CANONAE: SCALABLE POSE SYNCHRONIZATION FOR 3D SHAPE REPRESENTATION LEARNING

Anonymous authors
Paper under double-blind review

ABSTRACT

Many 3D generative pipelines rely on autoencoders (AEs) whose reconstruction quality depends on training-shape orientations. Pose variation complicates learning, while semantic alignment need not optimize geometric representation. We propose CanonAE, which learns canonical orientations from collection geometry and the chosen AE's reconstruction objective without prescribed semantic frames. Geometric pose synchronization initializes consistent poses, followed by joint pose and AE refinement. Our analysis characterizes local pose sensitivity to changes in matching losses and gives conditions for small AE-driven pose adjustments. Motivated by this analysis, our hierarchical, symmetry-aware synchronization retains multiple alignment hypotheses and scales to hundreds of thousands of shapes. Experiments on Thingi10K, CanoVerse, and protein densities show lower test reconstruction error across two AE backbones. Ablations attribute most gains to synchronization, with modest benefits from joint refinement. Thus, scalable geometric canonicalization is an effective initialization for 3D representation learning, including on semantically aligned collections.

1 INTRODUCTION

The availability of 3D shape repositories, including Thingi10K Zhou & Jacobson (2016), ShapeNet Chang et al. (2015), and Objaverse Deitke et al. (2023b;a), has accelerated 3D generative modeling. Much of this progress relies on increasingly expressive neural representations Qi et al. (2017a;b); Hanocka et al. (2019); Park et al. (2019); Chen & Zhang (2019); Mescheder et al. (2019); Sinha et al. (2016); Wu et al. (2016); Chen et al. (2020), while typically treating training-shape poses as fixed. Yet pose variation complicates the distribution a model must learn. Thingi10K and protein datasets lack consistent orientations, while curated datasets such as ShapeNet adopt primarily semantic canonical poses. These conventions need not maximize geometric compatibility: semantic and geometric alignments can differ by approximate symmetries, such as a half-turn exchanging an object's front and back. Rotation-invariant architectures address pose variation; prior results Deng et al. (2021) also suggest that explicit canonicalization can improve representation quality when feasible.

These observations motivate our central question: can optimizing a collection's canonical poses improve representation learning? Our central idea is to learn canonical orientations from collection geometry and the reconstruction objective of the chosen autoencoder (AE), without a user-defined semantic frame. Pose synchronization reconciles pairwise geometric relationships into consistent orientations, which are then refined jointly with the AE parameters. With shape barycenters normalized to the origin, we estimate one rotation $R_i \in \mathrm{SO}(3)$ per shape Kazhdan et al. (2003) and jointly optimize poses and AE parameters:

$$\mathcal{L}_{\mathrm{AE}}(\theta, \{R_i\}) = \sum_{i=1}^{N} D\left(R_i S_i, g^{\theta_g}\left(e^{\theta_e}(R_i S_i)\right)\right) \tag{1}$$

where $\theta = (\theta_e, \theta_g)$ collects the encoder and decoder parameters, and $D : \mathcal{S} \times \mathcal{S} \to \mathbb{R}\geq 0$ measures geometric discrepancy. Joint optimization is challenging because poses and network parameters depend on each other, while approximate rotational symmetries create competing alignments that local optimization may fail to resolve. To obtain reliable initial poses efficiently for large, heterogeneous collections, we first synchronize rotations on a graph connecting geometrically similar shapes:

$$\mathcal{L}_{\mathrm{PS}}(\{R_i\}) = \sum_{(i,j)\in\mathcal{E}} w_{ij} D(R_i S_i, R_j S_j), \qquad R_1 = I_3, \tag{2}$$

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

where $w_{ij} \in (0, 1]$ measures edge affinity. For discrepancies invariant to a common rotation, $R_1 = I_3$ removes the global rotational ambiguity. The objective aligns neighboring shapes through relative rotations induced by a shared set of object poses, using graph consistency to reconcile ambiguous local evidence.

In other words, CanonAE treats canonicalization as selecting a rotational representative for each shape, conditioned on the collection, matching discrepancy, and reconstruction model. The discrepancy defines local geometric compatibility, synchronization reconciles alignment choices across the collection, and joint AE optimization adapts the resulting poses to the reconstruction objective. Canonicalization thus becomes part of learning the shape representation.

Our theoretical analysis examines the stability of geometric canonicalization and its relationship to AE learning. We characterize local sensitivity to matching-loss changes, relating predicted pose displacement to graph connectivity, objective curvature, and changes in edge gradients. We also establish sufficient conditions for global optimality of a geometric stationary point. A separate local quadratic analysis accounts for AE parameter adaptation and identifies conditions under which joint refinement predicts only small pose adjustments. These results provide a conditional justification for geometric synchronization as an initialization for representation learning.



Figure 1: Original poses (top) and optimized poses (bottom) along a path of geometrically similar shapes. Geometric alignment can substantially change individual poses, as illustrated by the middle object.

To scale to hundreds of thousands of objects, we introduce a hierarchical, three-stage synchronization framework. First, we construct overlapping clusters of geometrically similar objects, balancing coverage, connectivity, and computational cost. Second, we estimate coherent initial poses through synchronization within and across clusters. Within each cluster, multiple candidate relative rotations are combined through rotation propagation, aggregation, and Markov random field-based (MRF) selection to resolve competing alignment hypotheses. Shared objects supply candidate relative rotations between clusters, enabling synchronization and reconciliation of their object-pose estimates. Third, we continuously refine object rotations within each cluster, synchronize the refined clusters, and aggregate their estimates into one pose per object. Retaining multiple hypotheses addresses symmetry-induced ambiguity, while distributing refinement across overlapping clusters scales beyond a single collection-wide optimization.

We evaluate on two unaligned datasets, Thingi10K and ProteinDensity, and the semantically aligned CanoVerse Jin et al. (2026). Our approach reduces geometric alignment losses and AE test reconstruction errors across all reported comparisons (Table 1). Relative to the original poses, alignment losses decrease by 25.9%, 29.3%, and 22.6%, respectively. Test reconstruction errors decrease by 2.7%–22.5% on Thingi10K, 3.3%–8.8% on ProteinDensity, and 0.7%–7.4% on CanoVerse, with gains depending on the dataset and backbone. Ablations indicate that synchronization accounts for most observed gains: fixing synchronized poses during AE training increases test error by only 0.3%–0.5% relative to joint refinement. These findings support geometric canonicalization for representation learning, including semantically aligned collections, with modest additional benefits from joint refinement. We also assess stability across matching functions, distinguishing changes in individual poses from changes in alignment energy under approximate symmetries.

2 Related Works

Canonicalization of 3D collections. Canonicalization assigns consistent object-centric orientations to a shape collection; the desired consistency may be semantic, geometric, or determined by a downstream learning objective. Classical approaches often rely on category-specific semantic cues, such as landmarks, part segmentations, or cross-shape correspondences Golovinskiy & Funkhouser (2009); Huang et al. (2011); Kim et al. (2012; 2013). More recent methods learn canonicalization directly from unaligned 3D data: Canonical Capsules Sun et al. (2021) and ConDor Sajnani

et al. (2022) learn self-supervised canonical frames for point clouds, Canonical Fields Agaram et al. (2023) extends this idea to neural fields, while ShapeMatcher Di et al. (2024) and GenAnalysis Yang et al. (2025) jointly reason about shape correspondence and segmentation. Recent work further incorporates semantic references or scales canonicalization to large collections, including the Canonical Objaverse Dataset (COD) Jin et al. (2025), Objaverse-OA Lu et al. (2025), and CanoVerse Jin et al. (2026). We couple collection-level geometric synchronization with reconstruction-driven pose learning: pairwise shape relations initialize globally coherent orientations, and joint AE optimization adapts them to the chosen representation model. A hierarchical construction scales this approach to large, heterogeneous collections with ambiguous alignments.

Joint shape-and-pose learning. A related line of work jointly learns canonical 3D representations and instance-specific pose. CMR Kanazawa et al. (2018), C3DPO Novotný et al. (2019), U-CMR Goel et al. (2020), and DOVE Wu et al. (2023) infer category-specific canonical shape representations from images or videos while treating pose and deformation as latent variables. Related generative models explicitly disentangle pose from content: DC-IGN Kulkarni et al. (2015) introduces pose-specific latent variables, while HoloGAN Nguyen-Phuoc et al. (2019) and PlatonicGAN Henzler et al. (2019) incorporate explicit 3D transformations into the generative process. These methods jointly model intrinsic shape variation and pose, making optimization susceptible to pose ambiguity and poor local minima. Our approach instead uses pose synchronization to obtain globally consistent initial poses before jointly refining the poses and AE. Rather than modeling the full pose-augmented shape space, we learn the AE on a canonical cross-section of this space, without relying on a category-specific template.

Pose synchronization. Pose synchronization recovers globally consistent poses from noisy pairwise transformations. Classical approaches include low-rank and convex formulations Huang & Guibas (2013); Wang & Singer (2013); Bandeira et al. (2014); Chen et al. (2014); Chaudhury et al. (2015); Briales & Gonzalez-Jimenez (2017); Rosen et al. (2019); Carlone & Calafiore (2018); Ling (2023), spectral methods Torsello et al. (2011); Pachauri et al. (2013); Shen et al. (2016); Arrigoni et al. (2016); Ling (2022); He et al. (2025a), and robust non-convex optimization Zhou et al. (2015); Hartley et al. (2011); Chatterjee & Madhav Govindu (2013); Huang et al. (2017); Arrigoni et al. (2015); Sünderhauf & Protzel (2012); Lee et al. (2013). These methods generally assume a single relative pose per object pair. K-best synchronization Sun et al. (2019) instead retains multiple pose candidates, which is important under ambiguous pairwise alignments and self-symmetries. Sun & Huang (2023) handle multiple relative poses per edge through mixture-model diffusion and clustering, MRF selection, and continuous refinement. We replace mixture-based candidate construction with weighted rotation propagation and geometric clustering. Following distributed synchronization Ni et al. (2007); Eriksson et al. (2016); Zhang et al. (2017); Zhu et al. (2018); Li et al. (2024), which coordinates local and global estimates, we group geometrically similar shapes into overlapping clusters and reconcile competing orientations at object and cluster levels. The synchronized poses initialize joint pose and AE learning, linking scalable geometric canonicalization to representation learning.

3 Pose Synchronization and Joint AE Learning

We analyze the local sensitivity of synchronized poses to changes in matching losses (Sec. 3.1) and their predicted displacement under a local quadratic model of joint AE refinement (Sec. 3.2). Under the stated assumptions, these results identify conditions for small pose adjustments and provide a conditional justification for geometric synchronization as an initialization for representation learning. Technical details appear in App. C–E.

3.1 Stability of Pose Synchronization

Different matching losses emphasize different shape features and can favor different pairwise rotations. Synchronization couples these decisions via the cycle-consistency constraint, where compositions of relative rotations along cycles should be the identity rotation.

On a fixed connected weighted graph with Laplacian L, App. C analyzes a change from loss D_0 to D_1 around a stationary D_0 alignment. With the common rotation removed and under the quadratic model assumption, the predicted node rotation increments $\boldsymbol{x}_q$ satisfy

$$\|\boldsymbol{x}_q\| \leq \frac{\eta}{\gamma\sqrt{\lambda_2(L)}}. \tag{3}$$

Here η measures the edge-gradient change after cancellation at shared nodes, $\gamma > 0$ is the minimum D_1 curvature over compatible edge increments, and $\lambda_2(L) > 0$ is the graph spectral gap. Complementary edge constraints can keep γ positive even when individual pairs have unconstrained directions. A larger spectral gap further limits collective node drift. App. D separately provides conditions certifying global optimality of a geometric stationary point in a specified region. Fig. 2 shows stability across our matching loss and ICP.

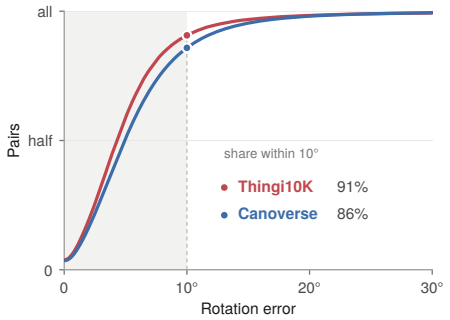


Figure 2: Rotation differences between poses synchronized using our spherical harmonic coefficient alignment and ICP Besl & McKay (1992). Median differences of $4.3°$ (Thingi10K) and $4.9°$ (CanoVerse) indicate stability across shape-matching functions.

3.2 Local Pose Stability under Joint AE Learning

Geometric alignment compares shapes with one another, whereas an AE compares each shape with its reconstruction. Eliminating shared network parameters reveals how the shape poses are coupled. As a simple example, suppose we have a constant-output AE $T_\mu(S) = \mu$. Eliminating its learned output gives

$$\min_{\mu} \frac{1}{2} \sum_i \|R_i S_i - \mu\|^2 = \frac{1}{2n} \sum_{i<j} \|R_i S_i - R_j S_j\|^2. \tag{4}$$

Thus fitting poses and one shared template is exactly complete-graph pose synchronization. More expressive AEs learn input-dependent reconstructions; this example illustrates how shared parameters can still couple shape poses.

For AEs with smooth squared reconstruction loss, App. E starts from a geometric stationary alignment under the same loss D_1. After fitting the network at these poses, its parameters can adapt during pose refinement. Under the stated local regularity and identifiability assumptions, the local quadratic model predicts a shape rotation displacement $\boldsymbol{x}_{\mathrm{AE},q} \in \mathbb{R}^{3n}$ satisfying

$$\|\boldsymbol{x}_{\mathrm{AE},q}\| \le \frac{\|a\|}{c_{\mathrm{AE}}}. \tag{5}$$

Here $\boldsymbol{a}$ is the pose gradient driven by rotation-correlated reconstruction errors, and $c_{\mathrm{AE}} > 0$ is the minimum pose curvature after shared parameter adjustment. Accurate reconstruction can reduce $\|\boldsymbol{a}\|$, while incompatible parameter updates across shapes can preserve c_{AE} (App. E). Small residual forcing relative to this curvature predicts little pose adjustment; at exact reconstruction, the initial poses are already optimal. Fig. 3 shows small pose changes during joint AE learning.

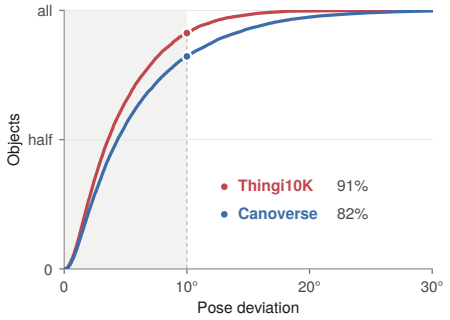


Figure 3: Median pose deviations of joint AE training from our pose sync. output are $3.58°$ and $4.35°$ on Thing10K and CanoVerse, respectively. The keys denote the percent shapes within $10°$.

4 Scalable Rotation Synchronization

We introduce a scalable approach to solving Eq. (2) that addresses two challenges: finding reliable initial rotations for continuous optimization and handling large object collections efficiently. As shown in Fig. 4, our approach has three stages: grouping similar objects into overlapping clusters, synchronizing rotations within and across clusters to obtain globally consistent initial estimates, and refining these estimates through distributed continuous optimization. We first review the necessary mathematical tools, then describe the three stages in Secs. 4.2–4.4. Note that the output provides initialization to Eq. (1) which is solved via alternating minimization of the network parameters and the training shape poses.

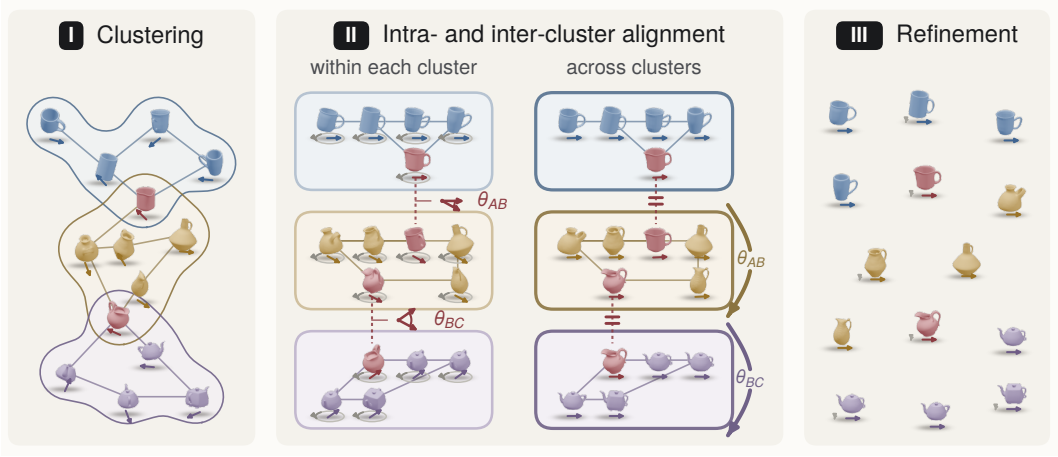


Figure 4: Method pipeline. Overlapping clusters with sparse object graphs form a connected cluster graph through shared objects (I); rotations are synchronized within and then across clusters (II); refinement gives the final poses (III).

4.1 Preliminaries

We summarize the tools used in our approach; details are provided in Appendix F.

Shape descriptors and relative rotations. Following He et al. (2025b), we represent each object by spherical harmonic (SH) coefficients F of its signed distance function or electron density. The coefficient norms at each frequency and spherical sample form a rotation-invariant descriptor $\boldsymbol{d}$. We estimate relative rotations by minimizing $\|F_i(R) - F_j\|_{\mathcal{F}}^2$ over $R \in \mathrm{SO}(3)$, where $F_i(R)$ denotes the coefficients after rotation. We retain multiple local minima to account for self-symmetries.

