Combinative Matching for Geometric Shape Assembly

Nahyuk Lee[1*] Juhong Min[1,2*] Junhong Lee[1] Chunghyun Park[1] Minsu Cho[1,3]
[1]POSTECH [2]Samsung Research America [3]RLWRLD
https://nahyuklee.github.io/cmnet

Abstract

*This paper introduces a new shape-matching methodology, **combinative matching**, to combine interlocking parts for geometric shape assembly. Previous methods for geometric assembly typically rely on aligning parts by finding identical surfaces between the parts as in conventional shape matching and registration. In contrast, we explicitly model two distinct properties of interlocking shapes: 'identical surface shape' and 'opposite volume occupancy.' Our method thus learns to establish correspondences across regions where their surface shapes appear identical but their volumes occupy the inverted space to each other. To facilitate this process, we also learn to align regions in rotation by estimating their shape orientations via equivariant neural networks. The proposed approach significantly reduces local ambiguities in matching and allows a robust combination of parts in assembly. Experimental results on geometric assembly benchmarks demonstrate the efficacy of our method, consistently outperforming the state of the art.*

1. Introduction

Geometric shape assembly, the task of reconstructing a target object from multiple fractured parts, plays a crucial role in diverse fields such as archaeology [25, 27, 35], medical imaging [16, 23, 50], robotics [9, 37, 49], and industrial manufacturing [2, 3]. Reliable assembly requires not only identifying common interfaces where parts align (*e.g.*, mating surfaces) but also establishing robust feature correspondences that account for how different parts combine with each other. This combinative process involves challenges in analyzing parts such as incomplete semantics, shape ambiguity, variations in orientation, and complexity in matching.

To address the challenges, prior work [5, 14] has predominantly relied on aligning parts by finding identical surfaces between parts as in conventional shape matching and registration. The methods typically extract visual features and maximize similarity for positive matches at interfaces under the assumption of their high visual resemblance. While technically sound, this approach often suffers from local ambiguities, where visually similar shapes from different parts are incorrectly matched, as shown in Fig. 1. This conventional matching on pure shape similarity often results in incorrect matching and pose estimations, as it overlooks intrinsic properties between matching for registration and that for assembly[1]. This naturally raises the question: What do we miss in matching to address the challenges of geometric shape assembly?

Drawing inspiration from construction and civil engineering, where male and female components combine to form stable structures, techniques such as mortise and tenon joints, tongue and groove connections, and dovetail

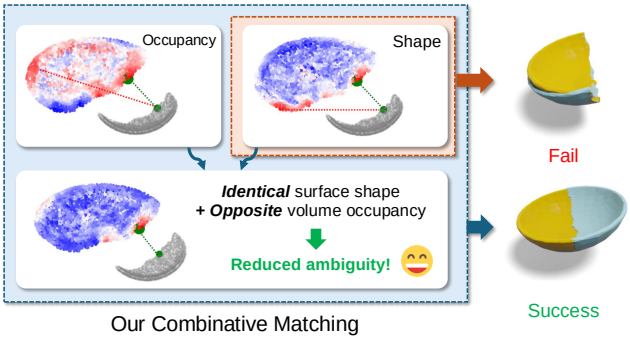

Figure 1. **Combinative matching.** In contrast to conventional approaches to matching solely based on shape similarity, our combinative matching explicitly models two distinct properties of interlocking shapes, 'identical surface shape' and 'opposite volume occupancy,' and learns to establish correspondences across regions where their surface shapes appear identical but their volumes occupy the inverted space to each other. The figure shows the assembly of source (gray) and target (red & blue) parts, with a true match shown by green dots (•) connected by line. The color gradient on target points indicates correlation scores with the green source point, ranging from red (high) to blue (low). Incorporating the volume occupancy (shown in this example), reduces visual ambiguities, achieving accurate assembly.

*Equal contribution

[1]In this manuscript, we use term "matching" to denote *local* pairwise-compatibility test, reserving "assembly" for *global* placement of all parts.

joints [13, 24, 29] demonstrate how stability and precision are achieved not merely through visual resemblance but through combinative properties between parts. Unlike surfaces designed to mirror each other, as in general scene/object alignment or registration tasks [11, 31], the mating parts in geometric assembly [34] are to be combined with each other, requiring attention to their mutual relationship. Let us assume two corresponding points on the mating surfaces of two interlocking parts. The two points share identical surface shapes in their vicinities, but have opposite volume occupancy, *i.e.*, the volume around one point occupies the inverted volume around the other point and vice versa. This observation reveals two distinct properties of interlocking shapes: *identical surface shape* and *opposite volume occupancy*. Reliable shape assembly thus needs to reflect both of the two properties in matching.

To this end, we introduce a new shape-matching methodology for geometric assembly, dubbed *combinative matching*, which learns to match interlocking regions of parts. Unlike conventional matching for registration and assembly [11, 14, 31, 45–47], which commonly relies on shape similarity, combinative matching establishes correspondences across regions where their surface shapes appear identical but their volumes occupy the inverted space to each other. Specifically, we train our model to learn: (1) shape orientations for consistent directional alignment, (2) surface shape descriptors for identical-shape matching, and (3) volume occupancy descriptors for inverted-volume matching. These three objectives jointly help the model reduce local ambiguities, enhance its understanding of interlocking structures, and improve the overall accuracy of assembly. Central to this approach is the use of equivariant and invariant descriptors, allowing both occupancy and shape descriptors to recognize orientation relationships through equivariance, while maintaining robustness to absolute pose through invariance. Experimental results validate that our multi-faceted matching enables a robust, interlocking-aware geometric assembly, addressing the limitations of conventional matching.

2. Related Work

Shape assembly from parts. A common approach to reconstructing a target shape from its parts involves point cloud registration [11, 31, 46], *i.e.*, object or scene alignment tasks, which focus on localizing overlapping interfaces and establishing dense feature correspondences to predict alignment poses. Shape assembly can be viewed as a challenging registration problem under extremely low-overlap conditions, *i.e.*, surface overlap. Existing assembly approaches can be broadly divided into two categories: (1) first category includes methods that rely on direct pose regression using global embeddings for each part [5, 10, 15, 18, 33, 42, 43]. While efficient, these methods often lack fine-grained local detail, leading to inaccuracies. Addressing this limitation, (2) methods such as Jigsaw [21] and PMTR [14] employ dense feature matching to identify reliable correspondences, predicting poses based on the dense matches rather than direct regression, similar to those used in registration approaches [11, 31, 46]. The dense matching methods [11, 14, 21, 31, 46] are built on the assumption that mating interfaces exhibit high visual resemblance, leading to employ training objectives that maximize feature similarity for positive matches. However, unlike general scene alignment, the assembly task requires more than resemblance-based matching alone: Mating interfaces are shaped to interlock rather than mirror each other, demanding a deeper, context-aware understanding of structural complementarity beyond naïve feature similarity.

Civil engineering and construction. A combinative design plays an essential role in creating durable assemblies as demonstrated by civil engineering techniques such as mortise and tenon joints [19, 28, 29, 44], tongue and groove connections [4, 24, 26], dovetail joints [13], rabbet joints [41, 48, 51], and bridle joints [1]. These methods share two key properties of combining parts: *surface resemblance* that ensures that mating parts align smoothly, and *volumetric complementarity* that reflects the design intention for parts to interlock in a structurally sound manner. Although existing shape assembly methods [14, 21] incorporate visual resemblance learning in their objectives, they typically lack the necessary learning to model the volumetric aspects of interlocking parts, which is essential for combinative matching for reliable assembly.

