speech recognition via large-scale weak supervision. In *ICML*, 2023. 2
- [30] Yifei Shi, Junwen Huang, Hongjia Zhang, Xin Xu, Szymon Rusinkiewicz, and Kai Xu. Symmetrynet: Learning to predict reflectional and rotational symmetries of 3d shapes from single-view rgb-d images. *ACM TOG*, 2020. 2, 5, 6

- [31] Yifei Shi, Zixin Tang, Xiangting Cai, Hongjia Zhang, Dewen Hu, and Xin Xu. Symmetrygrasp: Symmetry-aware antipodal grasp detection from single-view rgb-d images. *IEEE RA-L*, 2022. 2
- [32] Yifei Shi, Xin Xu, Junhua Xi, Xiaochang Hu, Dewen Hu, and Kai Xu. Learning to detect 3d symmetry from single-view rgb-d images with weak supervision. *IEEE TPAMI*, 2022. 2
- [33] Jiayang Tang, Zhaoxi Chen, Xiaokang Chen, Tengfei Wang, Gang Zeng, and Ziwei Liu. Lgm: Large multi-view gaussian model for high-resolution 3d content creation. In *ECCV*, 2024. 2
- [34] Jiayang Tang, Jiawei Ren, Hang Zhou, Ziwei Liu, and Gang Zeng. Dreamgaussian: Generative gaussian splatting for efficient 3d content creation. In *ICLR*, 2024. 2, 3, 5, 6, 7, 11
- [35] Matthias Sebastian Treder. Behind the looking-glass: A review on human symmetry perception. *Symmetry*, 2010. 2
- [36] Zhengyi Wang, Yikai Wang, Yifei Chen, Chendong Xiang, Shuo Chen, Dajiang Yu, Chongxuan Li, Hang Su, and Jun Zhu. Crm: Single image to 3d textured mesh with convolutional reconstruction model. In *ECCV*, 2024. 2
- [37] Jan D Wolter, Tony C Woo, and Richard A Volz. Optimal algorithms for symmetry detection in two and three dimensions. *The Visual Computer*, 1985. 2
- [38] Shangzhe Wu, Christian Ruppert, and Andrea Vedaldi. Unsupervised learning of probably symmetric deformable 3d objects from images in the wild. In *CVPR*, 2020. 2, 3
- [39] Shangzhe Wu, Ameesh Makadia, Jiajun Wu, Noah Snavely, Richard Tucker, and Angjoo Kanazawa. De-rendering the world’s revolutionary artefacts. In *CVPR*, 2021. 2, 3
- [40] Tong Wu, Jiarui Zhang, Xiao Fu, Yuxin Wang, Liang Pan, Jiawei Ren, Wayne Wu, Lei Yang, Jiaqi Wang, Chen Qian, Dahua Lin, and Ziwei Liu. Omniobject3d: Large-vocabulary 3d object dataset for realistic perception, reconstruction and generation. In *CVPR*, 2023. 2, 6
- [41] Yifan Xu, Tianqi Fan, Yi Yuan, and Gurprit Singh. Ladybird: Quasi-monte carlo sampling for deep implicit field based 3d reconstruction with symmetry. In *ECCV*, 2020. 2, 3
- [42] Allen Y. Yang, Kun Huang, Shankar Rao, Wei Hong, and Yi Ma. Symmetry-based 3-d reconstruction from perspective images. In *CVIU*, 2005. 3
- [43] Yuan Yao, Nico Schertler, Enrique Rosales, Helge Rhodin, Leonid Sigal, and Alla Sheffer. Front2back: Single view 3d shape reconstruction via front to back prediction. In *CVPR*, 2020. 2, 3
- [44] Hagit Zabrodsky, Shmuel Peleg, and David Avnir. Symmetry as a continuous feature. *IEEE TPAMI*, 1995. 2
- [45] Junyi Zhang, Charles Herrmann, Junhwa Hur, Luisa Polania Cabrera, Varun Jampani, Deqing Sun, and Ming-Hsuan Yang. A tale of two features: Stable diffusion complements dino for zero-shot semantic correspondence. In *NeurIPS*, 2023. 4
- [46] Zhaoxuan Zhang, Bo Dong, Tong Li, Felix Heide, Pieter Peers, Baocai Yin, and Xin Yang. Single depth-image 3d reflection symmetry and shape prediction. In *ICCV*, 2023. 2
- [47] Heng Zhao, Shenxing Wei, Dahu Shi, Wenming Tan, Zheyang Li, Ye Ren, Xing Wei, Yi Yang, and Shiliang Pu. Learning symmetry-aware geometry correspondences for 6d object pose estimation. In *ICCV*, 2023. 2
- [48] Yichao Zhou, Shichen Liu, and Yi Ma. NeRD: Neural 3d reflection symmetry detector. In *CVPR*, 2021. 2, 3, 4, 5, 6, 7, 8