Rotation synchronization. We use Laplacian synchronization Huang et al. (2019) to estimate absolute rotations from one relative rotation per graph edge. For multiple candidates, MRF synchronization jointly selects poses $R_i \in \mathcal{R}_i$ to minimize $\sum_{(i,j)\in\mathcal{E}} \phi_{ij}(R_i, R_j)$. For object nodes, we use

$$\phi_{ij}(R_i, R_j) = \rho_{\sigma_F}(\|F_i(R_i) - F_j(R_j)\|_{\mathcal{F}}), \qquad \rho_\sigma(t) = \frac{\sigma^2 t^2}{\sigma^2 + t^2}, \tag{6}$$

We set $\sigma_F = 0.6$. For clusters of objects, we use

$$\phi_{ij}(R_i, R_j) = \min_{R_{ij}^t \in \mathcal{R}_{ij}} c_{ij}^t \|R_{ij}^t - R_j^T R_i\|_{\mathcal{F}}^2, \tag{7}$$

where $\mathcal{R}_{ij}$ contains candidate relative rotations and $c_{ij}^t > 0$ is the cost associated with each candidate.

Rotation operations. We denote the Frobenius-norm projection of a matrix A onto $\mathrm{SO}(3)$ by $\mathcal{P}_{SO(3)}(A)$. We also cluster weighted rotation samples and compute a mean rotation for each cluster.

4.2 Intra-Cluster Rotation Synchronization

This stage partitions the input into overlapping clusters and constructs an object graph within each cluster.

4.2.1 Cluster Generation

Given rotation-invariant descriptors $\boldsymbol{d}_i$ for n objects, we construct K clusters $\mathcal{C}_i$, each containing the $m_i \leq m$ nearest objects to its center in descriptor space. These clusters group similar objects for relative rotation estimation and cover all inputs while minimizing total membership $\sum_{i=1}^{K} m_i$.

We use greedy selection and pruning. Farthest point sampling selects $\overline{K}$ candidate centers, each defining a cluster of its m nearest neighbors. We repeatedly select the cluster covering the most uncovered objects until coverage is complete.

We then prune memberships in decreasing neighbor rank while preserving coverage, freezing a cluster when its next removal would leave an object uncovered. We alternate pruning with recentering each cluster at its minimum enclosing sphere's center and reordering its objects by distance. Details are provided in App. G.

4.2.2 Object Graph of Each Cluster

For a cluster $\mathcal{C}$ of m objects, we construct a graph connecting similar objects while requiring its Laplacian to satisfy $\lambda_2(L) > \lambda_{\min} = 0.2$, promoting robustness to relative rotation errors Huang & Guibas (2013). The graphs constructed in our experiments are non-bipartite, consistent with the convergence assumptions in App. H. We initialize a k-nearest-neighbor graph in descriptor space, choosing k large enough to satisfy this constraint. We then greedily remove batches of b edges while preserving it. For a removal batch of edge set $\mathcal{E}_{\text{del}}$ with Laplacian contribution dL, the first-order approximation is

$$\lambda_2(L - dL) \approx \lambda_2(L) - \boldsymbol{u}^T dL \boldsymbol{u} = \lambda_2(L) - \sum_{(i,j)\in\mathcal{E}_{\text{del}}} (u_i - u_j)^2.$$

where $\boldsymbol{u}$ is the unit eigenvector associated with $\lambda_2(L)$. We therefore select the b edges with the smallest $|u_i - u_j|$, stopping before the spectral constraint of the recalculated $\lambda_2(L)$ is violated.

4.3 Global Distributed Rotation Synchronization

This stage initializes object rotations by synchronizing objects within clusters (Sec. 4.3.1), synchronizing overlapping clusters (Sec. 4.3.2), and combining their estimates into object rotations (Sec. 4.3.3).

4.3.1 Intra-Cluster Global Rotation Synchronization

For a cluster $\mathcal{C}$ with object graph $\mathcal{G} = (\mathcal{O}, \mathcal{E})$, we use Sec. 4.1 to compute candidate relative rotations $\mathcal{R}_{ij} = \{(w_{ij}, R_{ij})\}$ for each edge $(i,j) \in \mathcal{E}$, weighted by $w_{ij} = \exp(-\|F_i(R_{ij}) - F_j\|_{\mathcal{F}}^2 / 2\sigma_F^2)$ where $\sigma_F = 0.6$ in all of our experiments. A generalized belief propagation procedure then alternates pose propagation and aggregation to obtain candidate object rotations $\mathcal{R}_i^{in} = \{(w_{i1}, R_{i1}^{in}), \cdots, (w_{ik_i}, R_{ik_i}^{in})\}$.

Let $\mathcal{R}_i^{(t)}, 1 \le i \le m$ denote the candidate sets at iteration t. We initialize $\mathcal{R}_1^{(0)} = \{(1, I_3)\}$ and $\mathcal{R}_i^{(0)} = \emptyset, i \ge 2$, anchoring the first object at the identity with unit weight. Since R_{ij} maps object i to object j, votes from j to i use R_{ij}:

$$\mathcal{R}_i^{(t+1)} = \cup_{(i,j)\in\mathcal{E},(w_{ij},R_{ij})\in\mathcal{R}_{ij},(w_k,R_k)\in\mathcal{R}_j^{(t)}} (w_{ij}w_k, R_k R_{ij}). \tag{8}$$

We reduce $\mathcal{R}_i^{(t+1)}$ by rotation clustering (Sec. 4.1) and obtain $\mathcal{R}_i^{in} = \mathcal{R}_i^{(t_{\max})}$ after the final iteration. In our experiments, propagation on both graph levels starts from a single seed, with $t_{\max}$ extending beyond the rounds needed to populate all candidate sets. App. H establishes convergence for the single-seed, single-candidate, equal-weight specialization, providing insight into the general procedure.

We next use the MRF formulation in Sec. 4.1, with the cost in Eq. 6, to select $R_i^{in} \in \mathcal{R}_i^{in}$. Starting from R_i^{in}, Gauss-Newton refinement solves Eq. (2) with

$$D(R_i S_i, R_j S_j) = \rho_{\sigma_F}(\|F_i(R_i) - F_j(R_j)\|_{\mathcal{F}}). \tag{9}$$

The resulting rotations $R_i^{k,l}$ for each cluster c_k are the intra-cluster synchronization outputs.

4.3.2 Inter-Cluster Global Rotation Synchronization

We estimate an absolute rotation $R_{c_k}^{in}$ for each cluster c_k by extending the procedure in Sec. 4.3.1. For each pair of overlapping clusters $(k, k') \in \mathcal{E}_C$, every shared object $i \in c_k \cap c_{k'}$ induces a relative rotation $R_i^{k',l} R_i^{k,l^T}$. Rotation clustering (Sec. 4.1) consolidates these estimates into a candidate set $\mathcal{R}_{c_k c_k'}$.

The cluster-overlap graphs used in our experiments are also non-bipartite. We apply the same propagation and aggregation procedure to obtain candidate absolute rotations $\mathcal{R}_{c_k}$ for each cluster. An MRF (Sec. 4.1) with the potential in Eq. (7) then selects $R_{c_k}^{in}$.

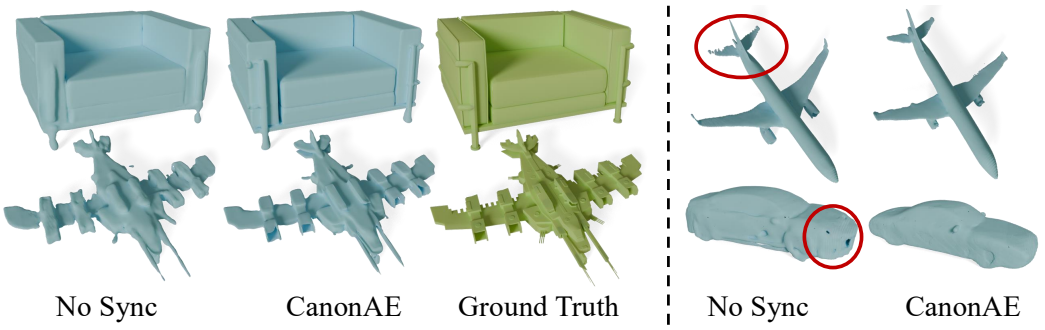


Figure 5: (Left) AE trained using CanonAE poses produces more faithful reconstructions than using original poses. (Right) Shapes generated using the AE trained on our poses contain fewer geometric artifacts than ones generated from AE trained on original poses. More examples are in App. A.

4.3.3 From Cluster Poses to Object Poses

For each object i, we cluster the induced rotations $\{R_{c_k}^{in} R_i^{k,l} | i \in c_k\}$ to obtain $\mathcal{R}_i^{in}$. An MRF with the cost in Eq. (6) then selects the initial object rotation $R_i^{in} \in \mathcal{R}_i^{in}$.

4.4 Local Distributed Rotation Synchronization

Starting from the rotations R_i^{in} obtained in stage II, joint first-order optimization of Eq. (2) converges slowly for hundreds of thousands of objects. We therefore apply Gauss-Newton refinement within each cluster c_k, initialized by R_i^{in}, to obtain $R_i^k, \forall i \in c_k$. The shared initialization encourages consistency across overlapping clusters.

For each overlapping pair $(k, k') \in \mathcal{E}_\mathcal{C}$, we estimate the relative cluster rotation as $R_{c_k c'_k} = \mathcal{P}_{SO(3)}\big(\sum_{i \in c_k \cap c_{k'}} R_i^{k'} R_i^{k^T}\big)$. Laplacian rotation synchronization (Sec. 4.1) then computes cluster rotations $R_{c_k}^{\text{Lap}}$ from $R_{c_k c_{k'}}$. Using these as initialization, we refine the cluster rotations by solving

$$\{R_{c_k}^\star\} = \underset{\{R_{c_k}\}}{\text{argmin}} \sum_{(k,k') \in \mathcal{E}_C} \|R_{c_{k'}} R_{c_k c_{k'}} - R_{c_k}\|_\mathcal{F}^2, \;\; s.t. \;\; R_{c_1} = I_3.$$

Finally, we aggregate the aligned estimates to obtain each object's rotation: $R_i^\star = \mathcal{P}_{SO(3)}\big(\sum_{k: i \in c_k} R_{c_k}^\star R_i^k\big)$.

5 Experimental Evaluation

We begin with the experimental setup in Sec. 5.1. We then evaluate reconstruction/generation results using the optimized poses for 3D shapes and protein densities in Sec. 5.2 and Sec. 5.3, respectively. Finally, we present an ablation study to evaluate various design choices of our approach in Sec. 5.4.

5.1 Experimental Setup

Datasets. We evaluate on three datasets: Thingi10K Zhou & Jacobson (2016), containing unaligned objects, many without obvious canonical semantic orientations; ProteinD, containing 44K inconsistently oriented density maps at 6 Å resolution curated from the Protein Data Bank Burley et al. (2025); and CanoVerse Jin et al. (2026), containing 320K shapes from Objaverse Deitke et al. (2023b;a) semantically aligned by human annotators. We use the official train/test splits of CanoVerse. For Thingi10K and ProteinD, we generated 90%/10% Train/Test splits randomly.

Implementation details. We retain the best relative rotation per edge on Thingi10K and CanoVerse, all relative rotations on ProteinD, and four pose candidates during propagation. We use two AE backbones: a point cloud encoder Achlioptas et al. (2018) with a DeepSDF decoder Park et al. (2019) (AEI), and 3DShape2VecSet Zhang et al. (2023) (AEII), used in commercial 3D generators Li et al. (2025); Tencent Hunyuan3D Team (2025). Synchronization and joint AE learning use only training data. After pose initialization (Sec. 4), we alternate network and training-pose updates to solve Eq. (1). At test time, each shape is aligned to its 16 nearest training shapes in descriptor space;

	Thingi10K			ProteinD			CanoVerse		
Loss	$r(\mathcal{E}_c)$	AEI-T	AEII-T	$r(\mathcal{E}_c)$	AEI-T	AEII-T	$r(\mathcal{E}_c)$	AEI-T	AEII-T
Our approach	**18.67**	**11.04**	**7.66**	**14.34**	**38.15**	**22.23**	**20.37**	**7.57**	**2.89**
No synchronization	25.20	11.35	9.89	20.27	41.82	22.99	26.33	7.62	3.12
Stage-III only	24.45	11.75	9.81	19.96	44.30	23.50	26.01	8.01	–
Zhu et al.	24.36	12.22	10.38	18.40	44.04	23.29	24.26	8.05	–

Table 1: Comparison on 3D shape and protein density datasets. We report mean alignment error per edge and test MAE ($\times 1000$) for PointNet+DeepSDF (AEI-T) and 3DShape2VecSet (AEII-T). AE results use one training run per configuration. '—': not evaluated.

rotation clustering initializes its pose, which is refined using the AE reconstruction loss. All methods use identical test-time procedures and optimization budgets, with AE parameters fixed.

Baselines. We compare original poses, their stage-III refinement, and distributed synchronization Zhu et al. (2018), all using the same joint AE training procedure. Synchronization uses Intel Xeon Platinum 8380 CPUs; AE training uses NVIDIA GH200 and RTX PRO 6000 Blackwell GPUs. Timings and concurrency are in App. B.

5.2 Evaluations on 3D Shape Datasets

We evaluate our optimized poses on Thingi10K and CanoVerse. Tab. 1 reports alignment errors and test reconstruction MAE; Fig. 5 and App. A.1 provide qualitative comparisons.

On Thingi10K and CanoVerse, our approach reduces alignment errors by 23.4% and 16.0% relative to the strongest baseline, Zhu et al. (2018), and by 25.9% and 22.6% relative to the original poses, respectively. The gains on semantically aligned CanoVerse show that semantic canonicalization need not minimize geometric alignment error.

Our optimized poses also reduce test reconstruction MAE on both datasets, with gains varying across architectures. AEI shows modest improvements, possibly because its limited capacity restricts its ability to exploit improved alignment. The more expressive AEII achieves 22.5% and 7.4% reductions on Thingi10K and CanoVerse relative to the original poses.

For man-made shapes, geometric and semantic alignments can differ by symmetry-related orientations, including front–back reversals. These ambiguities may hinder local optimization, potentially explaining why Stage-III underperforms our full approach in alignment residuals and test reconstruction. Notably, Stage-III only and Zhu et al. reduce alignment residuals relative to the original poses but increase test MAE in 9 of 10 comparisons. Thus, lower alignment residuals alone do not guarantee better reconstruction. Synchronizing CanoVerse's 320K shapes takes 18.2 h (App. B).

Qualitative results support the reconstruction metrics: AEs trained with our poses better preserve global structure and geometric details. For generation, we train latent rectified flow models Liu et al. (2023) on ShapeNet. For CanonAE, we encode shapes in their CanonAE poses using our AEII trained on CanoVerse; for the baseline, we encode them in their original poses using off-the-shelf AEII checkpoint. Qualitatively better shapes are sampled with our pipeline (Fig. 5, right; App. A.2), consistent with the improved reconstruction quality of the learned decoders.

5.3 Evaluations on Protein Densities

As shown in Tab. 1, alignment results on the protein density dataset follow similar trends to Thingi10K and CanoVerse, with comparable baseline rankings. Our approach reduces alignment error by 29.3% relative to the original poses, similar to the reduction on Thingi10K, which also lacks consistent orientations.

Test reconstruction gains show a different pattern. AEI reduces MAE by 8.8%, exceeding its gains on Thingi10K and CanoVerse, whereas AEII achieves a smaller reduction of 3.3%. This may reflect the comparatively limited variation in protein densities: improved alignment helps AEI capture this variation, while the more expressive AEII may already accommodate much of it without pose optimization. These results suggest that the benefits depend on model capacity relative to dataset complexity, rather than expressiveness alone.

5.4 Ablation Study

5.4.1 Pose synchronization

The ablation study concerns different components of the pose synchronization approach.

	Th.-10K	ProteinD	CanoVerse
Full	**18.67**	**14.34**	**20.37**
One Cand.	19.13	15.08	20.75
All Rel-Rot.	18.79	14.34	20.47
One Rel-Rot.	18.67	14.58	20.37
Two Rel-Rot.	18.74	14.44	20.43
No-Sta.-III	18.85	14.56	20.55
GMM	18.93	14.90	NA

Table 2: Pre-AE SH alignment errors.

Single candidate during pose propagation. Using one candidate rotation per object increases mean alignment error across all three datasets (Tab. 2). Multiple candidates allow the MRF to jointly select compatible poses, avoiding premature commitment to locally favorable but inconsistent rotations.

Varying the number of relative rotations per edge. One relative rotation per edge yields the lowest alignment error on Thingi10K and CanoVerse, with small differences across settings; additional candidates may introduce ambiguity. On ProteinDensity, retaining all candidates performs best, supporting alternative alignments for objects with pronounced self-symmetries.

No stage III. Removing stage III and directly using the stage-II MRF solution produces a modest but consistent increase in alignment error across all three datasets (Tab. 2). This indicates that continuous geometric refinement further improves the discrete pose estimates obtained through synchronization.

GMM candidate-generation ablation. Within the same hierarchy, we replace within-cluster pose propagation and rotation clustering with an isotropic GMM-based procedure adapted from Sun & Huang (2023), evaluating candidate generation. In our implementation, this variant is 10–100× slower and does not finish CanoVerse within one month. Alignment errors increase on Thingi10K and ProteinD (Tab. 2), while AE-II test MAE rises from 7.66 to 7.75 and 22.23 to 22.36, respectively. Our procedure improves computational efficiency, with small alignment and reconstruction gains.

Distance-based alignment. The final ablation replaces spherical harmonics pairwise alignment with ICP on the two shape datasets. The resulting object poses remain stable across these alignment methods (Fig. 2).

5.4.2 AE Learning

No alternating optimization. Holding the synchronized poses fixed during AE training increases test MAE by 0.3%–0.5% relative to joint refinement in the reported runs (Tab. 3). Similar small differences occur with the baseline initializations (App. B). These observations suggest a modest benefit from joint refinement. These results come from single runs.

	Th.-10K	ProteinD	CanoVerse
WO-Alter-AEI	+0.54%	+0.41%	+0.37%
WO-Alter-AEII	+0.32%	+0.36%	+0.33%
Random-AEI	-2.44%	-1.21%	+0.31%
Random-AEII	-3.12%	-1.83%	+0.69%

Table 3: Relative test MAE vs. joint refinement (WO-Alter) and unaugmented No Sync (Random).