Equivariance and invariance learning. Equivariance and invariance are essential properties in feature learning, especially for tasks involving spatial transformations, where understanding the relative pose relationship between parts is critical. Equivariance ensures transformations applied to the input are reflected in the output, allowing models to retain key orientation information [7, 30**?**] and individual point orientations [12, 22], resulting in structure-aware representations. Invariant descriptors, on the other hand, are widely used for maintaining consistent feature representations regardless of transformations. In geometric matching tasks, many efforts [38, 39, 45, 47] incorporate these descriptors to achieve rotation-invariant matching and alignment, demonstrating strong empirical performance. Inspired by complementary geometric design in civil engineering, our study goes further beyond invariance-based simple visual matching, underscoring that capturing structural complementarity requires equivariant learning to enable models to understand the relative orientations of parts and their interdependency. Our experiments show that leveraging both equivariance and invariance enhances the model's ability to capture essential features for combinative matching.

Complementary matching for assembly. A recent work

by Lu *et al*. [21] presents the concept of a primal-dual descriptor to reflect viewpoint-dependent characteristics for surface matching. Similarly to our motivation, they intend to capture the essence of complementary geometry arising from fracture assembly. However, focusing on the characteristics of a local surface from two different directions, inward and outward, they separate a descriptor into primal and dual ones, and train them to align by intercrossing in matching, *i.e*., encouraging the primal descriptor of one part to resemble the dual descriptor of the other part in matching. Despite a similar motivation, the primal-dual matching method implements the geometric complementarity of mating parts simply by switching two distinct descriptors in matching, and train them to resemble in the primal-dual pair between mating parts; there is no clear distinction between the primal and dual descriptors in terms of their roles and effects. This is clearly different from our approach that distinguishes the descriptor for surface shape, which is to be identical between mating parts, from that for volumetric occupancy, which is to be opposite between mating parts. As will be discussed in our experimental section and supplementary, the primal-dual descriptors fail to capture the interlocking properties of mating parts, and our combinative matching clearly outperforms in matching performance.

3. Proposed Approach

Problem setup. Following previous shape assembly methods [5, 10, 14, 15, 18, 21, 33, 42, 43], our method adopts a self-supervised learning approach: Given a holistic target object, we decompose it into multiple parts, each represented as a point cloud and each undergoing a random rigid transformation. The model takes this set of randomly transformed point clouds as input and predicts a corresponding set of transformation parameters, which are then applied to each part to reconstruct the original target object. Model performance is evaluated by measuring the distances between the ground-truth and predicted assembly configurations, as well as the accuracy of transformation parameters.

3.1. Combinative Matching

In this section, we introduce *Combinative Matching*, a novel approach that addresses the dual requirements of geometric assembly: *identical surface shape* and *opposite volume occupancy*. To ensure consistent assembly despite random transformations, our method first aligns local orientations between surface points, establishing a common reference frame. Within this frame, shape descriptors align to match identical surface shapes, whereas occupancy descriptors are inversely aligned to ensure opposite volume occupancy, enabling parts to interlock properly (Fig. 2).

Orientation alignment. For robust assembly, we require that surface points from different parts share a consistent

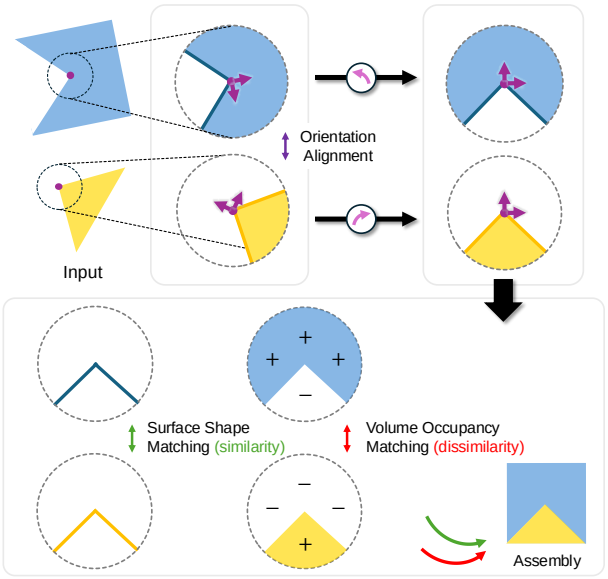

Figure 2. Main concept of our combinative matching.

orientation reference. This alignment ensures that subsequent shape and occupancy features can be compared meaningfully. To this end, we employ an equivariant network f_{d}, which takes a point cloud $\mathbf{P} \in \mathbb{R}^{N\times 3}$ or $\mathbf{Q} \in \mathbb{R}^{M\times 3}$ and predicts orientations $\mathbf{F}_{\mathrm{d}}^{\mathrm{P}} = f_{\mathrm{d}}(\mathbf{P}) \in \mathbb{R}^{N\times 3\times 3}$ and $\mathbf{F}_{\mathrm{d}}^{\mathrm{Q}} = f_{\mathrm{d}}(\mathbf{Q}) \in \mathbb{R}^{M\times 3\times 3}$ with $(\mathbf{F}_{\mathrm{d}}^{\mathrm{Q}})_i, (\mathbf{F}_{\mathrm{d}}^{\mathrm{P}})_i \in \mathrm{SO}(3)$, where N and M are the numbers of sampled points for the respective parts. The training loss for orientation alignment is defined as the difference between aligned orientations:

$$\mathcal{L}_{\mathrm{d}} = \frac{1}{|\mathcal{C}|} \sum_{(i,j)\in\mathcal{C}} \left\| (\mathbf{F}_{\mathrm{d}}^{\mathrm{P}})_i \, \mathbf{R}^{\mathrm{P}} \; - \; (\mathbf{F}_{\mathrm{d}}^{\mathrm{Q}})_j \, \mathbf{R}^{\mathrm{Q}} \right\|_F, \qquad (1)$$

where $\mathcal{C}$ is the set of indices for positive matches, $\mathbf{R}^{\mathrm{P}}$ and $\mathbf{R}^{\mathrm{Q}}$ are the ground-truth rotations of parts $\mathbf{P}$ and $\mathbf{Q}$, and $\|\cdot\|_F$ is the Frobenius norm. By minimizing this loss, the network learns to predict orientations that can be used to extract rigid transformation-invariant occupancy and shape descriptors in the subsequent matching steps, enabling stable assembly regardless of initial part positions.

Surface shape matching. For identical-shape matching, we require shape descriptors that capture *identical* surface characteristics. For this purpose, given the learned surface shape embeddings $\mathbf{F}_{\mathrm{s}}^{\mathrm{P}} = f_{\mathrm{s}}(\mathbf{P}) \in \mathbb{R}^{N\times d_{\mathrm{s}}}$ and $\mathbf{F}_{\mathrm{s}}^{\mathrm{Q}} = f_{\mathrm{s}}(\mathbf{Q}) \in \mathbb{R}^{M\times d_{\mathrm{s}}}$ and a set of all indices for all points on the mating surface $\mathcal{I}$, we employ the standard circle loss [36] without modification as follows:

$$\mathcal{L}_{\mathrm{s}} \propto \mathop{\mathbb{E}}_{i\sim\mathcal{I}} \left[\log\Big(\sum_{j\in\mathcal{E}_{\mathrm{p}}(i)} e^{(d_{ij}^{\mathrm{p}} - \Delta_{\mathrm{p}})^2} \cdot \sum_{k\in\mathcal{E}_{\mathrm{n}}(i)} e^{(\Delta_{\mathrm{n}} - d_{ik}^{\mathrm{n}})^2} \Big) \right], \qquad (2)$$

where $d_{ij}^{\mathrm{p}} = \|\hat{\mathbf{F}}_{\mathrm{s},i}^{\mathrm{P}} - \hat{\mathbf{F}}_{\mathrm{s},j}^{\mathrm{Q}}\|_2$ represents the L2 distances be-