9. Appendix

In this appendix, we present 1) a detailed description of our ground truth symmetry plane generation method; 2) more implementation details for our Reflect3D symmetry detector and our single-image 3D generation pipeline. All references and citations in this supplementary document refer to the main paper.

9.1. Dataset

In this section, we provide more details on generating ground truth reflection symmetry for our training and evaluation datasets. This generation method is also used in our *Shape to Symmetry* baseline method.

Automatic Ground Truth Generation. We propose a reliable protocol to efficiently generate accurate symmetry ground truth for arbitrary 3D datasets. First, for each mesh in the 3D dataset, we center it by its bounding sphere center and normalize it by scaling its bounding sphere to a unit sphere. This ensures the center of any potential ground truth symmetry plane is on the origin. Then we uniformly sample N_p points from this mesh. To find the symmetry plane, we generate a set of N_c unit vectors uniformly spanning a unit hemisphere as candidates for symmetry plane normals, analogous to the symmetry hypothesis in our Reflect3D single-image symmetry detector. These candidates only span a hemisphere rather than the entire unit sphere because a normal vector \mathbf{n} and its opposite $-\mathbf{n}$ represent the same plane. For each candidate normal, we derive its corresponding candidate plane as the plane passing through the origin and having a normal as the candidate normal. We choose a large enough number of sampled points N_p and symmetry candidates N_c , empirically we use $N_p = 50000$ and $N_c = 31$.

We reflect the point cloud by each candidate plane. Then by how well the reflected point cloud and the original point cloud align, we can infer whether the plane is a ground truth symmetry plane of the shape. We calculate a Chamfer distance between the original and reflected point clouds. We manually select a threshold to eliminate low-quality symmetry planes. We determine this threshold after comprehensively examining the results and making sure it aligns with human perception of symmetry. Meanwhile, for candidate planes that pass the threshold, we can refine them by registering the reflected point cloud to the original point cloud. We apply the iterative closest point (ICP) algorithm for this registration. Then we derive a refined symmetry plane from the registered point clouds.

9.2. Implementation Details

Reflect3D. Our method employs a frozen DINOv2 backbone with the ViT-L/14 architecture. The symmetry decoder consists of 4 layers with 256 channels. Our Adam

optimizer uses a weight decay of 0.05. The classification loss and regression loss are weighted at 1.0 and 0.1, respectively. Training is performed on 2 A100 GPUs for 2 days. For multi-view enhancement, we generate 8 views with the same elevation as the input view, uniformly spacing them in azimuth such that adjacent views are 45° apart. Finally, a clustering threshold of 30° is applied in the aggregation step.

Single-Image 3D Generation. During symmetry alignment, we disable the image-based MSE loss and apply only the symmetrically sampled SDS loss based on the detected symmetry. We run this optimization stage for 400 steps to obtain more accurate symmetry planes, though fewer steps often suffice in practice. We convert the rough Gaussian splatting to a mesh using the marching cubes algorithm. We remove internal surfaces via ambient occlusion, and uniformly sample 10,000 points from the mesh surface to align the symmetry plane, as discussed in Section 3.4. In the subsequent symmetric SDS optimization stage, we optimize for 500 steps following DreamGaussian [34]. Finally, we refine the texture for 50 steps.