Random rotation augmentation. Random rotation augmentation can mitigate overfitting Achlioptas et al. (2018). Table 3 reports relative changes in test reconstruction error compared with the corresponding no synchronization without augmentation. We observe substantial gains only for small, unaligned datasets such as Thingi10K; augmentation remains slightly less effective than training on semantically aligned shapes.

6 Conclusions and Future Work

CanonAE combines collection-level geometric pose synchronization with reconstruction-driven pose refinement for 3D representation learning. Its hierarchy reconciles competing alignment hypotheses and scales to hundreds of thousands of shapes. Across three datasets and two AE backbones, it reduces test reconstruction errors relative to training with original poses, including on semantically aligned CanoVerse. Ablations attribute most gains to synchronization, with modest benefits from joint pose and AE refinement.

Local stability analysis identifies conditions predicting small pose adjustments under matching-loss changes and AE refinement, supporting geometric initialization. These findings support scalable geometric canonicalization for representation learning. Future work will examine dependence on matching losses, model capacity, and dataset structure.

AI USE STATEMENT

Generative AI tools were used to polish the language of the manuscript, which was originally written by the authors, and to assist with coding for experimental implementation. All AI-assisted text, code, experiments, and results were reviewed and verified by the authors, who take full responsibility for the final manuscript.

REPRODUCIBILITY STATEMENT

Upon acceptance, we will release the code and scripts necessary to reproduce the main experiments and results presented in this paper.

ETHICS STATEMENT

This paper presents a geometric framework for 3D shape canonicalization to improve representation learning and generation. All experiments were conducted using publicly accessible datasets (Thingi10K, ShapeNet, and public protein density databases) that present no privacy or safety risks. We do not foresee any immediate or unique negative societal impacts or ethical concerns arising from our work, besides the potential for misuse which already exists for 3D generative pipelines.

REFERENCES

Panos Achlioptas, Olga Diamanti, Ioannis Mitliagkas, and Leonidas Guibas. Learning representations and generative models for 3d point clouds. In *Proceedings of the International Conference on Machine Learning (ICML)*, pp. 40–49, 2018.

Rohith Agaram, Shaurya Dewan, Rahul Sajnani, Adrien Poulenard, Madhava Krishna, and Srinath Sridhar. Canonical fields: Self-supervised learning of pose-canonicalized neural fields. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 4500–4510, June 2023.

B. Andres, T. Beier, and J.H. Kappes. OpenGM: A C++ library for discrete graphical models. *CoRR*, abs/1206.0111, 2012. URL http://arxiv.org/abs/1206.0111.

Federica Arrigoni, Andrea Fusiello, Beatrice Rossi, and Pasqualina Fragneto. Robust rotation synchronization via low-rank and sparse matrix decomposition. *CoRR*, abs/1505.06079, 2015.

Federica Arrigoni, Beatrice Rossi, and Andrea Fusiello. Spectral synchronization of multiple views in se (3). *SIAM Journal on Imaging Sciences*, 9(4):1963–1990, 2016.

Afonso S Bandeira, Moses Charikar, Amit Singer, and Andy Zhu. Multireference alignment using semidefinite programming. In *Proceedings of the 5th conference on Innovations in theoretical computer science*, pp. 459–470. ACM, 2014.

Paul J. Besl and Neil D. McKay. A method for registration of 3-d shapes. *IEEE Trans. Pattern Anal. Mach. Intell.*, 14(2):239–256, 1992. doi: 10.1109/34.121791. URL https://doi.org/10.1109/34.121791.

Jesus Briales and Javier Gonzalez-Jimenez. Initialization of 3d pose graph optimization using lagrangian duality. In *2017 IEEE International Conference on Robotics and Automation (ICRA)*, pp. 5134–5139. IEEE, 2017.

Stephen K. Burley, Rusham Bhatt, Charmi Bhikadiya, Chunxiao Bi, Alison Biester, Pratyoy Biswas, Sebastian Bittrich, Santiago Blaumann, Ronald Brown, Henry Chao, Vivek Reddy Chithari, Paul A. Craig, Gregg V. Crichlow, Jose M. Duarte, Shuchismita Dutta, Zukang Feng, Justin W. Flatt, Sutapa Ghosh, David S. Goodsell, Rachel Kramer Green, Vladimir Guranovic, Jeremy Henry, Brian P. Hudson, Michael Joy, Jason T. Kaelber, Igor Khokhriakov, Jhih-Siang Lai, Catherine L. Lawson, Yuhe Liang, Douglas Myers-Turnbull, Ezra Peisach, Irina Persikova, Dennis W. Piehl, Aditya Pingale, Yana Rose, Jared Sagendorf, Andrej Sali, Joan Segura, Monica Sekharan, Chenghua Shao, James Smith, Michael Trumbull, Brinda Vallat, Maria Voigt, Ben Webb, Shamara Whetstone, Amy Wu-Wu, Tongji Xing, Jasmine Y. Young, Arthur Zalevsky, and Christine Zardecki. Updated resources for exploring experimentally-determined PDB structures and computed structure models at the RCSB Protein Data Bank. *Nucleic Acids Research*, 53(D1):

D564–D574, 2025. doi: 10.1093/nar/gkae1091. URL https://doi.org/10.1093/nar/gkae1091.

Luca Carlone and Giuseppe C Calafiore. Convex relaxations for pose graph optimization with outliers. *IEEE Robotics and Automation Letters*, 3(2):1160–1167, 2018.

Angel X. Chang, Thomas A. Funkhouser, Leonidas J. Guibas, Pat Hanrahan, Qi-Xing Huang, Zimo Li, Silvio Savarese, Manolis Savva, Shuran Song, Hao Su, Jianxiong Xiao, Li Yi, and Fisher Yu. Shapenet: An information-rich 3d model repository, 2015. URL http://arxiv.org/abs/1512.03012.

Avishek Chatterjee and Venu Madhav Govindu. Efficient and robust large-scale rotation averaging. In *Proceedings of the IEEE International Conference on Computer Vision*, pp. 521–528, 2013.

Kunal N. Chaudhury, Yuehaw Khoo, and Amit Singer. Global registration of multiple point clouds using semidefinite programming. *SIAM Journal on Optimization*, 25(1):468–501, 2015.

Yuxin Chen, Leonidas Guibas, and Qixing Huang. Near-optimal joint object matching via convex relaxation. In *Proceedings of the 31st International Conference on Machine Learning*, volume 32 of *Proceedings of Machine Learning Research*, pp. 100–108. PMLR, 2014. URL https://proceedings.mlr.press/v32/chend14.html.

Zhiqin Chen and Hao Zhang. Learning implicit fields for generative shape modeling. In *IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2019, Long Beach, CA, USA, June 16-20, 2019*, pp. 5939–5948, New York, NY, USA, 2019. Computer Vision Foundation / IEEE. URL http://openaccess.thecvf.com/content_CVPR_2019/html/Chen_Learning_Implicit_Fields_for_Generative_Shape_Modeling_CVPR_2019_paper.html.

Zhiqin Chen, Andrea Tagliasacchi, and Hao Zhang. Bsp-net: Generating compact meshes via binary space partitioning. In *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition, CVPR 2020, Seattle, WA, USA, June 13-19, 2020*, pp. 42–51, Washington, DC, USA, 2020. Computer Vision Foundation / IEEE. doi: 10.1109/CVPR42600.2020.00012. URL https://openaccess.thecvf.com/content_CVPR_2020/html/Chen_BSP-Net_Generating_Compact_Meshes_via_Binary_Space_Partitioning_CVPR_2020_paper.html.

David J Crandall, Andrew Owens, Noah Snavely, and Daniel P Huttenlocher. Sfm with mrfs: Discrete-continuous optimization for large-scale structure from motion. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 35(12):2841–2853, 2013.

Matt Deitke, Ruoshi Liu, Matthew Wallingford, Huong Ngo, Oscar Michel, Aditya Kusupati, Alan Fan, Christian Laforte, Vikram Voleti, Samir Yitzhak Gadre, Eli VanderBilt, Aniruddha Kembhavi, Carl Vondrick, Georgia Gkioxari, Kiana Ehsani, Ludwig Schmidt, and Ali Farhadi. Objaverse-xl: A universe of 10m+ 3d objects. In Alice Oh, Tristan Naumann, Amir Globerson, Kate Saenko, Moritz Hardt, and Sergey Levine (eds.), *Advances in Neural Information Processing Systems 36: Annual Conference on Neural Information Processing Systems 2023, NeurIPS 2023, New Orleans, LA, USA, December 10 - 16, 2023*, Red Hook, NY, 2023a. Curran Associates. URL http://papers.nips.cc/paper_files/paper/2023/hash/70364304877b5e767de4e9a2a511be0c-Abstract-Datasets_and_Benchmarks.html.

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 *IEEE/CVF Conference on Computer Vision and Pattern Recognition, CVPR 2023, Vancouver, BC, Canada, June 17-24, 2023*, pp. 13142–13153, Washington, DC, USA, 2023b. IEEE. doi: 10.1109/CVPR52729.2023.01263. URL https://doi.org/10.1109/CVPR52729.2023.01263.

Congyue Deng, Or Litany, Yueqi Duan, Adrien Poulenard, Andrea Tagliasacchi, and Leonidas J Guibas. Vector neurons: A general framework for so(3)-equivariant networks. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 12200–12209, 2021.

Yan Di, Chenyangguang Zhang, Chaowei Wang, Ruida Zhang, Guangyao Zhai, Yanyan Li, Bowen Fu, Xiangyang Ji, and Shan Gao. Shapematcher: Self-supervised joint shape canonicalization segmentation retrieval and deformation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 21017–21028, June 2024.

Anders Eriksson, John Bastian, Tat-Jun Chin, and Mats Isaksson. A consensus-based framework for distributed bundle adjustment. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 1754–1762, June 2016.

Shubham Goel, Angjoo Kanazawa, and Jitendra Malik. Shape and viewpoint without keypoints. In Andrea Vedaldi, Horst Bischof, Thomas Brox, and Jan-Michael Frahm (eds.), *Computer Vision - ECCV 2020 - 16th European Conference, Glasgow, UK, August 23-28, 2020, Proceedings, Part XV*, volume 12360 of *Lecture Notes in Computer Science*, pp. 88–104. Springer, 2020. doi: 10.1007/978-3-030-58555-6_6. URL https://doi.org/10.1007/978-3-030-58555-6_6.

Aleksey Golovinskiy and Thomas Funkhouser. Consistent segmentation of 3D models. *Computers & Graphics*, 33(3):262–269, 2009. doi: 10.1016/j.cag.2009.03.010.

Rana Hanocka, Amir Hertz, Noa Fish, Raja Giryes, Shachar Fleishman, and Daniel Cohen-Or. Meshcnn: a network with an edge. *ACM Trans. Graph.*, 38(4):90:1–90:12, 2019. doi: 10.1145/3306346.3322959. URL https://doi.org/10.1145/3306346.3322959.

Richard I. Hartley, Khurrum Aftab, and Jochen Trumpf. L1 rotation averaging using the weiszfeld algorithm. In *CVPR*, pp. 3041–3048. IEEE Computer Society, 2011.

Zihang He, Hang Ruan, and Qixing Huang. A robust translation synchronization algorithm. In *International Conference on 3D Vision, 3DV 2025, Singapore, March 25-28, 2025*, pp. 276–285. IEEE, 2025a. doi: 10.1109/3DV66043.2025.00030. URL https://doi.org/10.1109/3DV66043.2025.00030.

Zihang He, Yuezhi Yang, Congyue Deng, Jiaxin Lu, Leonidas Guibas, and Qixing Huang. An efficient global-to-local rotation optimization approach via spherical harmonics. *Computer Graphics Forum*, 44(5):e70185, 2025b. doi: 10.1111/cgf.70185.

Philipp Henzler, Niloy J. Mitra, and Tobias Ritschel. Escaping plato's cave: 3d shape from adversarial rendering. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 9984–9993, 2019.

Qixing Huang and Leonidas Guibas. Consistent shape maps via semidefinite programming. *Computer Graphics Forum*, 32(5):177–186, 2013. doi: 10.1111/cgf.12184.

Qixing Huang, Vladlen Koltun, and Leonidas Guibas. Joint shape segmentation with linear programming. *ACM Transactions on Graphics*, 30(6):125:1–125:11, 2011.

Xiangru Huang, Zhenxiao Liang, Chandrajit Bajaj, and Qixing Huang. Translation synchronization via truncated least squares. In *Advances in neural information processing systems*, pp. 1459–1468, 2017.

Xiangru Huang, Zhenxiao Liang, Xiaowei Zhou, Yao Xie, Leonidas J. Guibas, and Qixing Huang. Learning transformation synchronization. In *IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2019, Long Beach, CA, USA, June 16-20, 2019*, pp. 8082–8091. Computer Vision Foundation / IEEE, 2019. doi: 10.1109/CVPR.2019.00827. URL http://openaccess.thecvf.com/content_CVPR_2019/html/Huang_Learning_Transformation_Synchronization_CVPR_2019_paper.html.

Li Jin, Yujie Wang, Wenzheng Chen, Qiyu Dai, Qingzhe Gao, Xueying Qin, and Baoquan Chen. One-shot 3d object canonicalization based on geometric and semantic consistency. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 16850–16859, June 2025.

Li Jin, Yuchen Yang, Weikai Chen, Yujie Wang, Dehao Hao, Tanghui Jia, Yingda Yin, Zeyu Hu, Runze Zhang, Keyang Luo, Li Yuan, Long Quan, Xin Wang, and Xueying Qin. CanoVerse: 3d object scalable canonicalization and dataset for generation and pose. In *European Conference on Computer Vision (ECCV)*, 2026.

Angjoo Kanazawa, Shubham Tulsiani, Alexei A. Efros, and Jitendra Malik. Learning category-specific mesh reconstruction from image collections. In Vittorio Ferrari, Martial Hebert, Cristian Sminchisescu, and Yair Weiss (eds.), *Computer Vision - ECCV 2018 - 15th European Conference, Munich, Germany, September 8-14, 2018, Proceedings, Part XV*, volume 11219 of *Lecture Notes in Computer Science*, pp. 386–402. Springer, 2018. doi: 10.1007/978-3-030-01267-0_23. URL https://doi.org/10.1007/978-3-030-01267-0_23.

JörgH. Kappes, Bjoern Andres, FredA. Hamprecht, Christoph Schnörr, Sebastian Nowozin, Dhruv Batra, Sungwoong Kim, BernhardX. Kausler, Thorben Kröger, Jan Lellmann, Nikos Komodakis, Bogdan Savchynskyy, and Carsten Rother. A comparative study of modern inference techniques for structured discrete energy minimization problems. *International Journal of Computer Vision*, 115(2):155–184, 2015. ISSN 0920-5691. doi: 10.1007/s11263-015-0809-x.

Michael M. Kazhdan, Thomas A. Funkhouser, and Szymon Rusinkiewicz. Rotation invariant spherical harmonic representation of 3d shape descriptors. In Leif Kobbelt, Peter Schröder, and Hugues Hoppe (eds.), *First Eurographics Symposium on Geometry Processing, Aachen, Germany, June 23-25, 2003*, ACM International Conference Proceeding Series, pp. 156–164. Eurographics Association, 2003. doi: 10.2312/SGP/SGP03/156-165. URL https://doi.org/10.2312/SGP/SGP03/156-165.

Vladimir G. Kim, Wilmot Li, Niloy J. Mitra, Stephen DiVerdi, and Thomas Funkhouser. Exploring collections of 3D models using fuzzy correspondences. *ACM Transactions on Graphics*, 31(4), 2012.

Vladimir G. Kim, Wilmot Li, Niloy J. Mitra, Siddhartha Chaudhuri, Stephen DiVerdi, and Thomas Funkhouser. Learning part-based templates from large collections of 3D shapes. *ACM Transactions on Graphics*, 32(4):70:1–70:12, 2013. doi: 10.1145/2461912.2461933.

Tejas D. Kulkarni, William F. Whitney, Pushmeet Kohli, and Joshua B. Tenenbaum. Deep convolutional inverse graphics network. In *Advances in Neural Information Processing Systems*, volume 28, 2015.

Gim Hee Lee, Friedrich Fraundorfer, and Marc Pollefeys. Robust pose-graph loop-closures with expectation-maximization. In *IROS*, pp. 556–563. IEEE, 2013.

Shaohan Li, Yunpeng Shi, and Gilad Lerman. Efficient detection of long consistent cycles and its application to distributed synchronization. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition, CVPR 2024, Seattle, WA, USA, June 16-22, 2024*, pp. 5260–5269. IEEE, 2024. doi: 10.1109/CVPR52733.2024.00503. URL https://doi.org/10.1109/CVPR52733.2024.00503.

Yangguang Li, Zi-Xin Zou, Zexiang Liu, Dehu Wang, Yuan Liang, Zhipeng Yu, Xingchao Liu, Yuan-Chen Guo, Ding Liang, Wanli Ouyang, and Yan-Pei Cao. TripoSG: High-fidelity 3D shape synthesis using large-scale rectified flow models. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pp. 1–18, 2025. doi: 10.1109/TPAMI.2025.3633512. URL https://doi.org/10.1109/TPAMI.2025.3633512.

Shuyang Ling. Near-optimal performance bounds for orthogonal and permutation group synchronization via spectral methods. *Applied and Computational Harmonic Analysis*, 60:20–52, 2022. doi: 10.1016/j.acha.2022.02.003. URL https://doi.org/10.1016/j.acha.2022.02.003.

Shuyang Ling. Solving orthogonal group synchronization via convex and low-rank optimization: Tightness and landscape analysis. *Mathematical Programming*, 200(1):589–628, 2023. doi: 10.1007/s10107-022-01896-3. URL https://doi.org/10.1007/s10107-022-01896-3.

Xingchao Liu, Chengyue Gong, and Qiang Liu. Flow straight and fast: Learning to generate and transfer data with rectified flow. In *International Conference on Learning Representations*, 2023. URL https://arxiv.org/abs/2209.03003.