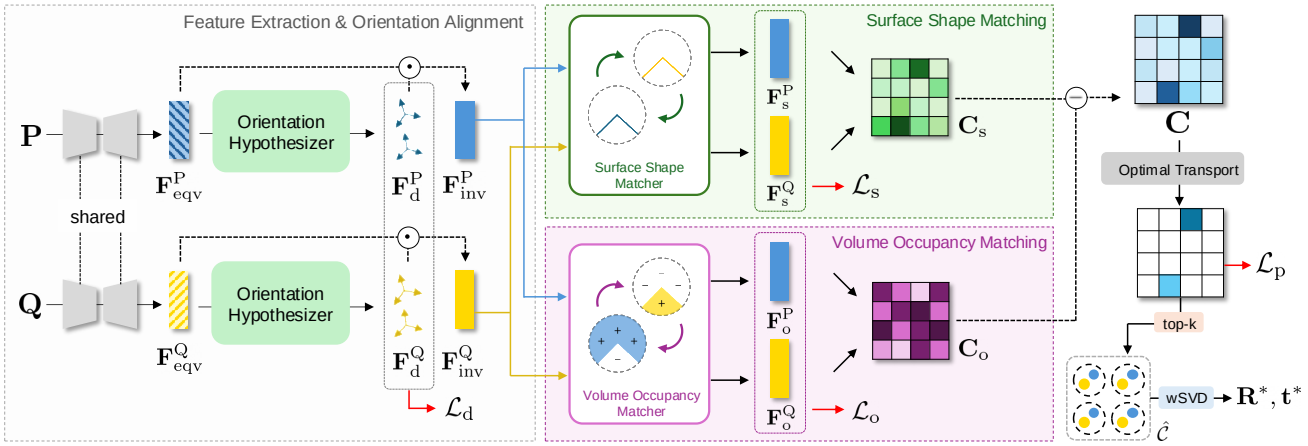

Figure 3. **Overall architecture.** Here, we show core components of (a) feature embedding network, (b) surface shape matching branch, (c) volume occupancy matching branch, and (d) transformation estimation. We refer the readers to Sec. 3.2 for details of each component.

tween shape features in the embedding space, where $\hat{\mathbf{F}}$ denotes the L2-normalized features, and $\mathcal{E}_{\mathrm{p}}(i), \mathcal{E}_{\mathrm{n}}(i)$ are positive/negative correspondences for index i, and $\Delta_{\mathrm{p}}, \Delta_{\mathrm{n}}$ are margin hyperparameters. This formulation encourages positive pairs to achieve distances close to the threshold Δ_{p} while guiding negative pairs toward distances around Δ_{n}, following the conventional design for distinguishing identical surface geometries.

Volume occupancy matching. A key insight for interlocking parts is that their local volumes must *occupy* opposite spaces at the interface. Specifically, if one region is occupied, the corresponding region in the mating part should be unoccupied, and vice versa—creating the interlocking relationship necessary for proper assembly. We encode this idea by learning volume occupancy descriptors $\mathbf{F}_{\mathrm{o}}^{\mathrm{P}} = f_{\mathrm{o}}(\mathbf{P}) \in \mathbb{R}^{N \times d_{\mathrm{o}}}$ and $\mathbf{F}_{\mathrm{o}}^{\mathrm{Q}} = f_{\mathrm{o}}(\mathbf{Q}) \in \mathbb{R}^{M \times d_{\mathrm{o}}}$ that are invariant to rigid transformations through the orientation alignment. To ensure that occupancy descriptors from corresponding regions have opposite values, we define the occupancy matching loss using a variant of circle loss [36]:

$$\mathcal{L}_{\mathrm{o}} \propto \mathop{\mathbb{E}}_{i \sim \mathcal{I}} \left[\log\Big(\sum_{j \in \mathcal{E}_{\mathrm{p}}(i)} e^{(s_{ij}^{\mathrm{p}} - \Delta_{\mathrm{p}})^2} \cdot \sum_{k \in \mathcal{E}_{\mathrm{n}}(i)} e^{(\Delta_{\mathrm{n}} - s_{ik}^{\mathrm{n}})^2} \Big) \right], \tag{3}$$

where $s_{ij}^{\mathrm{p}} = \|\hat{\mathbf{F}}_{\mathrm{o},i}^{\mathrm{P}} + \hat{\mathbf{F}}_{\mathrm{o},j}^{\mathrm{Q}}\|_2 \approx \cos(\mathbf{F}_{\mathrm{o},i}^{\mathrm{P}}, \mathbf{F}_{\mathrm{o},j}^{\mathrm{Q}})$ represents the cosine similarity measures in the occupancy embedding space. It is worth noting that while conventional circle loss typically uses distance metrics to bring positive pairs closer together, our approach leverages cosine similarity to explicitly encourage *opposite* occupancy between positive pairs. This approach treats occupied-unoccupied pairs as *positive* matches, encouraging their descriptors to be opposite, while penalizing non-matching pairs. Thus, the model learns to identify complementary volumes that interlock, rather than matching identical geometries.

Consequently, our Combinative Matching effectively achieves the two essential desiderata for geometric shape assembly: matching identical surface shapes at interfaces and ensuring opposite volume occupancy for proper interlocking, invariant to initial part orientations.

3.2. Combinative Matching Network

We now present the proposed framework that achieves combinative matching through the proposed objectives, capturing the multi-faceted aspects essential for assembly: orientation, shape, and occupancy. Figure 3 illustrates the overall architecture, which consists of five parts: (a) feature extraction and orientation alignment, (b) surface shape matching, (c) volume occupancy matching, (d) transformation estimation, and (e) training objective.

(a) Feature extraction and orientation alignment. For effective combinative matching, surface shape descriptors should ideally be rotation-invariant to ensure robustness across various orientations, while volume occupancy descriptors must retain direction-aligned information within their embedding space to enable complementary alignment. Therefore, prior to applying shape and occupancy matching, we require rotation-invariant features that also encode orientation-consistent information.

To address these requirements, we design a feature embedding network that can embed both clues of orientation-consistency and invariance into a unified representation. We employ VN-EdgeConvs [7], which takes as input a pair of point clouds $\mathbf{P}$ and $\mathbf{Q}$ and provides rotation-equivariant features $\mathbf{F}_{\mathrm{eqv}}^{\mathrm{P}}, \mathbf{F}_{\mathrm{eqv}}^{\mathrm{Q}} \in \mathbb{R}^{K \times D \times 3}$ for each input where K corresponds to N or M depending on the input point cloud, $\mathbf{P}$ and $\mathbf{Q}$, respectively. Next, an orientation hypothesizer, implemented with VN-Linear [7], processes the equivariant features, followed by Gram-Schmidt process with cross-product operation to provide orientations for each point, denoted as $\mathbf{F}_{\mathrm{d}}^{\mathrm{P}}, \mathbf{F}_{\mathrm{d}}^{\mathrm{Q}} \in \mathbb{R}^{K \times 3 \times 3}$. We obtain rotation-invariant

features by taking the dot-product between the equivariant features and the orientation matrices: $\mathbf{F}^{\mathrm{P}}_{\mathrm{inv}} = \mathbf{F}^{\mathrm{P}}_{\mathrm{eqv}} \cdot \mathbf{F}^{\mathrm{P}\top}_{\mathrm{d}}$[2], with the same calculation for $\mathbf{F}^{\mathrm{Q}}_{\mathrm{inv}}$. To ensure these invariant features are aligned consistently with orientation information, we employ the orientation training objective $\mathcal{L}_{\mathrm{d}}$ from Eq. 1, which encourages the features to encode orientation-aligned information while maintaining rotation-invariance, thus ensuring complementary alignment and visual consistency for both shape and occupancy descriptor learning.