Yichong Lu, Yuzhuo Tian, Zijin Jiang, Yikun Zhao, Yuanbo Yang, Hao Ouyang, Haoji Hu, Huimin Yu, Yujun Shen, and Yiyi Liao. Orientation matters: Making 3d generative models orientation-aligned. In D. Belgrave, C. Zhang, H. Lin, R. Pascanu, P. Koniusz, M. Ghassemi, and N. Chen (eds.), *Advances in Neural Information Processing Systems*, volume

38, Main Conference, pp. 148548–148583. Curran Associates, Inc., 2025. doi: 10.52202/085713-4968. URL https://proceedings.neurips.cc/paper_files/paper/2025/file/da75d2bbf862b86f10241d0887613b41-Paper-Conference.pdf.

Lars M. Mescheder, Michael Oechsle, Michael Niemeyer, Sebastian Nowozin, and Andreas Geiger. Occupancy networks: Learning 3d reconstruction in function space. In *IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2019, Long Beach, CA, USA, June 16-20, 2019*, pp. 4460–4470, Washington, DC, USA, 2019. Computer Vision Foundation / IEEE. doi: 10.1109/CVPR.2019.00459. URL http://openaccess.thecvf.com/content_CVPR_2019/html/Mescheder_Occupancy_Networks_Learning_3D_Reconstruction_in_Function_Space_CVPR_2019_paper.html.

Thu Nguyen-Phuoc, Chuan Li, Lucas Theis, Christian Richardt, and Yong-Liang Yang. HoloGAN: Unsupervised learning of 3d representations from natural images. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 7588–7597, 2019.

Kai Ni, Drew Steedly, and Frank Dellaert. Out-of-core bundle adjustment for large-scale 3D reconstruction. In *Proceedings of the IEEE 11th International Conference on Computer Vision (ICCV)*, pp. 1–8. IEEE, 2007.

David Novotný, Nikhila Ravi, Benjamin Graham, Natalia Neverova, and Andrea Vedaldi. C3DPO: canonical 3d pose networks for non-rigid structure from motion. In *2019 IEEE/CVF International Conference on Computer Vision, ICCV 2019, Seoul, Korea (South), October 27 - November 2, 2019*, pp. 7687–7696. IEEE, 2019. doi: 10.1109/ICCV.2019.00778. URL https://doi.org/10.1109/ICCV.2019.00778.

Deepti Pachauri, Risi Kondor, and Vikas Singh. Solving the multi-way matching problem by permutation synchronization. In *Advances in neural information processing systems*, pp. 1860–1868, 2013.

Jeong Joon Park, Peter R. Florence, Julian Straub, Richard A. Newcombe, and Steven Lovegrove. Deepsdf: Learning continuous signed distance functions for shape representation. In *IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2019, Long Beach, CA, USA, June 16-20, 2019*, pp. 165–174, New York, NY, USA, 2019. Computer Vision Foundation / IEEE. doi: 10.1109/CVPR.2019.00025. URL http://openaccess.thecvf.com/content_CVPR_2019/html/Park_DeepSDF_Learning_Continuous_Signed_Distance_Functions_for_Shape_Representation_CVPR_2019_paper.html.

Charles Ruizhongtai Qi, Hao Su, Kaichun Mo, and Leonidas J. Guibas. Pointnet: Deep learning on point sets for 3d classification and segmentation. In *2017 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017, Honolulu, HI, USA, July 21-26, 2017*, pp. 77–85, Washington, DC, USA, 2017a. IEEE Computer Society. doi: 10.1109/CVPR.2017.16. URL https://doi.org/10.1109/CVPR.2017.16.

Charles Ruizhongtai Qi, Li Yi, Hao Su, and Leonidas J Guibas. Pointnet++: Deep hierarchical feature learning on point sets in a metric space. In I. Guyon, U. Von Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett (eds.), *Advances in Neural Information Processing Systems*, volume 30, Red Hook, NY, USA, 2017b. Curran Associates, Inc. URL https://proceedings.neurips.cc/paper_files/paper/2017/file/d8bf84be3800d12f74d8b05e9b89836f-Paper.pdf.

David M Rosen, Luca Carlone, Afonso S Bandeira, and John J Leonard. Se-sync: A certifiably correct algorithm for synchronization over the special euclidean group. *The International Journal of Robotics Research*, 38(2-3):95–125, 2019.

Rahul Sajnani, Adrien Poulenard, Jivitesh Jain, Radhika Dua, Leonidas J. Guibas, and Srinath Sridhar. Condor: Self-supervised canonicalization of 3d pose for partial shapes. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2022.

Yanyao Shen, Qixing Huang, Nati Srebro, and Sujay Sanghavi. Normalized spectral map synchronization. In *Advances in Neural Information Processing Systems*, pp. 4925–4933, 2016.

Ayan Sinha, Jing Bai, and Karthik Ramani. Deep learning 3d shape surfaces using geometry images. In Bastian Leibe, Jiri Matas, Nicu Sebe, and Max Welling (eds.), *Computer Vision - ECCV 2016 - 14th European Conference, Amsterdam, The Netherlands, October 11-14, 2016, Proceedings,*

Part VI, volume 9910 of *Lecture Notes in Computer Science*, pp. 223–240, New York, NY, USA, 2016. Springer. doi: 10.1007/978-3-319-46466-4_14. URL https://doi.org/10.1007/978-3-319-46466-4_14.

Weiwei Sun, Andrea Tagliasacchi, Boyang Deng, Sara Sabour, Soroosh Yazdani, Geoffrey E. Hinton, and Kwang Moo Yi. Canonical capsules: Self-supervised capsules in canonical pose. In *Advances in Neural Information Processing Systems*, volume 34, 2021.

Yifan Sun and Qixing Huang. Pose synchronization under multiple pair-wise relative poses. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition, CVPR 2023, Vancouver, BC, Canada, June 17-24, 2023*, pp. 13072–13081. IEEE, 2023. doi: 10.1109/CVPR52729.2023.01256. URL https://doi.org/10.1109/CVPR52729.2023.01256.

Yifan Sun, Jiacheng Zhuo, Arnav Mohan, and Qixing Huang. K-best transformation synchronization. In *2019 IEEE/CVF International Conference on Computer Vision, ICCV 2019, Seoul, Korea (South), October 27 - November 2, 2019*, pp. 10251–10260. IEEE, 2019. doi: 10.1109/ICCV.2019.01035. URL https://doi.org/10.1109/ICCV.2019.01035.

Niko Sünderhauf and Peter Protzel. Towards a robust back-end for pose graph slam. In *2012 IEEE International Conference on Robotics and Automation*, pp. 1254–1261. IEEE, 2012.

Tencent Hunyuan3D Team. Hunyuan3D 2.0: Scaling diffusion models for high resolution textured 3D assets generation, 2025. URL https://arxiv.org/abs/2501.12202.

Andrea Torsello, Emanuele Rodola, and Andrea Albarelli. Multiview registration via graph diffusion of dual quaternions. In *CVPR 2011*, pp. 2441–2448. IEEE, 2011.

Martin J. Wainwright, Tommi S. Jaakkola, and Alan S. Willsky. A new class of upper bounds on the log partition function. *IEEE Transactions on Information Theory*, 51(7):2313–2335, 2005.

Lanhui Wang and Amit Singer. Exact and stable recovery of rotations for robust synchronization. *Information and Inference: A Journal of the IMA*, 2(2):145–193, 2013.

Jiajun Wu, Chengkai Zhang, Tianfan Xue, Bill Freeman, and Josh Tenenbaum. Learning a probabilistic latent space of object shapes via 3d generative-adversarial modeling. In Daniel D. Lee, Masashi Sugiyama, Ulrike von Luxburg, Isabelle Guyon, and Roman Garnett (eds.), *Advances in Neural Information Processing Systems 29: Annual Conference on Neural Information Processing Systems 2016, December 5-10, 2016, Barcelona, Spain*, pp. 82–90, 2016. URL https://proceedings.neurips.cc/paper/2016/hash/44f683a84163b3523afe57c2e008bc8c-Abstract.html.

Shangzhe Wu, Tomas Jakab, Christian Rupprecht, and Andrea Vedaldi. DOVE: Learning deformable 3d objects by watching videos. *International Journal of Computer Vision*, 131(10): 2623–2634, 2023. doi: 10.1007/s11263-023-01819-5.

Yuezhi Yang, Haitao Yang, Kiyohiro Nakayama, Xiangru Huang, Leonidas Guibas, and Qixing Huang. Genanalysis: Joint shape analysis by learning man-made shape generators with deformation regularizations. *ACM Transactions on Graphics*, 44(4):61:1–61:19, 2025. doi: 10.1145/3731164.

Biao Zhang, Jiapeng Tang, Matthias Nießner, and Peter Wonka. 3DShape2VecSet: A 3D shape representation for neural fields and generative diffusion models. *ACM Trans. Graph.*, 42(4), jul 2023. ISSN 0730-0301. doi: 10.1145/3592442. URL https://doi.org/10.1145/3592442.

Runze Zhang, Siyu Zhu, Tian Fang, and Long Quan. Distributed very large scale bundle adjustment by global camera consensus. In *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, pp. 29–38, 2017.

Qingnan Zhou and Alec Jacobson. Thingi10k: A dataset of 10,000 3d-printing models, 2016. URL https://arxiv.org/abs/1605.04797.

Xiaowei Zhou, Menglong Zhu, and Kostas Daniilidis. Multi-image matching via fast alternating minimization. In *Proceedings of the IEEE International Conference on Computer Vision*, pp. 4032–4040, 2015.

Siyu Zhu, Runze Zhang, Lei Zhou, Tianwei Shen, Tian Fang, Ping Tan, and Long Quan. Very large-scale global sfm by distributed motion averaging. In *2018 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2018, Salt Lake City, UT, USA, June 18-22, 2018*, pp. 4568–4577. Computer Vision Foundation / IEEE Computer Society, 2018. doi: 10.1109/CVPR.2018.00480. URL http://openaccess.thecvf.com/content_cvpr_2018/html/Zhu_Very_Large-Scale_Global_CVPR_2018_paper.html.

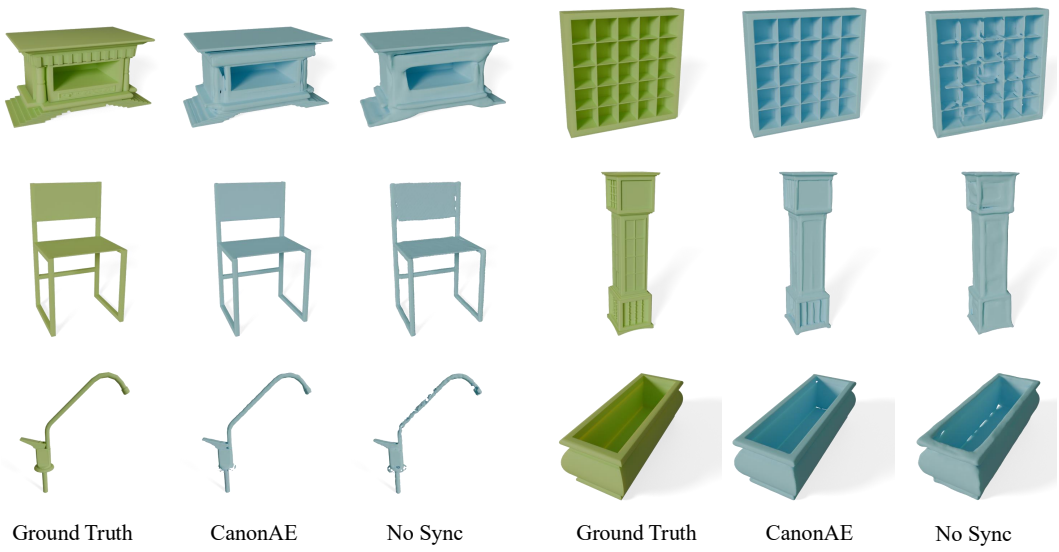


Figure 6: More qualitative results comparing the reconstruction quality of the auto-encoder trained using CanonAE poses vs unsynchronized poses on held-out test samples.

A Qualitative Results

A.1 Qualitative AE reconstruction results on test data

Across the presented examples in Fig. 6, the reconstructions by the AE trained via our approach recover more detail and feature artifacts less often than the reconstructions by the control.

Training Details for PointNet+DeepSDF-based autoencoders (AE-I). The encoder maps 2048 points through shared layers of widths $(64, 128, 1024)$, global max pooling, and fully connected layers $(512, 64)$. The decoder uses eight 1024-wide hidden layers, eight Fourier frequency bands, Softplus activations, weight normalization, and a midpoint input skip. We train with AdamW ($\beta = (0.9, 0.999)$), encoder/decoder learning rates $10^{-4}/(5 \times 10^{-4})$, and gradient clipping at 1. Shape runs use 200 expected data passes, batch size 96, five-pass warmup, and cosine decay. ProteinD uses 200 epochs, batch size 8, constant learning rates, density-valued encoder inputs, and a linear density output.

Training Details for 3DShape2VecSet-based autoencoders (AE-II). The training is conducted for 170 epochs with a learning rate of 5×10^{-5} and an effective batch size of 224. We use the AdamW optimizer with $\beta = (0.9, 0.95)$ and weight decay of 0.05. We use the architectural variant that uses furthest point sample queries in the encoder that produces 1024 latent sets of 32 dimensions each. The decoder contains 24 self-attention layers.

A.2 Details on generative model and further qualitative results

Fig. 7 presents further qualitative results from the generative model trained on the latent space encoded by CanonAE and the baseline method (No Sync) on a set of classes with wide variation in complexity. Shapes generated by CanonAE show fewer artifacts and better preservation of complex structures than the baseline. Implementation details are as below.

Generator. The generator is a permutation-equivariant Diffusion Transformer over the latent set. Relative to image DiT we remove patchification and positional embeddings: the network operates directly on the 1024×32 VecSet token set with self-attention only, making it equivariant to token permutation. Each block is a standard adaLN-Zero transformer block, with per-head query/key LayerNorm for stability. We use the 'B' configuration (hidden dim 768, depth 12, 12 heads; ≈130M parameters). Timestep and class conditioning are injected through adaptive LayerNorm. For class-conditional generation we use a 55-way class embedding with a null token and 10% label dropout,

enabling classifier-free guidance (CFG) at inference. We train with a rectified flow loss. Timesteps are sampled from a logit-normal distribution.

Optimization. Latents are pre-computed with the frozen encoder. We use AdamW (lr $= 10^{-4}$, an effective batch size of 256, $\beta = (0.9, 0.95)$, weight decay 0), 2,000-step warmup, and a warmup–stable–decay (WSD) schedule: constant learning rate after warmup, with a cosine decay to 10^{-6} over the final 20% of steps. Models are trained for 100k optimizer steps on ShapeNet-55.

Data. We use latents derived from ShapeNet-55 surface point clouds to train both models. The primary difference between No Sync and CanonAE is in how the latents were derived. In No Sync, latents are encoded from shapes in their default orientations by the public off-the-shelf VecSet encoder, whereas in CanonAE, latents were derived from shapes in their CanonAE poses with a VecSet autoencoder trained on CanonAE poses on Canoverse.

B Timings and Additional Ablations

B.1 Timing

Pose synchronization. Each CPU machine provided 80 physical cores and 256 GB RAM. Thingi10K and ProteinD used 24 workers on one machine. CanoVerse used 12 machines for matching and patch solves, eight for local refinement, and one for other stages. Node-hours sum allocated machine time; wall time accounts for parallel execution.

Dataset	Concurrency	Node-hours	Wall time
Thingi10K	1	0.594	00:35:39
ProteinD	1	3.193	03:11:33
CanoVerse	1–12	119.497	18:10:12

Table 4: RotSync computational cost; wall time is hh:mm:ss. Different stages of the CanoVerse processing used a different concurrency; the computation of the pairwise candidate rotations between shapes connected by an edge after clustering was highly parallelizable and hence used 12 nodes.

B.2 Additional Ablations

Table 5 shows that relative MAE test errors without pose optimization during AE training. We can see that the MAE test errors only increase slightly and are consistent with the relative differences in Tab. 3. This again indicates that the output of pose synchronization is largely unchanged during joint AE training.

	Thingi10K		ProteinD		CanoVerse	
Loss	AEI-T	AEII-T	AEI-T	AEII-T	AEI-T	AEII-T
No synchronization	0.52%	0.24%	0.61%	0.26%	0.45%	0.32%
Stage-III only	0.31%	0.27%	0.46%	0.35%	0.29%	–
Zhu et al.	0.41%	0.34%	0.65%	0.48%	0.53%	–

Table 5: This table reports the relative MAE test errors without pose optimization during AE training.

C Stability under a Change of Matching Loss

Let $S_1, \dots, S_n$ be shapes with rotations $R_i \in SO(3)$, and write $R = (R_1, \dots, R_n)$. Consider a fixed connected undirected graph $G = (V, E, w)$, where $V = \{1, \dots, n\}$, $n \geq 2$, and $w_{ij} > 0$ for every edge. Orient each edge once. For $e = (i, j)$ and $U \in SO(3)$, define

$$\phi_{k,e}(U) = D_k(US_i, S_j), \qquad U_{ij} = R_j^\top R_i, \qquad k \in \{0, 1\}.$$

Assume $D_k(Y, Z) = D_k(Z, Y)$ and $D_k(QY, QZ) = D_k(Y, Z)$ for every $Q \in SO(3)$. The synchronization objective is

$$F_k(R) = \sum_{(i,j)\in E} w_{ij}\phi_{k,ij}(R_j^\top R_i) = \sum_{(i,j)\in E} w_{ij} D_k(R_i S_i, R_j S_j).$$

Changing the matching loss can change the preferred relative rotation on each edge. These changes are coupled through the shared node rotations. Starting from a local minimizer R^0 under D_0, we use

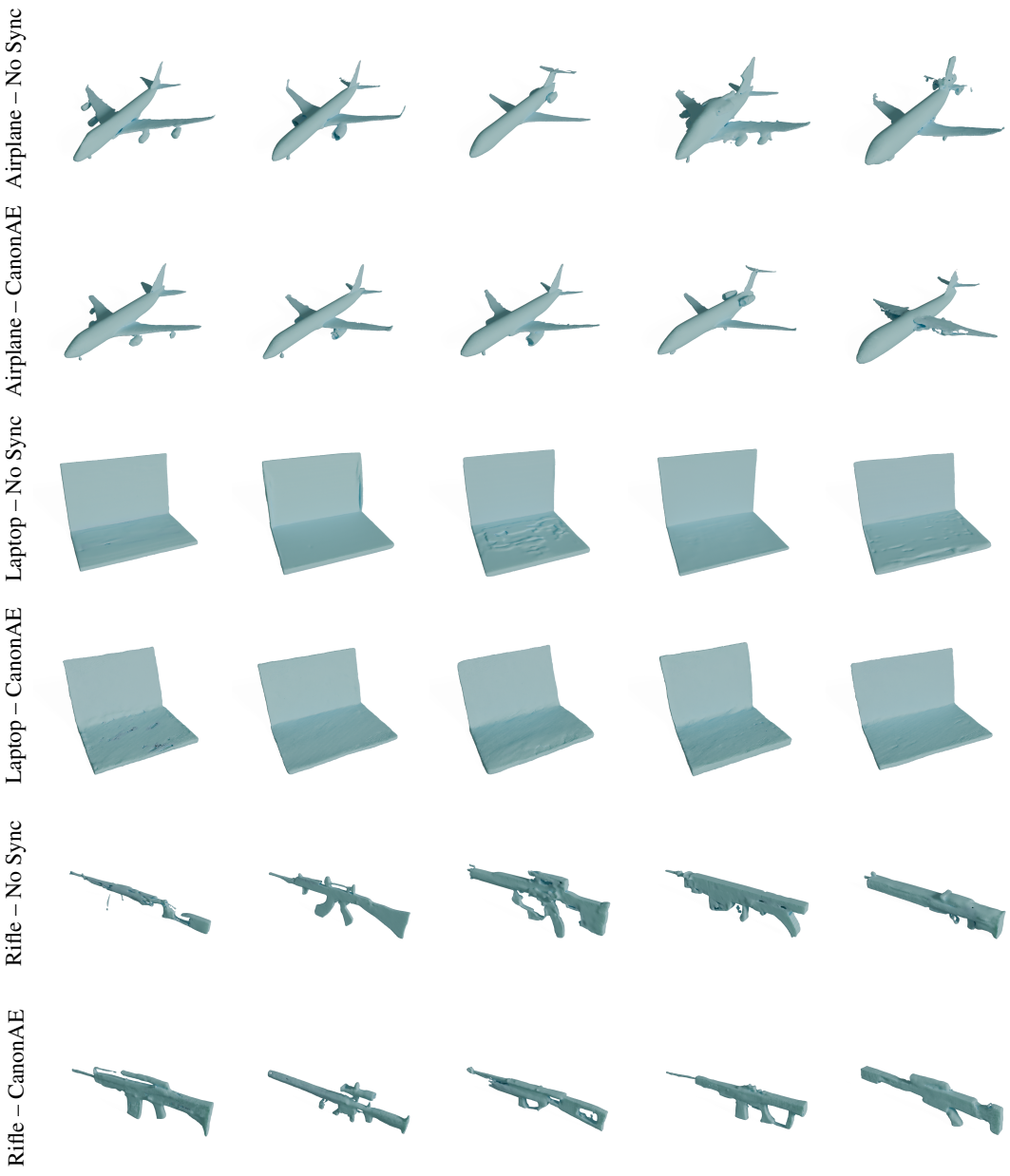


Figure 7: Qualitative comparison of generated samples across ShapeNet object classes. Each category has No Sync examples above and CanonAE samples below. Shapes are randomly sampled. Best viewed zoomed-in.

a quadratic model to bound the node displacement under D_1. The graph enters this bound through cancellation of incident edge gradients and curvature contributed by neighboring edges.

C.1 LOCAL QUADRATIC MODEL

Let $x_i \in \mathbb{R}^3$ be the rotation increment of node i, and stack them into $x \in \mathbb{R}^{3n}$. For $u \in \mathbb{R}^3$, $[u]_\times$ is the skew-symmetric matrix satisfying $[u]_\times v = u \times v$. The matrix exponential gives the rotation $\exp([u]_\times)$. Vector norms below are Euclidean. Write

$$R_i(x_i) = \exp([x_i]_\times) R_i^0, \qquad Y_i = R_i^0 S_i.$$

The fixed shapes Y_i express the reference alignment in a common coordinate frame. Define the rotational response

$$\ell_{k,e}(\omega) = D_k(\exp([\omega]_\times) Y_i, Y_j)\,,$$

assume it is C^2 near $\omega = 0$, and evaluate

$$a_{k,e} = \nabla \ell_{k,e}(0), \qquad C_e = \nabla^2 \ell_{1,e}(0).$$

Here $a_{k,e} \in \mathbb{R}^3$ is an edge gradient, and $C_e \in \mathbb{R}^{3\times 3}$ is the target loss's full Hessian with respect to the edge's relative rotation.

To first order, the relative rotation increment on edge (i, j), expressed in this common frame, is $x_i - x_j$. Define the weighted incidence operator B by

$$(Bx)_{ij} = \sqrt{w_{ij}}(x_i - x_j), \qquad B^\top B = L \otimes I_3 =: M,$$

Here $L_{ii} = \sum_{j\in\mathcal{N}_i} w_{ij}$ and $L_{ij} = -w_{ij}$ on edges, with $\mathcal{N}_i$ the neighbors of i; other off-diagonal entries are zero. The symbol $\otimes$ denotes the Kronecker product and I_3 the 3×3 identity. A vector of weighted edge increments $z \in \mathbb{R}^{3|E|}$ is compatible if $z = Bx$ for some node increments x. Stack the edge Hessians and gradient changes:

$$C = \mathrm{blkdiag}(C_e), \qquad \xi = \left(\sqrt{w_e}(a_{1,e} - a_{0,e})\right)_{e\in E}.$$

Here blkdiag forms a block-diagonal matrix. Stationarity of R^0 gives $B^\top(\sqrt{w_e}a_{0,e})_{e\in E} = 0$. Substituting the linearized increments into the second-order expansions of the target edge losses defines the quadratic model

$$\begin{aligned} q(x) &= F_1(R^0) + \xi^\top Bx + \tfrac{1}{2}(Bx)^\top C(Bx) \\ &= F_1(R^0) + g^\top x + \tfrac{1}{2}x^\top Hx, \end{aligned}$$

where

$$\boxed{g = B^\top \xi, \qquad H = B^\top CB.} \tag{10}$$

The model q uses the full edge Hessians and the linearized relation $z = Bx$. Section C.5 retains the additional quadratic contribution from nonlinear rotation composition.

Remove the common infinitesimal rotation by restricting x to

$$\mathcal{V} = \left\{ x \in \mathbb{R}^{3n} : \sum_i x_i = 0 \right\}.$$

Matrix comparisons and inverses involving H are restricted to $\mathcal{V}$. The notation $\succeq$ denotes the positive-semidefinite ordering. Assume H is positive definite on $\mathcal{V}$. The quadratic minimizer and its predicted edge displacement are

$$\boxed{x_q = -H^{-1}g, \qquad z_q = -BH^{-1}g,} \tag{11}$$

The graph affects this response through both g and H.

C.2 THE GRADIENT: CANCELLATION AT SHARED NODES

At each node, incident edge gradients are added with their incidence signs:

$$g_i = \sum_{e \ni i} s_{ie} w_e (a_{1,e} - a_{0,e}),$$

where $s_{ie} = 1$ at the first endpoint of an oriented edge and -1 at the second. These vectors are summed before taking a norm. Large changes in individual edge gradients can therefore cancel at a shared node, leaving a small assembled gradient.

Define its size relative to the graph Laplacian by

$$\eta^2 = g^\top M^\dagger g, \qquad M^\dagger = L^\dagger \otimes I_3,$$

where † denotes the Moore–Penrose inverse. Defining $P = BM^\dagger B^\top$, we obtain

$$\boxed{\eta^2 = \xi^\top P \xi = \|P\xi\|^2 \le \|\xi\|^2.} \tag{12}$$

Since $M = B^\top B$, P is the orthogonal projector onto $\mathrm{range}(B)$, the space of compatible edge increments. The projected gradient change $P\xi$ is the part that contributes to changes in the node rotations. Thus the numerator η can be small even when individual edge-gradient changes are large, because their incompatible components produce no net node gradient. In particular, if $D_1 = cD_0$ for $c > 0$, stationarity under D_0 gives $g = 0$, even though individual edge gradients may change. Under the stated positive-curvature assumption, the predicted displacement is then zero.

C.3 THE HESSIAN: COMPLEMENTARY CONSTRAINTS ACROSS EDGES

For independent edge increments, C is the curvature matrix and $\lambda_{\min}(C)$ is its smallest eigenvalue. Its block structure gives

$$\lambda_{\min}(C) = \min_{e \in E} \lambda_{\min}(C_e).$$

A weak rotational direction on just one edge can therefore make this quantity very small. Synchronization assembles these Hessians through shared node variables:

$$H = B^\top C B, \qquad H_{ii} = \sum_{j \in \mathcal{N}_i} w_{ij} C_{ij}, \qquad H_{ij} = -w_{ij} C_{ij},$$

Here H_{ij} is the 3×3 block for nodes i, j. Symmetry and common-rotation invariance give $\ell_{k,ji}(\omega) = \ell_{k,ij}(-\omega)$, so the Hessian C_{ij} is unchanged when the edge orientation is reversed. Off-diagonal blocks vanish for nonedges. Different edges contribute curvature to the same node rotation. A direction weakly constrained by one edge may be constrained by others. The assembled matrix H determines the curvature of q.

The minimum curvature over compatible increments is

$$\boxed{\gamma = \min_{x \in \mathcal{V} \setminus \{0\}} \frac{x^\top H x}{x^\top M x} = \min_{\substack{z \in \mathrm{range}(B) \\ z \neq 0}} \frac{z^\top C z}{\|z\|^2}.} \tag{13}$$

Connectivity makes $x^\top M x > 0$ for nonzero $x \in \mathcal{V}$, and positive definiteness of H gives $\gamma > 0$. Restricting the minimum to compatible increments gives

$$\gamma \ge \lambda_{\min}(C),$$

and the inequality can be strict when the weak edge directions cannot occur together in a compatible increment. Complementary edge constraints can therefore keep the denominator's curvature factor γ positive even when the individual edge Hessians are singular, as the following example shows.

For example, on an unweighted triangle, let

$$C_{12} = \mathrm{diag}(0, 1, 1), \qquad C_{23} = \mathrm{diag}(1, 0, 1), \qquad C_{31} = \mathrm{diag}(1, 1, 0).$$

Every edge Hessian is singular, so $\lambda_{\min}(C) = 0$. The cycle constraint

$$z_{12} + z_{23} + z_{31} = 0$$

forces any increment in an unconstrained direction to be balanced by increments on the other edges. In each coordinate, the minimum Rayleigh quotient under this constraint is $1/3$, so $\gamma = 1/3$. On a tree, every edge increment is compatible, so $\gamma = \lambda_{\min}(C)$.

The curvature condition can equivalently be written as

$$\boxed{H \succeq L \otimes A, \qquad A = \gamma I_3.} \tag{14}$$

This choice maximizes $\lambda_{\min}(A)$ among all feasible positive-definite 3×3 matrices. Indeed, any feasible $\widetilde{A}$ satisfies

$$H \succeq L \otimes \widetilde{A} \succeq \lambda_{\min}(\widetilde{A})(L \otimes I_3),$$

so $\lambda_{\min}(\widetilde{A}) \leq \gamma$, with equality attained by $A = \gamma I_3$. This gives the strongest scalar lower bound relative to the graph Laplacian.

C.4 Node displacement

Since $g = B^\top \xi \in \mathcal{V}$, weighted Cauchy–Schwarz on $\mathcal{V}$ and $Hx_q = -g$ give

$$\gamma \|Bx_q\|^2 \leq x_q^\top H x_q = -x_q^\top g \leq \|Bx_q\| \sqrt{g^\top M^\dagger g}.$$

It follows that

$$\|z_q\| = \|Bx_q\| \leq \frac{\sqrt{g^\top (L^\dagger \otimes I_3) g}}{\gamma} = \frac{\eta}{\gamma}. \tag{15}$$

To control the node rotations themselves, use $x_q \in \mathcal{V}$ and connectivity:

$$\|Bx_q\|^2 = x_q^\top (L \otimes I_3) x_q \geq \lambda_2(L) \|x_q\|^2,$$

where $\lambda_2(L) > 0$ is the second-smallest eigenvalue of L, its spectral gap. Combining these inequalities gives the node-displacement bound

$$\boxed{\|x_q\| \leq \frac{\eta}{\gamma \sqrt{\lambda_2(L)}} = \frac{\|P\xi\|}{\gamma \sqrt{\lambda_2(L)}}.} \tag{16}$$

This controls the node rotation increments after fixing the common rotation by $\sum_i x_{q,i} = 0$. In the predicted configuration $R_i = \exp([x_{q,i}]_\times) R_i^0$, $\|x_{q,i}\|$ equals the angular displacement of shape i whenever it is less than π. The numerator measures the edge gradient change after cancellation at shared nodes. The denominator combines the curvature of compatible increments with graph connectivity. Their ratio bounds the predicted change in node poses.

C.5 Quadratic correction from rotation composition

The preceding model uses the linearized relative increment $x_i - x_j$. To retain the full second-order contribution in node coordinates, common-rotation invariance gives

$$D_1(\exp([x_i]_\times) Y_i, \exp([x_j]_\times) Y_j) = \ell_{1,e}(\omega_{ij}(x)), \qquad [\omega_{ij}(x)]_\times = \log(\exp(-[x_j]_\times) \exp([x_i]_\times)).$$

Expanding the rotation product and its logarithm at zero yields

$$\omega_{ij}(x) = x_i - x_j + \tfrac{1}{2} x_i \times x_j + O(\|x\|^3).$$

The cross term multiplies the edge gradient $a_{1,e}$ in the loss expansion and therefore contributes at quadratic order. Consequently,

$$\begin{aligned} q_{\text{full}}(x) &= q(x) + \tfrac{1}{2} \sum_{e=(i,j) \in E} w_e a_{1,e}^\top (x_i \times x_j) \\ &= F_1(R^0) + g^\top x + \tfrac{1}{2} x^\top \widehat{H} x, \qquad \widehat{H} = H + \Delta H, \\ F_1(R(x)) &= q_{\text{full}}(x) + o(\|x\|^2). \end{aligned}$$

Each oriented edge contributes the following symmetric block matrix to ΔH at its two endpoints:

$$(\Delta H)_e = \frac{w_e}{2} \begin{pmatrix} 0 & -[a_{1,e}]_\times \\ [a_{1,e}]_\times & 0 \end{pmatrix}.$$

Thus $g = B^\top \xi$ and its projector interpretation are unchanged, whereas the node Hessian includes an additional, possibly indefinite, term. This term vanishes if every target edge gradient is zero; stationarity of the assembled objective alone does not imply this edgewise condition. Shrinking x alone does not remove it relative to the other quadratic terms.

Let $\widehat{H}_{\mathcal{V}} v = \Pi_{\mathcal{V}} \widehat{H} v$ for $v \in \mathcal{V}$, where $\Pi_{\mathcal{V}}$ is the orthogonal projector onto $\mathcal{V}$, and define

$$\widehat{\gamma} = \min_{v \in \mathcal{V} \setminus \{0\}} \frac{v^\top \widehat{H} v}{v^\top M v}.$$

If $\widehat{\gamma} > 0$, the restricted quadratic model has the unique minimizer $x_{\mathrm{full},q} = -\widehat{H}_{\mathcal{V}}^{-1} g$. The same weighted Cauchy–Schwarz argument used above gives

$$\widehat{\gamma} \|B x_{\mathrm{full},q}\|^2 \le -x_{\mathrm{full},q}^\top g \le \eta \|B x_{\mathrm{full},q}\|, \qquad \boxed{\|x_{\mathrm{full},q}\| \le \frac{\eta}{\widehat{\gamma}\sqrt{\lambda_2(L)}}.}$$

The numerator therefore retains the same cancellation mechanism, while the denominator must account for the corrected curvature.

A sufficient condition can be obtained without diagonalizing $\widehat{H}$. Set

$$\delta = \sup_{v \in \mathcal{V} \setminus \{0\}} \frac{|v^\top \Delta H v|}{v^\top M v}, \qquad \alpha = \max_{e \in E} \|a_{1,e}\|, \qquad d_{\max} = \max_i \sum_{j \in \mathcal{N}_i} w_{ij}.$$

Since

$$|v^\top \Delta H v| \le \frac{\alpha}{2} \sum_{(i,j) \in E} w_{ij} (\|v_i\|^2 + \|v_j\|^2) \le \frac{\alpha d_{\max}}{2\lambda_2(L)} v^\top M v,$$

we have $\delta \le \alpha d_{\max}/(2\lambda_2(L))$ and $\widehat{\gamma} \ge \gamma - \delta$. Hence $\delta < \gamma$ suffices, and replacing $\widehat{\gamma}$ by $\gamma - \delta$ gives a conservative displacement bound. In particular, $\alpha d_{\max} < 2\gamma\lambda_2(L)$ suffices. Small edge gradients relative to the compatible curvature justify treating the correction as a perturbation. When this sufficient condition fails, the corrected curvature must be checked directly. Both bounds concern quadratic-model predictions.

D Global Optimality of Geometric Alignment

We identify a region containing every global minimizer of the geometric objective and prove that it contains exactly one stationary point. The proof has two steps: an objective-value comparison places all global minimizers in the region, and a uniform Hessian bound establishes uniqueness there. We compare configurations modulo a common rotation.