(b) Surface shape matching branch. Similar to the way mating surfaces of male and female parts exhibit compatible appearance, we require a distinct feature representation that effectively encodes appearance information to learn shape compatibility. To achieve this, we introduce a dedicated branch that takes the rotation-invariant features $\mathbf{F}^{\mathrm{P}}_{\mathrm{inv}}, \mathbf{F}^{\mathrm{Q}}_{\mathrm{inv}}$ to embed them into surface shape descriptors $\mathbf{F}^{\mathrm{P}}_{\mathrm{s}}, \mathbf{F}^{\mathrm{Q}}_{\mathrm{s}} \in \mathbb{R}^{K\times d_{\mathrm{s}}}$, using a three-layer MLP, followed by LeakyReLU. Applying the surface shape matching objective $\mathcal{L}_{\mathrm{s}}$ from Eq. 2 ensures that matching surfaces with similar appearance are correctly aligned, forming a reliable basis for the subsequent transformation estimation for assembly.

(c) Volume occupancy matching branch. To capture occupancy, we introduce another dedicated branch to learning occupancy descriptors, allowing the model to recognize complementary alignment requirements. This branch begins by taking the $\mathbf{F}^{\mathrm{P}}_{\mathrm{inv}}, \mathbf{F}^{\mathrm{Q}}_{\mathrm{inv}}$ which encode orientation-consistency information in their representations and provide occupancy descriptors $\mathbf{F}^{\mathrm{P}}_{\mathrm{o}}, \mathbf{F}^{\mathrm{Q}}_{\mathrm{o}} \in \mathbb{R}^{K\times d_{\mathrm{o}}}$, using a three-layer MLP with parameters distinct from those in shape matching branch, followed by a Tanh activation. To enforce correct alignment of complementary surfaces, we apply the volume occupancy matching objective $\mathcal{L}_{\mathrm{o}}$ from Eq. 3 which penalizes similarity between descriptors of complementary (occupied *vs.* unoccupied) regions, thereby ensuring corresponding surfaces interlock stably.

(d) Transformation estimation. With the surface shape and volume occupancy descriptor pairs obtained, we construct a cost matrix $\mathbf{C} \in \mathbb{R}^{N\times M}$ that encodes unified correlations across both shape and occupancy characteristics:

$$\mathbf{C} = (\mathbf{F}^{\mathrm{P}}_{\mathrm{s}} \cdot \mathbf{F}^{\mathrm{Q}\top}_{\mathrm{s}} - \mathbf{F}^{\mathrm{P}}_{\mathrm{o}} \cdot \mathbf{F}^{\mathrm{Q}\top}_{\mathrm{o}})/Z, \tag{4}$$

where Z is a normalization constant. In this formulation, the shape descriptors are learned to be similar for positive matches, meaning that their dot product reflects a direct measure of 'similarity' while the occupancy descriptors are trained with the opposite objective, implying their dot product instead represents the 'dissimilarity'. By negating the dissimilarity, $\mathbf{C}$ becomes a similarity measure, integrating the visual similarity with the volumetric complementarity, forming a comprehensive, *combinative* cost matrix.

[2]We refer the readers to the supplementary for a detailed proof.

To obtain a reliable set of correspondence indices $\hat{\mathcal{C}}$, we first apply an Optimal Transport (OT) layer [32] to encourage one-to-one correspondence, then collect the top-k correspondences ($k = 128$), resulting in $|\hat{\mathcal{C}}| = 128$. Finally, we estimate the transformation between the pair of point clouds using weighted SVD [6], formulated as follows:

$$\mathbf{R}^*, \mathbf{t}^* = \arg\min_{\mathbf{R},\mathbf{t}} \sum_{(i,j)\in\hat{\mathcal{C}}} w_{ij} \|\mathbf{R}\mathbf{P}_i + \mathbf{t} - \mathbf{Q}_j\|_2^2, \tag{5}$$

where w_{ij} represents the weight, *e.g.*, the output of OT, for match (i, j). For multi-part assembly, we adopt the same transformation estimation method as used in [14], of which implementation details are provided in the supplementary.

(e) Training objective. Following previous point cloud matching methods [14, 31, 45], we incorporate a point matching loss $\mathcal{L}_{\mathrm{p}}$ [31], cross-entropy loss between ground-truth and predicted match probabilities. By integrating orientation, shape, and occupancy losses along with point matching loss, the final training objective is formulated as:

$$\mathcal{L} = \lambda_{\mathrm{d}}\mathcal{L}_{\mathrm{d}} + \lambda_{\mathrm{s}}\mathcal{L}_{\mathrm{s}} + \lambda_{\mathrm{o}}\mathcal{L}_{\mathrm{o}} + \mathcal{L}_{\mathrm{p}}, \tag{6}$$

where $\lambda_{\mathrm{d}} = 0.1$, $\lambda_{\mathrm{s}} = 0.5$, and $\lambda_{\mathrm{o}} = 0.5$ are weighting coefficients, balancing contributions of different matchings.

4. Experiments

4.1. Dataset and Evaluation Metrics

Dataset. For our experiments, we use the large-scale, standard geometric assembly dataset, Breaking Bad [34], consisting of multiple fractured parts of target objects, categorized into two main subsets: `everyday` and `artifact`. For pairwise shape assembly, we focus specifically on its 2-part subset, while we utilize the entire dataset for multi-part assembly, which consists of objects with 2 to 20 parts. Our experiments are conducted on the volume-constrained Breaking Bad dataset in which the volume of every piece is at least 1/40 of the total shape volume, reducing extreme point density imbalance. For vanilla Breaking Bad benchmark evaluation results, we refer to the supplementary.

Evaluation Metrics. We use the evaluation metrics used in PMTR [14] to validate our method: (1) RMSE between ground truth and predicted rotation and translation parameters, (2) CoRrespondence Distance (CRD), the average distance between positive matches on mating surfaces, and (3) Chamfer Distance (CD) between the input and model-predicted assemblies. Note that CRD provides a more reliable measure than RMSE(R,T) as CRD directly assesses assembly quality while RMSE(R,T) measures relative pose differences without explicitly considering alignment accuracy. Following Lee *et al*. [14], our evaluation is performed with *relative poses* between parts instead of absolute ones to solely focus on the assembly, not absolute positioning.

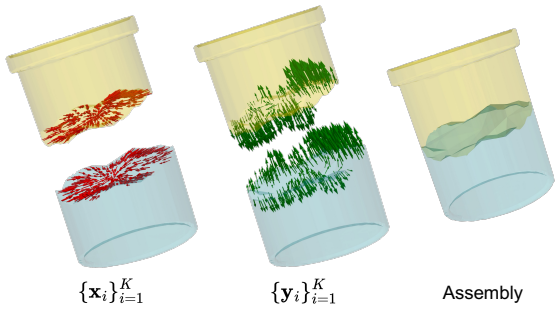

Figure 4. **Visualization of learned orientations.** We visualize learned vectors of $\{\mathbf{x}_i\}_{i\in\mathcal{I}}$ (left, red arrows) and $\{\mathbf{y}_i\}_{i\in\mathcal{I}}$ (middle, green arrows). The assembly result is shown on the right.