Use the shapes, rotations, and graph notation of Appendix C. On a connected graph $G = (V, E, w)$ with positive weights and Laplacian L, count each edge once and consider

$$F_D(R) = \sum_{(i,j) \in E} w_{ij} D(R_i S_i, R_j S_j).$$

Assume D is nonnegative, symmetric, continuous under rotations, satisfies $D(Y, Y) = 0$, and is invariant to a common rotation. Let d be a common-rotation-invariant metric on shapes and $\theta(Q) \in [0, \pi]$ the angle of $Q \in SO(3)$, in radians. The angular distance between P and Q is $\theta(P^\top Q)$.

D.1 Assumptions

We assume that neighboring shapes admit similar reference poses, self-matching distinguishes rotations, and the loss and its curvature vary continuously at controlled rates.

1. *Compatible reference poses.* There exist rotations $\bar{R}_i$ such that the reference-oriented shapes $Y_i = \bar{R}_i S_i$ satisfy

   $$d(Y_i, Y_j) \le \varepsilon_{ij} \quad ((i,j) \in E), \qquad \varepsilon = \max_{(i,j) \in E} \varepsilon_{ij}.$$

   These reference rotations align all neighboring pairs simultaneously. Choose their common orientation so that $\bar{R}_1 = I_3$.

2. *Rotational identifiability.* For some $\kappa > 0$, every reference shape Y satisfies
$$D(QY, Y) \geq \frac{\kappa}{2}\theta(Q)^2, \qquad Q \in SO(3). \tag{D.1}$$
Every nonidentity rotation has a positive self-matching cost, with a quadratic lower bound in its angle.
3. *Controlled loss and curvature changes.* For a pair of shapes, define
$$\ell_{Y,Z}(u, v) = D(\exp([u]_\times)Y, \exp([v]_\times)Z), \qquad \mathcal{H}_{Y,Z}(u, v) = \nabla^2_{(u,v)}\ell_{Y,Z}(u, v).$$
The two rotation increments $u, v \in \mathbb{R}^3$ act on the endpoints of an edge. The 6×6 Hessian $\mathcal{H}_{Y,Z}$ contains their individual curvature and their cross derivatives. When $Z = Y$, the two increments rotate separate copies of the same shape.
Fix $0 < r_c < \pi$. Assume each response is C^2 on an open neighborhood of $\|u\|, \|v\| \leq r_c$. For reference pairs Y, Z, suppose uniform constants $C_0 > 0$ and $C_2, K \geq 0$ satisfy the following bounds for all $P, Q \in SO(3)$ and all u, v in this region:
$$\begin{aligned} |D(PY, QZ) - D(PY, QY)| &\leq C_0 d(Y, Z), \\ \|\mathcal{H}_{Y,Z}(u, v) - \mathcal{H}_{Y,Y}(u, v)\|_{\mathrm{op}} &\leq C_2 d(Y, Z), \\ \|\mathcal{H}_{Y,Y}(u, v) - \mathcal{H}_{Y,Y}(0, 0)\|_{\mathrm{op}} &\leq K \max\{\|u\|, \|v\|\}. \end{aligned} \tag{D.2}$$
Here $\|\cdot\|_{\mathrm{op}}$ is the Euclidean operator norm. The first two inequalities control changes caused by replacing a shape; the third controls changes caused by rotation. The Hessians are taken with respect to (u, v).

D.2 Global optimality theorem

Write $R_i = Q_i \bar{R}_i$ and fix the common rotation by $Q_1 = I_3$. For $0 < r \leq r_c$, define
$$\mathcal{K}_r = \{z \in \mathbb{R}^{3n} : z_1 = 0,\ \max_i \|z_i\| \leq r\}, \qquad \mathcal{U}_r = \left\{R_i = \exp([z_i]_\times)\bar{R}_i :\ z \in \mathcal{K}_r\right\}.$$
Thus $\mathcal{U}_r$ allows each node to rotate by up to r from its reference pose, with node 1 fixed. Interiors and stationary points are understood in these fixed-node coordinates: $\operatorname{int}\mathcal{U}_r$ corresponds to $\max_i \|z_i\| < r$, and stationarity means zero derivatives with respect to the free variables $z_2, \ldots, z_n$. The coordinates are nonsingular because $r < \pi$, and $\mathcal{K}_r$ is convex. Define two graph constants:
$$\tau_G = \max_i \min_{p:1\rightsquigarrow i} \sum_{e\in p} \frac{1}{w_e}, \qquad \chi_G = \sup_{\substack{a\in\mathbb{R}^n\setminus\{0\} \\ a_1=0}} \frac{\sum_{(i,j)\in E} w_{ij}(a_i^2 + a_j^2)}{a^\top L a}.$$
Here $p : 1 \rightsquigarrow i$ is a graph path from node 1 to node i. Connectivity makes both constants finite. The path constant τ_G converts a bound on edge rotations into a bound on each node rotation. The factor χ_G bounds the accumulated Hessian perturbation relative to the graph energy.

Theorem D.1 (A region with a unique global optimum) *Under Assumptions 1–3, set*
$$\eta = C_0 \sum_{(i,j)\in E} w_{ij}\varepsilon_{ij}, \qquad \mu = \kappa - \chi_G(Kr + C_2\varepsilon).$$
If
$$\boxed{\eta < \frac{\kappa r^2}{4\tau_G}, \qquad \mu > 0,} \tag{D.3}$$
then every global minimizer of F_D lies in $\operatorname{int}\mathcal{U}_r$, and $\mathcal{U}_r$ contains exactly one stationary point. Consequently, this point is the unique global minimizer modulo common rotation. Any local minimizer of the original objective in this region is therefore globally optimal.

For $K > 0$, any $r \leq r_c$ satisfying
$$\sqrt{\frac{4\tau_G\eta}{\kappa}} < r < \frac{\kappa/\chi_G - C_2\varepsilon}{K}$$
meets the theorem's conditions. If $K = 0$, only the lower bound and $C_2\varepsilon < \kappa/\chi_G$ are needed. For $K > 0$ with the other constants fixed, as neighboring shapes become closer, the lower threshold tends to zero and the upper threshold tends to $\kappa/(K\chi_G)$. This gives a finite angular neighborhood whenever the interval is nonempty.

D.3 PROOF

STEP 1: ALL GLOBAL MINIMIZERS LIE IN THE REGION

Common-rotation invariance, Assumption 3, and Assumption 2 give, on each edge,

$$D(Q_iY_i, Q_jY_j) \geq D(Q_j^\top Q_iY_i, Y_i) - C_0\varepsilon_{ij} \geq \frac{\kappa}{2}\theta(Q_j^\top Q_i)^2 - C_0\varepsilon_{ij}.$$

With $\mathcal{E}(Q) = \sum_{(i,j)\in E} w_{ij}\theta(Q_j^\top Q_i)^2$, summing yields

$$F_D(R) \geq \frac{\kappa}{2}\mathcal{E}(Q) - \eta, \qquad F_D(\bar{R}) \leq \eta. \tag{D.4}$$

The second inequality follows by comparing each reference pair with zero-cost self-matching. Along a path p from the fixed node to i, the angular triangle inequality and weighted Cauchy–Schwarz give

$$\theta(Q_i)^2 \leq \left(\sum_{(a,b)\in p}\theta(Q_b^\top Q_a)\right)^2 \leq \left(\sum_{e\in p}\frac{1}{w_e}\right)\mathcal{E}(Q) \leq \tau_G\mathcal{E}(Q),$$

where the last bound uses a minimizing path. If $R \notin \operatorname{int}\mathcal{U}_r$, some node has angle at least r, and hence

$$F_D(R) \geq \frac{\kappa r^2}{2\tau_G} - \eta > \eta \geq F_D(\bar{R}).$$

Such a configuration cannot be globally optimal. Continuity on the compact rotation space guarantees existence of a global minimizer, so every global minimizer lies in the interior of the stated region.

STEP 2: THE STATIONARY POINT IS UNIQUE

We first bound the Hessian throughout $\mathcal{K}_r$. For self-matching, define the loss as a function of one rotation and its 3×3 Hessian

$$h_Y(\omega) = D(\exp([\omega]_\times)Y, Y), \qquad C_Y = \nabla^2 h_Y(0).$$

Nonnegativity and $h_Y(0) = 0$ give $\nabla h_Y(0) = 0$. For small ω, $\theta(\exp([\omega]_\times)) = \|\omega\|$. Comparing the Taylor expansion $h_Y(\omega) = \frac{1}{2}\omega^\top C_Y\omega + o(\|\omega\|^2)$ with (D.1) gives $C_Y \succeq \kappa I_3$.

To obtain the two-endpoint Hessian, use common-rotation invariance to write

$$\ell_{Y,Y}(u,v) = h_Y(\omega(u,v)), \qquad [\omega(u,v)]_\times = \log(\exp(-[v]_\times)\exp([u]_\times)).$$

Here $\log$ is the matrix logarithm near the identity. The derivative of the local relative-rotation vector ω at the origin is $[I_3\ -I_3]$. The second-derivative terms of ω in the chain rule are multiplied by $\nabla h_Y(0) = 0$, so they vanish at this self-matching configuration. Therefore,

$$\mathcal{H}_{Y,Y}(0,0) = \begin{pmatrix} I_3 \\ -I_3 \end{pmatrix} C_Y \begin{pmatrix} I_3 & -I_3 \end{pmatrix} = \begin{pmatrix} C_Y & -C_Y \\ -C_Y & C_Y \end{pmatrix}. \tag{D.5}$$

Its quadratic form is $(v_i - v_j)^\top C_Y(v_i - v_j)$. It penalizes relative rotation; rotating both copies together remains a zero direction.

For an edge between neighboring shapes at any $z \in \mathcal{K}_r$, add and subtract $\mathcal{H}_{Y_i,Y_i}(z_i, z_j)$. The two Hessian bounds in (D.2) imply

$$\begin{aligned} \mathcal{H}_{Y_i,Y_j}(z_i,z_j) &= \mathcal{H}_{Y_i,Y_i}(0,0) + \Delta_{ij}(z), \\ \|\Delta_{ij}(z)\|_{\mathrm{op}} &\leq C_2\varepsilon_{ij} + Kr. \end{aligned} \tag{D.6}$$

Summing the perturbed edge Hessians gives the node Hessian. Write

$$\widetilde{F}_D(z) = \sum_{(i,j)\in E} w_{ij}\ell_{Y_i,Y_j}(z_i,z_j), \qquad M = L\otimes I_3, \qquad v_{ij} = \begin{pmatrix} v_i \\ v_j \end{pmatrix}.$$

For every direction $v = (v_1, \dots, v_n) \in \mathbb{R}^{3n}$ with $v_1 = 0$, the node Hessian satisfies

$$\begin{aligned} v^\top \nabla^2 \widetilde{F}_D(z) v &= \sum_{(i,j)\in E} w_{ij} v_{ij}^\top \mathcal{H}_{Y_i,Y_j}(z_i, z_j) v_{ij} \\ &\geq \kappa\, v^\top M v - (Kr + C_2\varepsilon) \sum_{(i,j)\in E} w_{ij}(\|v_i\|^2 + \|v_j\|^2) \\ &\geq \mu\, v^\top M v. \end{aligned} \tag{D.7}$$

The middle inequality uses $h^\top \Delta h \geq -\|\Delta\|_{\mathrm{op}}\|h\|^2$; the last applies the definition of χ_G to each coordinate of v. Connectivity and $v_1 = 0$ make $v^\top M v > 0$ for every nonzero v. The lower bound therefore holds with positive curvature throughout $\mathcal{K}_r$.

Suppose two different stationary points z^a, z^b existed there, and set $s = z^b - z^a \neq 0$. Convexity of $\mathcal{K}_r$ keeps their connecting line segment inside it. Since $s_1 = 0$, stationarity and integration of (D.7) give

$$\begin{aligned} 0 &= s^\top \big(\nabla \widetilde{F}_D(z^b) - \nabla \widetilde{F}_D(z^a)\big) \\ &= \int_0^1 s^\top \nabla^2 \widetilde{F}_D(z^a + ts)\, s\, dt \\ &\geq \mu\, s^\top M s > 0, \end{aligned} \tag{D.8}$$

a contradiction. The interior global minimizer from Step 1 is stationary, and Step 2 shows it is the only stationary point in the region. This proves the theorem. □

The theorem applies to a quadratic prediction from Appendix C when that configuration lies in $\mathcal{U}_r$ and satisfies the stationarity equations of F_D.

E Stability from Geometric Alignment to AE Learning

Geometric alignment provides the initial poses for joint AE training. Appendix C bounds the quadratic pose response when the geometric loss changes from D_0 to D_1, and Appendix D gives conditions for global optimality of a geometric stationary point. We now compare geometric alignment and AE learning under the same loss D_1. The quadratic pose response depends on the reconstruction gradient and the curvature remaining after parameter fitting. Shared parameters couple the shapes and constrain which rotations the network can absorb.

E.1 Local quadratic model

Let R^1 be a stationary geometric alignment under the target loss D_1. Recenter the coordinates of Appendix C at this alignment:

$$R_i(x_i) = \exp([x_i]_\times) R_i^1, \qquad Y_i = R_i^1 S_i, \qquad Y_i(x_i) = \exp([x_i]_\times) Y_i.$$

Thus $x = 0$ denotes the geometric stationary alignment under D_1. We fix the common orientation and use the pose space $\mathcal{V} = \{x : \sum_i x_i = 0\}$ from Appendix C. All pose derivatives and inverses below are restricted to this space.

For this analysis, specialize D_1 to a squared reconstruction loss on a common d-dimensional vector representation. Let $T_\theta : \mathbb{R}^d \to \mathbb{R}^d$ be the encoder–decoder map, with local parameter coordinates $\theta \in \mathbb{R}^p$, and define

$$\begin{aligned} &e_i(\theta, x_i) = Y_i(x_i) - T_\theta(Y_i(x_i)), \\ &\mathcal{J}_i(\theta, x_i) = D_1\big(Y_i(x_i), T_\theta(Y_i(x_i))\big) = \tfrac{1}{2}\|e_i(\theta, x_i)\|^2, \qquad \mathcal{J} = \sum_i \mathcal{J}_i. \end{aligned}$$

A common rotation Q is a gauge symmetry when the network admits a smooth reparameterization $\theta \mapsto \theta_Q$ satisfying $T_{\theta_Q}(QY) = QT_\theta(Y)$ for all inputs Y. In this case, restricting poses to $\mathcal{V}$ fixes the common rotational freedom.

Fit the network at $x = 0$ to a parameter stationary point θ_0, so $\nabla_\theta \mathcal{J}(\theta_0, 0) = 0$, and assume the residual maps are twice continuously differentiable nearby. When a continuously differentiable branch $\theta_*(x)$ of local parameter minima passes through θ_0, define $\Phi(x) = \mathcal{J}(\theta_*(x), x)$. This profile

loss is the reconstruction loss after locally refitting the network at each pose. The two pose problems being compared are

$$\min_{x\in\mathcal{V}} F_1(R(x)) \quad \text{and} \quad \min_{x\in\mathcal{V}} \Phi(x), \qquad F_1(R) = \sum_{(i,j)\in E} w_{ij} D_1(R_i S_i, R_j S_j).$$

The first has a stationary point at $x = 0$. We analyze the second through its quadratic model at that point. The positive parameter curvature assumed below guarantees such a branch when the residual is nonzero.

Stack the residuals into $e \in \mathbb{R}^{nd}$. Evaluate all derivatives below at $(\theta_0, 0)$ unless stated otherwise. The Jacobians $J_x = \partial_x e$ and $J_\theta = \partial_\theta e$ map pose increments in $\mathcal{V}$ and parameter increments in $\mathbb{R}^p$, respectively, to residual changes. Their transposes are Euclidean adjoints. Write $a = J_x^\top e$ for the pose gradient. The joint Hessian and quadratic model are

$$\begin{aligned} \mathcal{H} &= \nabla^2_{(x,\theta)} \mathcal{J} = [J_x, J_\theta]^\top [J_x, J_\theta] + \sum_\alpha e_\alpha \nabla^2_{(x,\theta)} e_\alpha, \\ Q(x,h) &= \mathcal{J}(\theta_0, 0) + a^\top x + \tfrac{1}{2} \begin{pmatrix} x \\ h \end{pmatrix}^\top \mathcal{H} \begin{pmatrix} x \\ h \end{pmatrix}. \end{aligned} \tag{17}$$

Here $h \in \mathbb{R}^p$ is the shared increment in $\theta = \theta_0 + h$, and α indexes scalar residual entries. Parameter stationarity removes the linear term in h. The Hessian includes the residual-weighted second derivatives of the network and rotation maps.

E.2 The gradient: reconstruction errors that drive rotation

Parameter stationarity gives $J_\theta^\top e = 0$. Let $P_\theta = J_\theta J_\theta^\dagger$ project onto the residual changes attainable by adjusting the network, and set $J_\perp = (I - P_\theta) J_x$. Since $P_\theta e = 0$,

$$\boxed{a = J_x^\top e = J_\perp^\top e.} \tag{18}$$

$J_\perp$ contains the rotation responses that parameter changes cannot reproduce to first order. Their correlation with the residuals gives the pose gradient. The derivative J_x accounts for changes in both the rotated target and the network output evaluated at the rotated input. Before restricting to $\mathcal{V}$, its ith block acts on a direction $v = (v_1, \ldots, v_n)$ as

$$(J_x v)_i = \big(I - \mathrm{D}T_{\theta_0}(Y_i)\big)\, \mathrm{D}Y_i(0) v_i.$$

Here D denotes differentiation with respect to a map's argument. If $j_\alpha^\top$ denotes a row of $J_\perp$, then $a = \sum_\alpha e_\alpha j_\alpha$. Contributions to each pose coordinate are summed before taking a norm and can cancel. Residuals orthogonal to the effective rotation responses contribute zero pose gradient. The bound

$$\|a\| = \|J_\perp^\top e\| \le \|J_\perp\|_{\mathrm{op}} \|e\|$$

shows that accurate reconstruction gives a small numerator when the effective rotation responses remain bounded; here $\|\cdot\|_{\mathrm{op}}$ denotes the operator norm. Weak correlation between the residuals and these responses can further reduce the pose gradient. Whenever the local profile is differentiable, $\nabla\Phi(0) = a$.