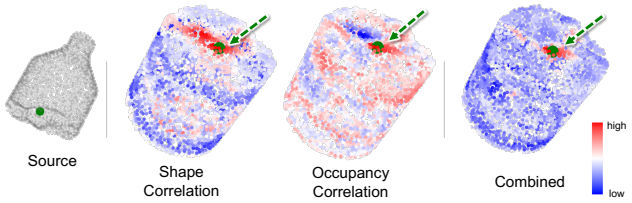

Figure 5. **Visualization of correlation distribution.** A green dot (●) on the left point cloud marks the source's i-th point, with corresponding true match points marked with green dots and arrows. Point colors represent correlation score magnitudes for the i-th point's similarity to each target point, with red and blue indicating high and low correlation scores, respectively.

4.2. Implementation Details

We implement our model using PyTorch Lightning [8]. All experiments were conducted on 4 NVIDIA GeForce RTX 3090 GPUs. We utilize the AdamW [20] optimizer with an initial learning rate of 1×10^{-2}, employing a cosine scheduler set for 90 and 120 epochs on the respective `everyday` and `artifact` subsets. Following previous work of [14, 21], we uniform-sample approximately 5,000 points on the surface per holistic object, with each part allocated a subset of points proportional to its surface area.

4.3. Experimental Results and Analyses

Analysis on learned orientation $\mathbf{F}_\mathrm{d}^\mathrm{P}, \mathbf{F}_\mathrm{d}^\mathrm{Q}$. We begin by analyzing the learned orientations $(\mathbf{F}_\mathrm{d}^\mathrm{P})_i, (\mathbf{F}_\mathrm{d}^\mathrm{Q})_i \in \mathrm{SO}(3)$ to observe the types of information captured through combinative matching, such as the directionality of occupied/unoccupied regions, magnitudes of local concavity or convexity, surface normal, and other relevant geometric properties. For the analysis, we represent each orientation $(\mathbf{F}_\mathrm{d})_i = [\mathbf{x}_i, \mathbf{y}_i, \mathbf{z}_i]$ for all i, where $\mathbf{x}_i, \mathbf{y}_i \in \mathbb{R}^3$ are orthonormal vectors, and $\mathbf{z}_i$ given by $\mathbf{x}_i \times \mathbf{y}_i$. We focus on visualizing the scaled[3] vectors $\mathbf{x}_i$ and $\mathbf{y}_i$, omitting $\mathbf{z}_i$ as it is redundant for interpretative purposes.

[3]They are scaled by the magnitudes of the input vectors for the Gram-Schmidt process, specifically for analysis purposes.

Figure 4 visualizes the vectors $\{\mathbf{x}_i\}_{i\in\mathcal{I}}$ and $\{\mathbf{y}_i\}_{i\in\mathcal{I}}$ for both the source $\mathbf{F}_\mathrm{d}^\mathrm{P}$ and target $\mathbf{F}_\mathrm{d}^\mathrm{Q}$. Through this visualization, we observe several notable patterns: For both $\mathbf{x}_i$ and $\mathbf{y}_i$, (1) source and target orientations are aligned in parallel, as enforced by our training objective $\mathcal{L}_\mathrm{d}$ (Eq. 1). For $\mathbf{x}_i$, we observe that (2) The learned $\mathbf{x}_i$ vectors are consistently directed toward the center of the mating surface, (3) staying parallel to the 2D plane of the mating surface lies, indicating our model has learned a stable orientation that respects the geometry of mating surfaces. For $\mathbf{y}_i$, we observe that: (4) vectors on convex regions (where the surface extends outward into occupied space) point outward, while those on concave regions (where the surface recedes) point inward. (5) The magnitudes of $\mathbf{y}_i$ correlate with the degree of convexity or concavity at each point, indicating an awareness of surface curvature. These results imply that the learned orientations not only differentiate between convex and concave structures but also capture complementarity and directional alignment *without any explicit supervisions dedicated to these aspects from (2) to (5)*, highlighting the efficacy of the proposed combinative matching in intuitive learning of integral properties for 'combining' elements.

Analysis on learned correlations. To validate how the proposed combinative matching resolves a limitation of conventional shape-based matching (*e.g.*, local ambiguity), we compare correlation matrices for shape, occupancy, and the combined cost matrix: specifically, $\mathbf{C}_\mathrm{s} = \mathbf{F}_\mathrm{s}^\mathrm{P}\mathbf{F}_\mathrm{s}^{\mathrm{Q}\top} \in \mathbb{R}^{N\times M}$, $\mathbf{C}_\mathrm{o} = \mathbf{F}_\mathrm{o}^\mathrm{P}\mathbf{F}_\mathrm{o}^{\mathrm{Q}\top} \in \mathbb{R}^{N\times M}$, and $\mathbf{C}$. For this analysis, we select a single point on the source's mating surface (index i) and examine its similarity distributions of these correlations: $(\mathbf{C}_\mathrm{s})_i, (\mathbf{C}_\mathrm{o})_i, \mathbf{C}_i \in \mathbb{R}^M$. These distributions are visualized as heatmaps, with red and blue colors indicating high and low similarities, respectively (we invert the color for $\mathbf{C}_\mathrm{o}$ to reflect its representation of dissimilarity). Figure 5 presents the visualized distributions.

Based solely on the surface shape distribution, the best target match for the i-th source point is located in a broad area due to the similar appearance of surrounding points, resulting in local ambiguity. The volume occupancy distribution, on the other hand, shows large scores are nearly uniformly spread across the surface, with a slightly higher score near the true match, indicating it provides complementary information yet lacks distinct localization. By combining shape and occupancy information, the local ambiguity and match confidence uncertainty are resolved; the score at the true match is significantly higher, enabling precise alignment of the source point with its correct match on the target, verifying that the combinative matching effectively enhances precision by resolving the local ambiguity.

Ablation studies. To assess the contributions of key components in our method, we conduct ablation studies on `everyday` subset. First, Tab. 1 (a) examines the choice of affinity metric (s^p and s^n in Eq. 3) and the impact of ori-

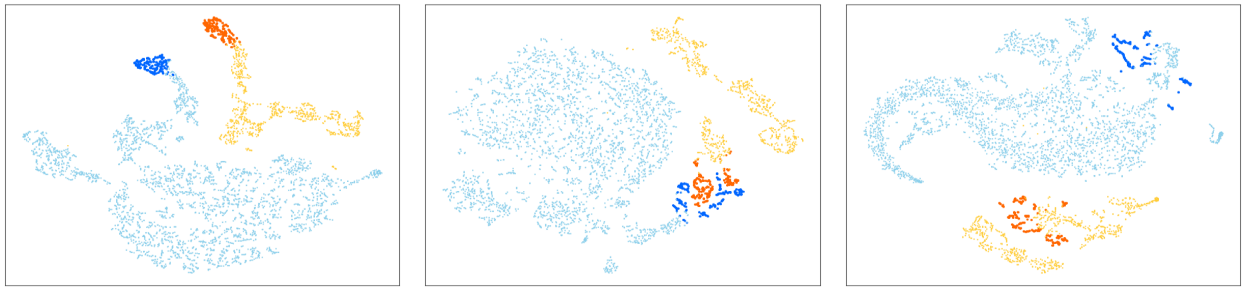

Figure 6. **Feature visualization via t-SNE.** Mating surface points are displayed in blue (●) for the source and orange (●) for the target, while non-mating surface points are colored in skyblue (●) for the source and yellow (●) for the target.

Occupancy Affinity	Orientation Loss ($\mathcal{L}_d$)	CRD ↓ (10^{-2})	CD ↓ (10^{-3})	RMSE(R) ↓ (°)	RMSE(T) ↓ (10^{-2})
L2 dist	✗	0.42	0.31	14.88	4.31
cosine	✗	0.31	0.21	14.58	4.44
L2 dist	✓	0.38	0.30	13.29	3.81
cosine	✓	**0.28**	**0.17**	**12.88**	**3.78**

(a) Ablation study on combinative matching.

Equivariant Embedding	Shape Matching	Occupancy Matching	CRD ↓ (10^{-2})	CD ↓ (10^{-3})	RMSE(R) ↓ (°)	RMSE(T) ↓ (10^{-2})
✗	✓	✓	0.74	0.53	38.74	11.88
✓	✗	✓	0.38	0.28	13.17	3.86
✓	✓	✗	0.35	0.25	14.01	4.24
✓	✓	✓	**0.28**	**0.17**	**12.88**	**3.78**

(b) Ablation study on model components.

Table 1. Ablation studies of the proposed approach.

Method	CRD ↓ (10^{-2})	CD ↓ (10^{-3})	RMSE(R) ↓ (°)	RMSE(T) ↓ (10^{-2})
`everyday` → `artifact`				
NSM [5]	19.95	6.88	84.16	21.74
Wu et al. [43]	19.13	7.98	85.27	22.96
GeoTransformer [31]	1.01	0.78	33.14	9.75
Jigsaw [21]	10.36	2.48	56.98	10.36
PMTR [14]	0.82	0.59	29.63	9.21
Ours	**0.74**	**0.54**	**25.67**	**7.73**
`artifact` → `everyday`				
NSM [5]	21.34	8.52	85.46	23.58
Wu et al. [43]	20.70	11.67	85.81	22.96
GeoTransformer [31]	0.80	0.53	41.65	13.23
Jigsaw [21]	11.00	3.04	70.88	10.75
PMTR [14]	0.64	**0.44**	33.23	10.97
Ours	**0.62**	0.46	**26.91**	**8.30**

Table 2. Transferability experiments on Breaking Bad [34].

entation loss $\mathcal{L}_o$. Using L2 distance instead of cosine similarity, or omitting orientation loss during training, results in consistent performance drops, implying the importance of both complementarity and orientation learning in assembly. Second, Tab. 1 (b) evaluates the impact of equivariant network [7] and surface shape & volume occupancy matching branches. When the equivariant backbone is replaced with a standard point embedding network, *e.g.*, DGCNN [40], we observe a substantial drop in CRD, verifying the importance of learning orientation-awareness and rotation-invariance in assembly. Consistent accuracy drops in the absence of either shape or occupancy matching branch demonstrate that both branches work synergistically to enhance alignment.

Learned descriptor analysis. In the proposed network, $\mathbf{F}_s$ and $\mathbf{F}_o$ are optimized through the ***opposing*** objectives: $\mathcal{L}_s$ clusters mating surface features for positive matches while separating negative ones, whereas $\mathcal{L}_o$ penalizes features for positive matches, each within its respective embedding space. To examine their clustering behavior, we project the invariant features $\mathbf{F}_{inv}$, shape descriptors $\mathbf{F}_s$, and occupancy descriptors $\mathbf{F}_o$ into a 2D space using t-SNE and visualize the results in Fig. 6.

For invariant features of mating surface, we observe neither separation nor adhesion in their embedding space, implying that the invariance property alone does not provide significant feature distinction. In contrast, the shape descriptors lying on mating surface are tightly clustered as enforced by $\mathcal{L}_s$, supported by the visual resemblance of the interface. Meanwhile, the occupancy descriptors on mating surface are more widely dispersed, guided by $\mathcal{L}_o$. The results collectively highlight the efficacy of proposed learning objectives in shaping the embedding space, reflecting both visual and volumetric properties of mating surfaces.

4.4. Model generalizability

To demonstrate the generalizability of our approach, we conduct transferability experiments within the Breaking Bad dataset [34]. Specifically, we evaluate our model, trained on the `everyday` subset, on the `artifact` subset, and vice versa. The `everyday` subset primarily consists of common objects relevant to computer vision and robotics applications, whereas the `artifact` subset focuses on archaeological objects, representing a notable domain shift between the two subsets.

Table 2 summarizes the transferability results. The proposed method consistently outperforms state-of-the-art baselines, achieving higher CRDs across cross-subset evaluations. This highlights the robustness of our model in adapting to different data domains, underscoring its efficacy in capturing task-oriented features of orientation, shape, and occupancy generalizable across distinct object categories.

Method	CRD ↓ (10^{-2})	CD ↓ (10^{-3})	RMSE(R) ↓ (°)	RMSE(T) ↓ (10^{-2})
everyday				
NSM [5]	21.71	11.09	83.38	23.71
Wu et al. [43]	20.65	11.66	84.58	22.90
GeoTransformer [31]	0.61	0.51	22.81	7.28
Jigsaw [21]	5.48	1.34	38.73	**2.73**
PMTR [14]	0.39	0.25	17.14	5.53
Ours	**0.28**	**0.17**	**12.88**	3.78
artifact				
NSM [5]	19.44	6.33	83.22	21.41
Wu et al. [43]	19.17	7.97	85.04	20.90
GeoTransformer [31]	0.89	0.70	33.23	10.30
Jigsaw [21]	6.36	1.45	39.71	**3.02**
PMTR [14]	0.60	0.42	23.28	7.27
Ours	**0.49**	**0.34**	**18.77**	5.57

Table 3. Pairwise shape assembly results. Numbers in **bold** indicate the best performance and underlined ones are the second best.

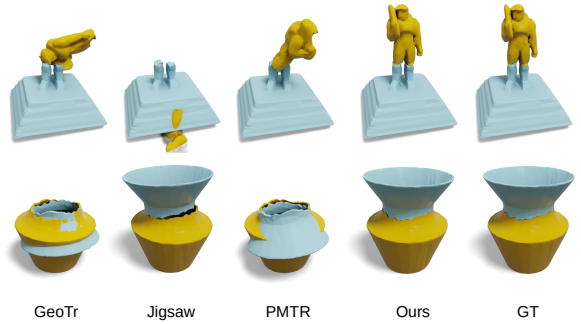

Figure 7. Qualitative comparison for pairwise shape assembly.

4.5. Comparison with State of the Arts

To validate the efficacy of the proposed method, we compare it with recent baselines on pairwise assembly in Tab. 3 and Fig. 7. In terms of CRD and CD, our method outperforms all the baselines in both `everyday` and `artifact` subsets, achieving relative CRD improvements of 28% and 18%, respectively, over the previous state of the art [14].

We further evaluate our method on multi-part assembly with additional metrics, Part Accuracy (PA) [14, 17], and compare the results in Tab. 4 and Fig. 8, where ours consistently shows superior numbers compared to baselines. The results confirm that, unlike methods that rely solely on visual cues [14, 31], our combinative matching enables more reliable shape assembly, as evident from Figs. 7 and 8. We refer to the supplementary for additional qualitative results.