If the network reconstructs the geometric initialization exactly, then $e = 0$, $a = 0$, and every residual-weighted Hessian term vanishes. Hence

$$Q(x,h) = \tfrac{1}{2}\|J_x x + J_\theta h\|^2 \ge 0 = Q(0,0).$$

The initialization attains the minimum value of zero for both the quadratic model and the original reconstruction objective. Zero pose displacement is therefore an optimal response.

E.3 The Hessian: complementary constraints from shared parameters

For $e \neq 0$, assume the parameter Hessian $K = \mathcal{H}_{\theta\theta}$ is positive definite in the chosen local parameter coordinates. Directions corresponding to parameter symmetries are removed before imposing this assumption. Minimizing Q over h defines $Q_{\mathrm{eff}}(x) = \min_h Q(x,h)$ and gives

$$h_*(x) = -K^{-1}\mathcal{H}_{\theta x} x, \qquad Q_{\mathrm{eff}}(x) = \mathcal{J}(\theta_0, 0) + a^\top x + \tfrac{1}{2} x^\top H_{\mathrm{AE}} x,$$

where

$$\boxed{\begin{aligned} H_{\mathrm{AE}} &= \mathcal{H}_{xx} - \mathcal{H}_{x\theta} K^{-1} \mathcal{H}_{\theta x}, \\ v^\top H_{\mathrm{AE}} v &= \min_h \begin{pmatrix} v \\ h \end{pmatrix}^\top \mathcal{H} \begin{pmatrix} v \\ h \end{pmatrix}. \end{aligned}} \tag{19}$$

Differentiating the parameter stationarity equation $\nabla_\theta \mathcal{J}(\theta_*(x), x) = 0$ gives $\mathrm{D}\theta_*(0) = -K^{-1}\mathcal{H}_{\theta x}$. Parameter stationarity also gives $\nabla\Phi(x) = \partial_x \mathcal{J}(\theta_*(x), x)$; differentiating this identity yields $\nabla^2\Phi(0) = H_{\mathrm{AE}}$. This Schur complement is the curvature remaining after the optimal shared parameter response.

Each shape contributes to this curvature through its reconstruction loss. Before restricting the pose variables, write its full Hessian as

$$\mathcal{H}_i = \nabla^2_{(x_i,\theta)} \mathcal{J}_i(\theta_0, 0) = \begin{pmatrix} C_i & W_i^\top \\ W_i & K_i \end{pmatrix}, \qquad c_i(v_i, h) = \begin{pmatrix} v_i \\ h \end{pmatrix}^\top \mathcal{H}_i \begin{pmatrix} v_i \\ h \end{pmatrix}.$$

Here $C_i \in \mathbb{R}^{3\times 3}$ is the pose block, $K_i \in \mathbb{R}^{p\times p}$ the parameter block, and $W_i \in \mathbb{R}^{p\times 3}$ their cross derivative; $K = \sum_i K_i$. The quadratic form $c_i(v_i, h)$ is the curvature contributed by shape i along pose and parameter increments (v_i, h). By additivity, the minimum in Eq. (19) equals $\min_h \sum_i c_i(v_i, h)$.

First minimize each shape's curvature contribution independently. For a prescribed pose direction v_i, define $m_i(v_i) = \min_h c_i(v_i, h)$. Assume the minimum is attained and nonnegative for every $v_i \in \mathbb{R}^3$ and every shape, equivalently $\mathcal{H}_i \succeq 0$. The parameter response with minimum Euclidean norm is $h_i(v_i) = -K_i^\dagger W_i v_i$. Completing the square gives

$$\begin{aligned} m_i(v_i) &= v_i^\top (C_i - W_i^\top K_i^\dagger W_i) v_i, \\ c_i(v_i, h) &= m_i(v_i) + \|h - h_i(v_i)\|^2_{K_i}, \qquad \|u\|^2_{K_i} = u^\top K_i u. \end{aligned} \tag{20}$$

The quantity $\|u\|_{K_i}$ is a seminorm when K_i is singular. All minimizing responses then form the set $h_i(v_i) + \ker K_i$, where $\ker K_i$ is the null space of K_i. Positive semidefiniteness of $\mathcal{H}_i$ gives $K_i \succeq 0$ and $W_i v_i \in \mathrm{range}(K_i)$, which justify the square completion in this case. In particular, an independent parameter update may remove all curvature in a pose direction, giving $m_i(v_i) = 0$.

A shared AE uses the same parameter increment for all shapes. Substituting Eq. (20) into Eq. (19) gives

$$\boxed{v^\top H_{\mathrm{AE}} v = \sum_i m_i(v_i) + \mathcal{C}(v), \qquad \mathcal{C}(v) = \min_h \sum_i \|h - h_i(v_i)\|^2_{K_i}.} \tag{21}$$

Different shapes may require different parameter responses to compensate for their rotations. The term $\mathcal{C}(v)$ measures the discrepancy between a shared response and the individual responses $h_i(v_i)$, weighted by K_i. This term can be positive even when every $m_i(v_i)$ is zero. Shared adaptation can therefore retain pose curvature that independent per-shape adaptation would remove. This contribution depends on incompatible parameter responses to rotation; nonzero reconstruction error alone does not guarantee a positive minimum pose curvature.

Assume the following implication holds for every $v \in \mathcal{V}$ and $h \in \mathbb{R}^p$:

$$\boxed{\left.\begin{aligned} m_i(v_i) &= 0 \quad \text{for every } i, \\ \|h - h_i(v_i)\|_{K_i} &= 0 \quad \text{for every } i \end{aligned}\right\} \implies v = 0.} \tag{22}$$

This identifiability condition states that every nonzero relative pose direction either has $\sum_i m_i(v_i) > 0$ or admits no shared parameter response attaining all independent minima.

E.4 Node displacement

Under these conditions, define

$$\boxed{c_{\mathrm{AE}} = \min_{\substack{v\in\mathcal{V} \\ \|v\|=1}} \left[\sum_i m_i(v_i) + \mathcal{C}(v) \right] > 0.} \tag{23}$$

To see why $c_{\mathrm{AE}} > 0$, suppose the sum in brackets were zero for a unit vector v. Since $K \succ 0$, the minimum defining $\mathcal{C}(v)$ is attained at some shared increment h. Every nonnegative term must then vanish, so Eq. (22) gives $v = 0$, a contradiction. The sum equals the continuous quadratic form $v^\top H_{\mathrm{AE}} v$; compactness of the unit sphere therefore gives a strictly positive minimum. Hence $H_{\mathrm{AE}} \succeq c_{\mathrm{AE}} I_{\mathcal{V}}$, where $I_{\mathcal{V}}$ is the identity on $\mathcal{V}$. The quadratic model Q_{eff} has the unique pose minimizer

$$\boxed{x_{\mathrm{AE},q} = -H_{\mathrm{AE}}^{-1} a, \qquad \|x_{\mathrm{AE},q}\| \le \frac{\|a\|}{c_{\mathrm{AE}}} = \frac{\|J_\perp^\top e\|}{c_{\mathrm{AE}}}.} \tag{24}$$

Using $H_{\mathrm{AE}} x_{\mathrm{AE},q} = -a$ and Cauchy–Schwarz,

$$c_{\mathrm{AE}} \|x_{\mathrm{AE},q}\|^2 \le x_{\mathrm{AE},q}^\top H_{\mathrm{AE}} x_{\mathrm{AE},q} = -x_{\mathrm{AE},q}^\top a \le \|x_{\mathrm{AE},q}\| \, \|a\|.$$

Dividing by $\|x_{\mathrm{AE},q}\|$ proves the bound when it is nonzero; the zero case is immediate. The predicted poses are $R_i = \exp([x_{\mathrm{AE},q,i}]_\times) R_i^1$. When $\|x_{\mathrm{AE},q,i}\| < \pi$, this norm equals the angular displacement of shape i from the geometric alignment under D_1.

The numerator measures the residual forcing on pose: accurate reconstruction and weak correlation with effective rotation responses can make it small. The denominator is the minimum curvature remaining after shared parameter adjustment. Under the stated identifiability condition, individual shape curvature and incompatible parameter responses jointly constrain every relative pose direction. Small residual forcing relative to this remaining curvature gives a small predicted change from the geometric initialization.

F Details of Preliminaries

This section introduces the details of the tools we summarized in Sec. 4.1.

F.1 Spherical Harmonics and Pose Estimation

We adopt the approach of He et al. (2025b) to estimate relative rotations between objects. Specifically, we represent each object o by spherical harmonic (SH) coefficients $F \in \mathbb{R}^{l_{\max}^2 \times n_r}$. We first construct a volumetric function $f : \mathbb{R}^3 \to \mathbb{R}$, defined as the signed distance function for a 3D shape or the spatial electron density for a protein. We then restrict f to n_r concentric spheres centered at the barycenter of o, with uniformly spaced radii. The columns of F contain the SH coefficients of these spherical functions, retaining $l_{\max}$ frequency bands. Throughout this paper, we use $l_{\max} = 16$ and $n_r = 8$, with a common maximum radius chosen to enclose every input object.

Let $\boldsymbol{f}_{kl} \in \mathbb{R}^{2k-1}$ denote the coefficient vector of the l-th spherical function in the k-th frequency band, corresponding to SH degree $k-1$. Its Euclidean norm $|\boldsymbol{f}_{kl}|$ is invariant under rotations of f. Concatenating these norms across all frequency bands and radii yields a rotation-invariant descriptor $d \in \mathbb{R}^{l_{\max} n_r}$. This descriptor supports effective shape comparison and is used to identify the nearest neighbors of each object.

Let $F(R)$ denote the SH coefficients of the volumetric function f after applying a rotation $R \in \mathrm{SO}(3)$. Given the SH representations F_i and F_j of objects o_i and o_j, we estimate the relative rotation that aligns o_i with o_j by solving

$$\min_{R \in \mathrm{SO}(3)} \quad |F_i(R) - F_j|_{\mathcal{F}}^2, \tag{25}$$

where $|\cdot|_{\mathcal{F}}$ denotes the Frobenius norm.

The method of He et al. (2025b) solves Eq. (25) using continuous optimization. Because the objective is non-convex, the solution depends on the initial rotation. To obtain promising initializations, the method exploits the Wigner-D representation of rotations to efficiently evaluate the objective on a grid of samples in $\mathrm{SO}(3)$. Local minima on this grid serve as initial rotations, each of which is refined through continuous optimization. By retaining multiple local minima, the method captures alternative alignments, making it particularly suitable for objects with self-symmetries.

F.2 $SO(3)$ Projection, Synchronization, and Clustering

$SO(3)$ **projection.** The projection of a matrix $A \in \mathbb{R}^{3\times 3}$ in SO(3) is given by

$$\mathcal{P}_{SO(3)}(A) = \operatorname*{argmin}_{R \in SO(3)} \|A - R\|_{\mathcal{F}}^2 = U \mathrm{diag}(1, 1, s) V^T \tag{26}$$

where $s = \text{sign}(\det(UV^T))$ and $U\Sigma V^T$ is the SVD of A.

$SO(3)$ **synchronization.** Our approach will utilize two types of rotation synchronization techniques, both of which estimate the absolute poses of n nodes $\mathcal{O} = \{o_1, \cdots, o_n\}$ from the relative poses associated with the edges of a graph $\mathcal{G} = \{\mathcal{O}, \mathcal{E}\}$. The first is Laplacian rotation synchronization Huang et al. (2019), which takes as input a single relative rotation R_{ij} associated with each edge $(i, j) \in \mathcal{E}$ and outputs the pose of each object as

$$R_i = \hat{R}_1^T \hat{R}_i, \qquad \hat{R}_i = \mathcal{P}_{SO(3)}(S_i)^T$$

where $S = (S_1; \cdots; S_n) \in \mathbb{R}^{3n\times 3}$ concatenates the three eigenvectors corresponding to the smallest eigenvalues of the connection Laplacian $L \in \mathbb{R}^{3n\times 3n}$ where

$$L_{ij} = \begin{cases} d_i I_3 & i = j \\ -R_{ij}^T & (i,j) \in \mathcal{E} \\ 0 & \text{otherwise} \end{cases}$$

where d_i is the number of neighbors of o_i. Note that if $\det(S_1) < 0$, we multiply one column of the entire eigenvector matrix S by -1.

The second assumes that each edge is associated with a cost function $\phi_{ij}(R_i, R_j)$ and each node o_i has m_i candidate poses $\mathcal{R}_i = \{R_{i1}, \cdots, R_{im_i}\}$. Let $x_i \in \{1, \cdots, m_i\}$ be the selected pose index of o_i. MRF-based pose synchronization amounts to solving the following optimization problem

$$\min_{\{x_i\}} \quad \sum_{(i,j)\in\mathcal{E}} \phi_{ij}(R_{ix_i}, R_{jx_j}). \tag{27}$$

There are many MRF solvers. We use the OPENGM's Andres et al. (2012); Kappes et al. (2015) implementation of tree-reweighted belief propagation Wainwright et al. (2005), which applies to multi-labels on sparse graphs.

We differentiate between two scenarios when defining the cost function ϕ_{ij}. When o_i and o_j are 3D objects, we define

$$\phi_{ij}(R_{ik}, R_{jl}) = \frac{\sigma_F^2 \|F_i(R_{ik}) - F_j(R_{jl})\|_{\mathcal{F}}^2}{\sigma_F^2 + \|F_i(R_{ik}) - F_j(R_{jl})\|_{\mathcal{F}}^2}. \tag{28}$$

When o_i and o_j represent clusters of objects, where there are m_{ij} candidate poses $\mathcal{R}_{ij} = \{(R_{ij}^1, \cdots, R_{ij}^{m_{ij}})$ between them, we define

$$\phi_{ij}(R_{ik}, R_{jl}) = \min_{R_{ij}^t \in \mathcal{R}_{ij}} c_{ij}^t \|R_{ij}^t - R_{jl}^T R_{ik}\|_{\mathcal{F}}^2 \tag{29}$$

where $c_{ij}^t > 0$ is the cost associated with R_{ij}^t.

The performance of MRF-based rotation synchronization depends on the quality of the candidate poses $\mathcal{R}_i$. Ideally, the size of $\mathcal{R}_i$ should be small, and one of its elements shall approximate the underlying ground-truth. These two properties are not satisfied in prior work Crandall et al. (2013); Sun et al. (2019), which is based on sampling SO(3) and leads to a large $\mathcal{R}_i$, resulting in generally inaccurate samples. As described in the main paper, a key contribution of this paper is an iterative approach that produces accurate samples.

$SO(3)$ **clustering.** Finally, we describe an efficient procedure to cluster a collection of rotation samples $\mathcal{R} = \{R_1, \cdots, R_N\}$, where w_i is the weight of R_i, and compute a mean rotation from each cluster. This procedure uses three simple hyper-parameters, i.e., a radius parameter δ of each cluster, the maximum number of clusters $K_{\max}$, and k nearest neighbors used in computing sample densities. Our key idea is to compute seed rotations $\mathcal{S} = \{R_{s_1}, \cdots, R_{s_K}\}$ and generate a cluster for each seed as

$$C_{s_i} = \{j | i = \underset{1\le l\le K}{\text{argmin}} \|R_{s_l} - R_j\|_{\mathcal{F}}\}.$$

The seeds are expected to be located in regions of dense samples. Therefore, we compute the density of each R_i as

$$c(R_i) = \frac{1}{k} \sum_{j\in\mathcal{N}_i} \|R_i - R_j\|_{\mathcal{F}} \tag{30}$$

where $\mathcal{N}_i$ is the k-nearest neighbors of R_i. We then iteratively pick seeds s_l that have the smallest $c(R_{s_l})$. s_1 is given by the smallest value of $c(R_i)$. Suppose that we have picked l

seeds $R_{s_1}, \cdots, R_{s_l}$. s_{l+1} is given by the sample that has the smallest $c(R_i)$ and whose distance $\min_{1 \le j \le l} \|R_i - R_{s_j}\|_{\mathcal{F}} \ge \delta$.

For each cluster $C \subset \mathcal{R}$, we compute its mean rotation as $R_C = \mathcal{P}_{SO(3)}(\sum_{R_i \in C} w_i R_i)$ where w_i is the weight of R_i. We also set an upper limit $K_{\max}$ on the number of clusters by sorting the clusters based on the sum of rotation weights of each cluster and keeping the top $K_{\max}$ clusters.

In this paper, we set $\delta = 0.2$ in all of our experiments.

G Details of the Object Cluster Generation Stage

The overlapping clusters are constructed in three stages. In the first stage, we perform farthest-point sampling in the descriptor space to select $\overline{K}$ seed objects. For each seed object o_{i_1}, we retrieve its $\overline{n}$ nearest neighbors in the descriptor space to form an initial cluster $\mathcal{C}_i^{in}$. We choose $\overline{K}$ and $\overline{n}$ sufficiently large so that the initial clusters collectively cover all input objects. Empirically, we set

$$\overline{K} = \overline{n} = \sqrt{Cn},$$

where n is the number of input objects and $C = 30$.

The second step selects a subset of clusters so that the coverage constraint remains valid. We do this in an iterative manner. We start with the first $\mathcal{C}_1 = \mathcal{C}_1^{in}$. Suppose we have decided $\mathcal{C}_1, \cdots, \mathcal{C}_k$. We define the distance between $\mathcal{C}_j^{in}$ and the existing clusters as

$$d(\mathcal{C}_j^{in}, \{\mathcal{C}_i, 1 \le i \le k\}) = \frac{1}{m} \sum_{o_{j_l} \in \mathcal{C}_j} \min_{1 \le i \le k} \|\boldsymbol{d}_{j_l} - \boldsymbol{d}_{i_1}\|.$$

We also set $\mathcal{C}_{k+1}$ as the $\mathcal{C}_j^{in}$ that has the maximum number of uncovered objects and then maximize the distance $d(\mathcal{C}_j^{in}, \{\mathcal{C}_i, 1 \le i \le k\})$ to break the tie.