5. Limitations and Future Work

While our method demonstrates robust performance in most assembly scenarios, it still fails in certain challenging scenarios, as illustrated in Fig. 9. First, given extremely low overlap between mating surfaces (a,b), the occupancy cues become too weak to provide reliable guidance. Second, when fracture surfaces are visually indistinguishable (b,c),

Method	CRD ↓ (10^{-2})	CD ↓ (10^{-3})	RMSE(R) ↓ (°)	RMSE(T) ↓ (10^{-2})	PA_{CRD} ↑ (%)	PA_{CD} ↑ (%)
everyday						
Global [15, 33]	27.79	15.30	55.42	15.31	36.42	37.90
LSTM [42]	27.69	15.23	54.78	15.24	36.74	38.97
DGL [10]	27.90	13.23	55.76	15.33	36.99	39.70
Wu et al. [43]	28.18	19.70	54.98	15.59	35.66	36.28
Jigsaw [21]	14.13	11.82	41.12	11.74	52.48	60.26
PMTR [14]	6.51	5.56	31.57	9.95	66.95	70.56
Ours	**5.18**	**3.65**	**27.11**	**8.13**	**73.88**	**77.88**
artifact						
Global [15, 33]	26.42	14.92	54.41	14.48	36.67	36.97
LSTM [42]	28.15	14.61	53.59	15.49	36.67	37.25
DGL [10]	27.48	13.91	54.66	15.10	36.66	37.40
Wu et al. [43]	26.02	15.81	54.35	14.27	36.63	37.02
Jigsaw [21]	16.10	9.53	42.01	17.47	56.93	65.58
PMTR [14]	5.67	4.33	31.58	10.08	66.96	71.61
Ours	**4.56**	**3.04**	**29.21**	**8.99**	**71.02**	**76.32**

Table 4. Multi-part assembly results. Numbers in **bold** indicate the best performance and underlined ones are the second best.

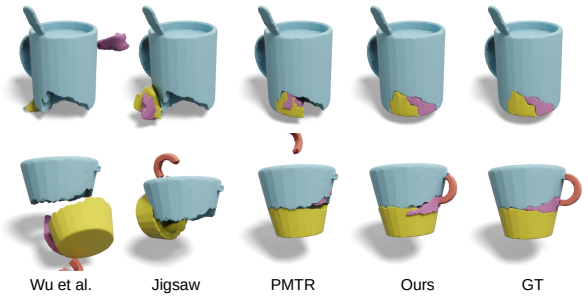

Figure 8. Qualitative comparison for multi-part assembly.

the pairwise matching scores become less discriminative, often resulting in incorrect part permutations. Integrating additional information such as texture & color cues, or enforcing cycle-consistency, could address such ambiguities.

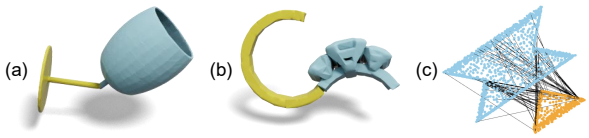

Figure 9. **(a-b)** Representative failure cases on Breaking Bad. **(c)** Visualization of top-k matches on a toy example ($k = 128$).

6. Conclusion

We have introduced combinative matching that incorporates the multi-faceted, task-oriented properties, which demonstrated the superiority over recent baselines by capturing intrinsic properties of assembly, such as degrees of convexity/concavity and orientation of mating surfaces, even without explicit supervision. Although this paper explores learning orientation, shape, and occupancy matching, the method can be further expanded to incorporate properties like physical compatibility or functional constraints, paving the way for more versatile assembly frameworks.

References

[1] M Altınok, I Kureli, S Doganay, and A Onduran. Mechanical performance of woodwork joinery produced by industrıal methods. *Journal of Applied Mechanical Engineering*, 5:228, 2016. 2

[2] Federico Campi, Claudio Favi, Michele Germani, and Marco Mandolini. Cad-integrated design for manufacturing and assembly in mechanical design. *International Journal of computer integrated manufacturing*, 35(3):282–325, 2022. 1

[3] Carlo Canali, Ferdinando Cannella, Fei Chen, Giuseppe Sofia, Amit Eytan, and Darwin G Caldwell. An automatic assembly parts detection and grasping system for industrial manufacturing. In *2014 IEEE international conference on automation science and engineering (CASE)*, pages 215–220. IEEE, 2014. 1

[4] Olcay Ersel Canyurt, Cemal Meran, and Mine Uslu. Strength estimation of adhesively bonded tongue and groove joint of thick composite sandwich structures using genetic algorithm approach. *International Journal of Adhesion and Adhesives*, 30(5):281–287, 2010. 2

[5] Yun-Chun Chen, Haoda Li, Dylan Turpin, Alec Jacobson, and Animesh Garg. Neural shape mating: Self-supervised object assembly with adversarial shape priors. In *CVPR*, 2022. 1, 2, 3, 7, 8

[6] Christopher Choy, Wei Dong, and Vladlen Koltun. Deep global registration. In *CVPR*, 2020. 5

[7] 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 *ICCV*, 2021. 2, 4, 7

[8] William Falcon and The PyTorch Lightning team. Pytorch lightning, 2019. 6

[9] Kensuke Harada, Kazuyuki Nagata, Juan Rojas, Ixchel G Ramirez-Alpizar, Weiwei Wan, Hiromu Onda, and Tokuo Tsuji. Proposal of a shape adaptive gripper for robotic assembly tasks. *Advanced Robotics*, 30(17-18):1186–1198, 2016. 1

[10] Jialei Huang, Guanqi Zhan, Qingnan Fan, Kaichun Mo, Lin Shao, Baoquan Chen, Leonidas Guibas, and Hao Dong. Generative 3d part assembly via dynamic graph learning. In *NeurIPS*, 2020. 2, 3, 8

[11] Shengyu Huang, Zan Gojcic, Mikhail Usvyatsov, Andreas Wieser, and Konrad Schindler. Predator: Registration of 3d point clouds with low overlap. In *CVPR*, 2021. 2

[12] Seungwook Kim, Chunghyun Park, Yoonwoo Jeong, Jaesik Park, and Minsu Cho. Stable and consistent prediction of 3d characteristic orientation via invariant residual learning. In *ICML*, 2023. 2

[13] Jiří Kunecký, Anna Arciszewska-Kędzior, Václav Sebera, and Hana Hasníková. Mechanical performance of dovetail joint related to the global stiffness of timber roof structures. *Materials and Structures*, 49:2315–2327, 2016. 2

[14] Nahyuk Lee, Juhong Min, Junha Lee, Seungwook Kim, Kanghee Lee, Jaesik Park, and Minsu Cho. 3d geometric shape assembly via efficient point cloud matching. In *ICML*, 2024. 1, 2, 3, 5, 6, 7, 8

[15] Jun Li, Chengjie Niu, and Kai Xu. Learning part generation and assembly for structure-aware shape synthesis. In *AAAI*, 2020. 2, 3, 8

[16] Jianning Li, Antonio Pepe, Gijs Luijten, Christina Schwarz-Gsaxner, Jens Kleesiek, and Jan Egger. Anatomy completor: A multi-class completion framework for 3d anatomy reconstruction. In *International Workshop on Shape in Medical Imaging*, pages 1–14. Springer, 2023. 1

[17] Yichen Li, Kaichun Mo, Lin Shao, Minhyuk Sung, and Leonidas Guibas. Learning 3d part assembly from a single image. In *ECCV*, 2020. 8

[18] Yichen Li, Kaichun Mo, Yueqi Duan, He Wang, Jiequan Zhang, and Lin Shao. Category-level multi-part multi-joint 3d shape assembly. In *CVPR*, 2024. 2, 3

[19] Keyang Liu, Yao Du, Xiaohong Hu, Hualei Zhang, Luhao Wang, Wenhao Gou, Li Li, Hongguang Liu, and Bin Luo. Investigating the influence of tenon dimensions on white oak (quercus alba) mortise and tenon joint strength. *Forests*, 15 (9):1612, 2024. 2

[20] Ilya Loshchilov and Frank Hutter. Decoupled weight decay regularization. In *International Conference on Learning Representations*, 2019. 6

[21] Jiaxin Lu, Yifan Sun, and Qixing Huang. Jigsaw: Learning to assemble multiple fractured objects. In *NeurIPS*, 2023. 2, 3, 6, 7, 8

[22] Shitong Luo, Jiahan Li, Jiaqi Guan, Yufeng Su, Chaoran Cheng, Jian Peng, and Jianzhu Ma. Equivariant point cloud analysis via learning orientations for message passing. In *CVPR*, 2022. 2

[23] William Marande and Gertraud Burger. Mitochondrial dna as a genomic jigsaw puzzle. *Science*, 318(5849):415–415, 2007. 1

[24] Karel Matouš and George J Dvorak. Analysis of tongue and groove joints for thick laminates. *Composites Part B: Engineering*, 35(6-8):609–617, 2004. 2

[25] Jonah C McBride and Benjamin B Kimia. Archaeological fragment reconstruction using curve-matching. In *CVPRW*, 2003. 1

[26] Amir Mollahassani, A Hemmasi, H Khademi Eslam, Amir Lashgari, and Behzad Bazyar. Dynamic and static comparison of beech wood dovetail, tongue and groove, halving, and dowel joints. *BioResources*, 15(2):3787, 2020. 2

[27] Xiaolei Niu, Qifeng Wang, Bin Liu, and Jianxin Zhang. An automatic chinaware fragments reassembly method framework based on linear feature of fracture surface contour. *ACM Journal on Computing and Cultural Heritage*, 16(1): 1–22, 2022. 1

[28] Keita Ogawa, Yasutoshi Sasaki, and Mariko Yamasaki. Theoretical estimation of the mechanical performance of traditional mortise–tenon joint involving a gap. *Journal of Wood Science*, 62:242–250, 2016. 2

[29] Shigefumi Okamoto, Makoto Nakatani, Nobuhiko Akiyama, Kei Tanaka, and Takuro Mori. Verification of the shear performance of mortise and tenon joints with top and bottom notches at the beam end. *Journal of Wood Science*, 67:1–23, 2021. 2

[30] Charles R Qi, Hao Su, Kaichun Mo, and Leonidas J Guibas. Pointnet: Deep learning on point sets for 3d classification and segmentation. In *CVPR*, 2017. 2

[31] Zheng Qin, Hao Yu, Changjian Wang, Yulan Guo, Yuxing Peng, and Kai Xu. Geometric transformer for fast and robust point cloud registration. In *CVPR*, 2022. 2, 5, 7, 8

[32] Paul-Edouard Sarlin, Daniel DeTone, Tomasz Malisiewicz, and Andrew Rabinovich. Superglue: Learning feature matching with graph neural networks. In *CVPR*, 2020. 5

[33] Nadav Schor, Oren Katzir, Hao Zhang, and Daniel Cohen-Or. Componet: Learning to generate the unseen by part synthesis and composition. In *ICCV*, 2019. 2, 3, 8

[34] Silvia Sellán, Yun-Chun Chen, Ziyi Wu, Animesh Garg, and Alec Jacobson. Breaking bad: A dataset for geometric fracture and reassembly. In *NeurIPS Datasets and Benchmarks Track*, 2022. 2, 5, 7

[35] Kilho Son, Eduardo B Almeida, and David B Cooper. Axially symmetric 3d pots configuration system using axis of symmetry and break curve. In *CVPR*, 2013. 1

[36] Yifan Sun, Changmao Cheng, Yuhan Zhang, Chi Zhang, Liang Zheng, Zhongdao Wang, and Yichen Wei. Circle loss: A unified perspective of pair similarity optimization. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 6398–6407, 2020. 3, 4

[37] Garrett Thomas, Melissa Chien, Aviv Tamar, Juan Aparicio Ojea, and Pieter Abbeel. Learning robotic assembly from cad. In *ICRA*, 2018. 1

[38] Haiping Wang, Yuan Liu, Zhen Dong, and Wenping Wang. You only hypothesize once: Point cloud registration with rotation-equivariant descriptors. In *ACM International Conference on Multimedia*, 2022. 2

[39] Haiping Wang, Yuan Liu, Qingyong Hu, Bing Wang, Jianguo Chen, Zhen Dong, Yulan Guo, Wenping Wang, and Bisheng Yang. Roreg: Pairwise point cloud registration with oriented descriptors and local rotations. *IEEE TPAMI*, 45(8): 10376–10393, 2023. 2

[40] Yue Wang, Yongbin Sun, Ziwei Liu, Sanjay E Sarma, Michael M Bronstein, and Justin M Solomon. Dynamic graph cnn for learning on point clouds. *ACM TOG*, 38(5): 1–12, 2019. 7

[41] Ze Wang and Mingfu Liao. Stiffness and deformation characteristics of the rabbet joint structure in the central tie-rod fastening rotor. In *2020 Global Reliability and Prognostics and Health Management (PHM-Shanghai)*, pages 1–7. IEEE, 2020. 2

[42] Rundi Wu, Yixin Zhuang, Kai Xu, Hao Zhang, and Baoquan Chen. Pq-net: A generative part seq2seq network for 3d shapes. In *CVPR*, 2020. 2, 3, 8

[43] Ruihai Wu, Chenrui Tie, Yushi Du, Yan Zhao, and Hao Dong. Leveraging se (3) equivariance for learning 3d geometric shape assembly. In *ICCV*, 2023. 2, 3, 7, 8

[44] Qifang Xie, Baozhuang Zhang, Lipeng Zhang, Tiantian Guo, and Yajie Wu. Normal contact performance of mortise and tenon joint: theoretical analysis and numerical simulation. *Journal of Wood Science*, 67:1–21, 2021. 2

[45] Runzhao Yao, Shaoyi Du, Wenting Cui, Canhui Tang, and Chengwu Yang. Pare-net: Position-aware rotation-equivariant networks for robust point cloud registration. In *ECCV*, 2024. 2, 5

[46] Hao Yu, Fu Li, Mahdi Saleh, Benjamin Busam, and Slobodan Ilic. Cofinet: Reliable coarse-to-fine correspondences for robust pointcloud registration. In *NeurIPS*, 2021. 2

[47] Hao Yu, Zheng Qin, Ji Hou, Mahdi Saleh, Dongsheng Li, Benjamin Busam, and Slobodan Ilic. Rotation-invariant transformer for point cloud matching. In *CVPR*, 2023. 2

[48] Wei Yue, Qing Mei, Dayi Zhang, et al. Robust design method of rabbet joint structure in high speed assemble rotor. *Journal of Aerospace Power*, 32(7):1754–1761, 2017. 2

[49] Kevin Zakka, Andy Zeng, Johnny Lee, and Shuran Song. Form2fit: Learning shape priors for generalizable assembly from disassembly. In *ICRA*, 2020. 1

[50] Yubo Zhang, Jan-Michael Frahm, Samuel Ehrenstein, Sarah K McGill, Julian G Rosenman, Shuxian Wang, and Stephen M Pizer. Colde: a depth estimation framework for colonoscopy reconstruction. *arXiv preprint arXiv:2111.10371*, 2021. 1

[51] Tianyi Zhou, Hang Gao, Xuanping Wang, Lun Li, Jianfeng Chen, and Can Peng. Prediction method of aeroengine rotor assembly errors based on a novel multi-axis measuring and connecting mechanism. *Machines*, 10(5):387, 2022. 2