In the third stage, we prune redundant object–cluster memberships to reduce the cluster sizes while preserving coverage. We sort all object–cluster relations according to the rank of each object within its cluster and process them from lower-priority relations to higher-priority ones. We remove a relation whenever doing so preserves the coverage constraint. Once removing a relation from a cluster would violate coverage, we fix that cluster and remove its remaining relations from the pruning queue.

We further refine the clusters by recomputing each cluster center as the center of its minimum enclosing sphere in the descriptor space. The objects within each cluster are then reordered according to their distances to the updated center, after which the pruning procedure is applied again. We alternate between recentering, reordering, and pruning until the cluster configuration stabilizes.

Finally, we require the cluster-overlap graph $\overline{G}$ to be sufficiently well connected to support robust inter-cluster pose propagation and aggregation. In particular, we monitor its second-smallest Laplacian eigenvalue $\lambda_2(\overline{G})$. Across all three datasets considered in our experiments, we find that $\lambda_2(\overline{G})$ is typically larger than a prescribed threshold $\lambda_{\min}$. If $\lambda_2(\overline{G}) < \lambda_{\min}$ after stage I, we add additional clusters until this condition is satisfied. During cluster pruning, an object–cluster relation is removed only if the resulting cluster-overlap graph continues to satisfy

$$\lambda_2(\overline{G}) \ge \lambda_{\min}.$$

Thus, the pruning procedure reduces redundancy while explicitly preserving both object coverage and the connectivity of the cluster-overlap graph. The resulting cluster-overlap graphs in our experiments are also non-bipartite.

H Convergence Analysis

We analyze the single-seed, single-candidate, equal-weight specialization of propagation and aggregation. This simplified setting has all the ingredients of the general setting, e.g., propagation, aggregation, and projection, and the convergence analysis provides insights on the behavior of the general setting.

There is one fixed relative-rotation measurement per edge, at most one candidate per node, and all incoming votes receive equal weight at every iteration. We compare configurations after optimal common alignment: finite propagation enters a small neighborhood of synchronization, the update contracts distances within this neighborhood, and reciprocity of the measurements ensures convergence of the actual rotations.

Let $\mathcal{G} = (\mathcal{O}, \mathcal{E})$ be a fixed undirected, connected, non-bipartite graph on n objects, with neighbor sets $\mathcal{N}_i$ and degrees d_i. The node variable R_i denotes the canonicalizing rotation used in the main text, and R_{ij} maps object i to object j. Thus $R_j^{gt} R_{ij}^{gt} = R_i^{gt}$, and a vote from j to i is $R_j^{(t)} R_{ij}$, as in Eq. (8). Set

$$R_{ij}^{gt} = {R_j^{gt}}^T R_i^{gt}, \qquad R_{ij} = \exp(-[\boldsymbol{c}_{ij}]_\times) R_{ij}^{gt}, \qquad \varepsilon = \max_{\{i,j\}\in\mathcal{E}} \|\boldsymbol{c}_{ij}\|_2. \tag{31}$$

Here $[\boldsymbol{c}]_\times \boldsymbol{v} = \boldsymbol{c} \times \boldsymbol{v}$. Each undirected edge is measured once in one orientation; the reverse measurement is $R_{ji} = R_{ij}^T$.

Choose any $R_1^{(0)} \in \mathrm{SO}(3)$ as the sole initial candidate. Let $\mathcal{V}_t$ contain the nodes with a candidate at round t. The synchronous update is

$$\begin{aligned} \mathcal{V}_0 = \{1\}, \qquad \mathcal{V}_{t+1} = \{i : \mathcal{N}_i \cap \mathcal{V}_t \neq \emptyset\}, \\ R_i^{(t+1)} = \mathcal{P}_{\mathrm{SO}(3)} \left(\sum_{j \in \mathcal{N}_i \cap \mathcal{V}_t} R_j^{(t)} R_{ij} \right), \qquad i \in \mathcal{V}_{t+1}. \end{aligned} \tag{32}$$

Nodes receiving no votes have no candidate in that round. All nodes, including node 1, follow this rule.

Definition 1 *Let T be the first round with $\mathcal{V}_T = \{1, \dots, n\}$. From this round onward, every node has a candidate and the update uses all neighbors. Thus, T marks the point from which convergence can be analyzed on the entire graph.*

Let $P_{ij} = 1/d_i$ for $j \in \mathcal{N}_i$ and zero otherwise, and define

$$\pi_i = \frac{d_i}{\sum_j d_j}, \qquad \pi_{\min} = \min_i \pi_i, \qquad q_0 = \max_{k \geq 2} |\lambda_k(P)| < 1, \quad \lambda_1(P) = 1.$$

For rotation configurations $X, Y \in \mathrm{SO}(3)^n$, use the weighted Frobenius norm (also for their differences) and aligned distance

$$\|X\|_\pi^2 = \sum_i \pi_i \|X_i\|_F^2, \qquad d(X, Y) = \min_{Q \in \mathrm{SO}(3)} \|X - QY\|_\pi, \qquad \mathbf{I} = (I_3, \dots, I_3).$$

The distance identifies configurations differing by a common left rotation. Alignment is used only to compare iterates; it adds no operation to Eq. (32).

Theorem 1 (Single-seed convergence) *For every $q \in (q_0, 1)$, define*

$$\rho = \frac{(q - q_0)\sqrt{\pi_{\min}}}{8}, \qquad \varepsilon_0 = \frac{\rho}{2\sqrt{2}} \min\left\{ \frac{1}{T}, 1 - q \right\}. \tag{33}$$

If $\varepsilon \leq \varepsilon_0$, all nonempty updates are uniquely defined and the iterates converge to a fixed configuration $\overline{R}$. For a constant $C > 0$ depending only on the graph and q, and for every $t \geq T$,

$$\begin{aligned} \max_i \|R_i^{(t)} - \overline{R}_i\|_F \leq C\varepsilon q^{t-T}, \\ \max_i \|{R_1^{(t)}}^T R_i^{(t)} - R_i^\star\|_F \leq 2C\varepsilon q^{t-T}, \qquad R_i^\star = \overline{R}_1^T \overline{R}_i, \quad R_1^\star = I_3. \end{aligned} \tag{34}$$

These bounds hold for any seed rotation. Equation (33) is a conservative sufficient noise condition.

The second bound follows from convergence of the individual rotations. The relative rotation $(R_1^{(t)})^T R_i^{(t)}$ is invariant under common left multiplication.

Proof. 1. Finite propagation. Remove the ground-truth and seed rotations by setting

$$Q_{\mathrm{s}} = R_1^{(0)} R_1^{gt\,T}, \qquad \widetilde{R}_i^{(t)} = Q_{\mathrm{s}}^T R_i^{(t)} R_i^{gt\,T}, \qquad \widetilde{R}_{ij} = R_j^{gt} R_{ij} R_i^{gt\,T}.$$

Rotation projection commutes with left and right rotations, so the transformed votes are $\widetilde{R}_j^{(t)} \widetilde{R}_{ij}$ and $\widetilde{R}_1^{(0)} = I_3$. Moreover,

$$\widetilde{R}_{ij} = \exp(-[R_j^{gt} \boldsymbol{c}_{ij}]_\times), \qquad \widetilde{R}_{ji} = \widetilde{R}_{ij}^T, \qquad \|\widetilde{R}_{ij} - I_3\|_2 \le \varepsilon.$$

The transformed iteration is independent of the seed. A node has a candidate at round t exactly when a length-t walk reaches it from node 1. An odd-cycle detour changes a walk's parity, and an edge followed by its reverse adds two steps. Connectivity therefore gives all sufficiently large walk lengths to every node, so T is finite. Once every node has a candidate, this remains true in subsequent rounds.

Let $\theta(R) \in [0, \pi]$ denote the rotation angle. Projection preserves an angular ball of radius $b < \pi/2$: if $\theta(Q_k) \le b$, $w_k \ge 0$, and $\sum_k w_k = 1$, then the projection of $M = \sum_k w_k Q_k$ is unique and has angle at most b. Indeed, write $M = S + [\boldsymbol{s}]_\times$. Rodrigues' formula gives $S \succeq \cos(b) I_3$ and $\|\boldsymbol{s}\|_2 \le \sin b$. The path $S + t[\boldsymbol{s}]_\times$ is nonsingular, so $\det M > 0$. Along a unit rotation axis $\boldsymbol{u}$, the projection objective is a constant plus $A\cos\theta + B\sin\theta$, where

$$A = \operatorname{tr} S - \boldsymbol{u}^T S \boldsymbol{u} \ge 2\cos b, \qquad |B| = |2\boldsymbol{u}^T \boldsymbol{s}| \le 2\sin b.$$

Its maximizing signed angle has magnitude at most $\arctan(|B|/A) \le b$, proving the claim.

Normalize each nonempty vote sum by its number of votes, which leaves the projection unchanged. The angular triangle inequality and induction give

$$\varepsilon \le \frac{\pi}{4T} \quad \Longrightarrow \quad \max_{i \in \mathcal{V}_t} \theta(\widetilde{R}_i^{(t)}) \le t\varepsilon, \qquad 0 \le t \le T. \tag{35}$$

All these projections are unique. Since $\|R - I_3\|_F = 2\sqrt{2}\sin(\theta(R)/2)$, this yields the explicit propagation bound

$$\max_i \|\widetilde{R}_i^{(T)} - I_3\|_F \le \sqrt{2}\, T\varepsilon. \tag{36}$$

In particular, this error is at most a prescribed $\delta > 0$ if $\varepsilon \le \min\{\pi/(4T), \delta/(\sqrt{2}T)\}$. Equation (33) ensures $T\varepsilon \le \rho/(2\sqrt{2}) < \pi/4$ and $d(\widetilde{R}^{(T)}, \mathbf{I}) \le \rho/2$.

2. Local contraction. For $t \ge T$, the transformed update is

$$\mathcal{F}_i(X) = \mathcal{P}_{\mathrm{SO}(3)} \left(\sum_{j \in \mathcal{N}_i} P_{ij} X_j \widetilde{R}_{ij} \right).$$

Dividing the vote sum by d_i does not change its projection. The map satisfies $\mathcal{F}(QX) = Q\mathcal{F}(X)$. Work in $\mathcal{B}_\rho = \{X \in \mathrm{SO}(3)^n : d(X, \mathbf{I}) \le \rho\}$. After common alignment, graph averaging contributes q_0, the residual mean and measurement errors add $2\rho + \varepsilon$, and projection amplifies the discrepancy by at most s^{-1}. We establish these bounds next.

We use one polar-projection inequality. If $M = S_M U$ and $N = S_N V$ are polar decompositions with $U, V \in \mathrm{SO}(3)$ and $S_M, S_N \succeq sI_3$, then $\langle M - N, U - V \rangle_F \ge s\|U - V\|_F^2$. This follows by expanding the inner product and using $I_3 - (UV^T + VU^T)/2 \succeq 0$. Cauchy–Schwarz gives

$$\|\mathcal{P}_{\mathrm{SO}(3)}(M) - \mathcal{P}_{\mathrm{SO}(3)}(N)\|_F \le s^{-1}\|M - N\|_F. \tag{37}$$

For $X \in \mathcal{B}_\rho$, align X to $\mathbf{I}$ so that $\|X - \mathbf{I}\|_\pi \le \rho$. Put $a = \rho/\sqrt{\pi_{\min}}$. Then $\max_i \|X_i - I_3\|_2 \le a$, and every normalized vote matrix is within $a + \varepsilon$ of I_3 in operator norm. It has positive determinant and minimum singular value at least $s = 1 - a - \varepsilon > 3/4$. These properties are unchanged by common left rotation.

Take $X, Y \in \mathcal{B}_\rho$. First align X to $\mathbf{I}$, then align Y optimally to X. Set $D_i = X_i - Y_i$, $d_0 = \|D\|_\pi = d(X, Y) \le 2\rho$, and $\overline{D} = \sum_i \pi_i D_i$. Optimal alignment makes $\sum_i \pi_i X_i Y_i^T$ symmetric. Using orthogonality gives

$$\overline{D} = \sum_i \pi_i D_i (I_3 - X_i^T) + \frac{1}{2} \sum_i \pi_i D_i D_i^T,$$

$$\|\overline{D}\|_F \le \rho d_0 + \frac{1}{2} d_0^2 \le 2\rho d_0.$$

The identities $\pi^T P = \pi^T$ and $\pi_i P_{ij} = \pi_j P_{ji}$ make P self-adjoint in the weighted inner product. Its norm on zero-mean configurations is q_0: connectivity removes additional eigenvalues 1, and non-bipartiteness excludes -1. Splitting D into its mean and zero-mean parts, and applying weighted Jensen's inequality to the measurement perturbation, gives

$$\|PD\|_\pi \le (q_0 + 2\rho)d_0,$$
$$\left\| \left(\sum_j P_{ij} D_j (\widetilde{R}_{ij} - I_3) \right)_i \right\|_\pi \le \varepsilon d_0.$$

Together with Eq. (37), these imply the direct contraction estimate

$$d(\mathcal{F}(X), \mathcal{F}(Y)) \le \frac{q_0 + 2\rho + \varepsilon}{1 - \rho/\sqrt{\pi_{\min}} - \varepsilon}\, d(X, Y) \le q\, d(X, Y). \tag{38}$$

To verify the last inequality, put $\Delta = q - q_0$. Equation (33) gives $a = \Delta/8$, $\rho \le a$, and $\varepsilon \le a$, so $2\rho + qa + (1+q)\varepsilon \le 5\Delta/8 < \Delta$.

The nearest-point property of projection gives $d(\mathcal{F}(\mathbf{I}), \mathbf{I}) \le 2\sqrt{2}\varepsilon$, since each input mean is within $\sqrt{2}\varepsilon$ of I_3 in Frobenius norm. Hence, for $X \in \mathcal{B}_\rho$,

$$d(\mathcal{F}(X), \mathbf{I}) \le q\, d(X, \mathbf{I}) + 2\sqrt{2}\varepsilon \le \rho.$$

Together with Step 1, this shows that all iterates from time T onward remain in $\mathcal{B}_\rho$.

3. Summable updates. We show that each actual update is controlled by its aligned change:

$$\|\mathcal{F}(X) - X\|_\pi \le 3 d(\mathcal{F}(X), X), \qquad X \in \mathcal{B}_\rho. \tag{39}$$

By equivariance, assume $\|X - \mathbf{I}\|_\pi \le \rho$. Choose Q attaining $e = d(\mathcal{F}(X), X)$, and set $E_i = \mathcal{F}_i(X) - QX_i$. The polar decompositions satisfy

$$M_i = \sum_j P_{ij} X_j \widetilde{R}_{ij} = S_i \mathcal{F}_i(X), \qquad sI_3 \preceq S_i \preceq I_3,$$

since each M_i is a convex combination of rotations. Set $B = \sum_i \pi_i M_i X_i^T$ and $S = \sum_i \pi_i S_i$. Reciprocity and $\pi_i P_{ij} = \pi_j P_{ji}$ make B symmetric by pairing each edge term with its transpose. The bounds of Step 2 give

$$\|B - I_3\|_2 \le 2a + \varepsilon < \tfrac{1}{2},$$
$$B - SQ = \sum_i \pi_i S_i E_i X_i^T, \qquad \|B - SQ\|_F \le \sum_i \pi_i \|E_i\|_F \le e.$$

Thus $B \succ \frac{1}{2} I_3$ and $S \succeq sI_3$. The polar rotation factors of B and SQ are I_3 and Q. Equation (37) with lower bound $1/2$ gives $\|Q - I_3\|_F \le 2e$, so $\|\mathcal{F}(X) - X\|_\pi \le e + \|Q - I_3\|_F \le 3e$.

Write $X^{(t)} = \widetilde{R}^{(t)}$ and $e_t = d(X^{(t+1)}, X^{(t)})$. Contraction and the triangle inequality give, for $t \ge T$,

$$e_{t+1} \le q e_t,$$
$$e_T \le (1+q) d(X^{(T)}, \mathbf{I}) + 2\sqrt{2}\varepsilon \le 2\sqrt{2}(T+1)\varepsilon.$$

By Eq. (39), $\|X^{(t+1)} - X^{(t)}\|_\pi \le 3 e_T q^{t-T}$. Summing this geometric series proves convergence of the actual iterates to $X^{(\infty)}$ and

$$\|X^{(t)} - X^{(\infty)}\|_\pi \le \frac{3e_T}{1-q}\, q^{t-T}.$$

The limit is fixed by continuity. Returning to $\overline{R}_i = Q_{\mathrm{s}} X_i^{(\infty)} R_i^{gt}$ and converting the weighted norm to a maximum-node norm proves Eq. (34), with $C = 6\sqrt{2}(T+1)/((1-q)\sqrt{\pi_{\min}})$; its second bound follows by the triangle inequality.

Applying the invariant-neighborhood estimate to the fixed limit gives $d(X^{(\infty)}, \mathbf{I}) \le q\, d(X^{(\infty)}, \mathbf{I}) + 2\sqrt{2}\varepsilon$. Hence, under $\varepsilon \le \varepsilon_0$,

$$d(\overline{R}, R^{gt}) = d(X^{(\infty)}, \mathbf{I}) \le \frac{2\sqrt{2}\varepsilon}{1-q},$$
$$\min_{Q \in \mathrm{SO}(3)} \max_i \|\overline{R}_i - Q R_i^{gt}\|_F \le \frac{2\sqrt{2}\varepsilon}{(1-q)\sqrt{\pi_{\min}}}. \tag{40}$$

The second bound uses a Q attaining the weighted distance. Choosing $q = (1 + q_0)/2$ in Eq. (33) yields the explicit spectral bound $d(\overline{R}, R^{gt}) \leq 4\sqrt{2}\varepsilon/(1 - q_0)$. □

For Gaussian errors $\boldsymbol{c}_{ij} \sim \mathcal{N}(0, \sigma^2 I_3)$ sampled once independently per undirected edge, a coordinate-wise union bound ensures the theorem holds with probability at least $1 - \delta$ if $\sigma \leq \varepsilon_0/\sqrt{6\log(6|\mathcal{E}|/\delta)}$, for $0 < \delta < 1$. At zero noise, the fixed configuration is reached by time T.