Symmetry Informative and Agnostic Feature Disentanglement for 3D Shapes

Tobias Weißberg Weikang Wang Paul Roetzer Nafie El Amrani Florian Bernard
University of Bonn & Lamarr Institute

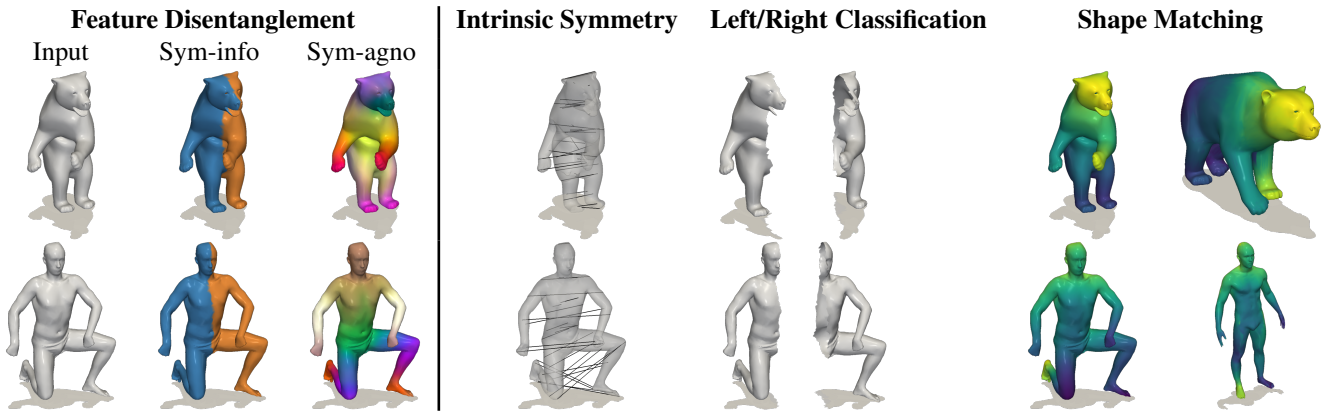

Figure 1. Left: Visualization of disentangled intrinsic symmetry-informative (abbreviated as sym-info) and intrinsic symmetry-agnostic (abbreviated as sym-agno) descriptors. Right: Results of various shape analysis tasks (intrinsic symmetry detection, left/right classification and shape matching) using our disentangled intrinsic symmetry-aware descriptor pairs.

Abstract

Shape descriptors, i.e., per-vertex features of 3D meshes or point clouds, are fundamental to shape analysis. Historically, various handcrafted geometry-aware descriptors and feature refinement techniques have been proposed. Recently, several studies have initiated a new research direction by leveraging features from image foundation models to create semantics-aware descriptors, demonstrating advantages across tasks like shape matching, editing, and segmentation. Symmetry, another key concept in shape analysis, has also attracted increasing attention. Consequently, constructing symmetry-aware shape descriptors is a natural progression. Although the recent method χ [53] successfully extracted symmetry-informative features from semantic-aware descriptors, its features are only one-dimensional, neglecting other valuable semantic information. Furthermore, the extracted symmetry-informative feature is usually noisy and yields small misclassified patches. To address these gaps, we propose a feature disentanglement approach which is simultaneously symmetry informative and symmetry agnostic. Further, we propose a feature refinement technique to improve the robustness of predicted symmetry informative features. Extensive experiments, including intrinsic symmetry detection, left/right classification, and shape matching, demonstrate the effectiveness of our proposed framework compared to various state-of-the-art methods, both qualitatively and quantitatively. Project page: https://tweissberg.github.io/chirality/

1. Introduction

Unlike image analysis, where features from foundational models (e.g. DINO-V2 [33], CLIP [36], StableDiffusion [43]) have already surpassed handcrafted features across a broad range of tasks [10, 20, 50, 60, 61], per-vertex features in shape analysis have still mostly been dominated by handcrafted descriptors such as WKS [3], HKS [48], SHOT [45], with different feature refinement methods [8, 46].

To address this gap, a recent work, Diff3F [15], has introduced a new semantic-aware shape descriptor construction pipeline, which aggregated pixel-level 2D foundation model features from surrounding multi-view images rendered from a shape. Works that directly use Diff3F features [2, 19, 42, 54, 59] or implement similar ideas together with other modules [11, 51, 65], have demonstrated advantages over handcrafted features across various shape analysis tasks, including matching [11, 42, 47, 51, 54, 65], editing [19, 59] and shape-image correspondence [2, 47].

However, as demonstrated in [60, 61], features from 2D

foundation models can be ambiguous between intrinsic symmetry structures, such as left and right eyes of a cat. Since Diff3F [15] aggregates features from these 2D foundation models, it inherits left/right ambiguity which can lead to incorrect predictions in downstream tasks as shown in [53].

In order to address this intrinsic symmetry ambiguity in Diff3F [15] features, the recent work [53] has proposed an intrinsic symmetry feature extraction pipeline to distill a one-dimensional symmetry-informative feature out of input semantic-aware shape descriptors. However, despite its impressive accuracy on left/right classification and the ability to boost matching performance, certain drawbacks exist. First, [53] only extracts a single one-dimensional symmetry-informative feature, rather than disentangling input feature into symmetry informative and agnostic parts. Secondly, it has limited ability to perform tasks such as intrinsic symmetry detection due to lack of sophisticated losses guiding the training process. In addition, as noted in [53], resulting symmetry-informative features are noisy, i.e. contain small mispredicted patches, see Fig. 2 for an illustration.

In order to improve upon these drawbacks, we propose a feature disentanglement approach which simultaneously extracts symmetry informative and symmetry agnostic features. The former encodes only the left/right information, while the latter contains the remaining information of the input features. Ideally, the symmetry-informative descriptor tells whether a vertex belongs to the left or right part of the shape, and the symmetry-agnostic descriptor should be similar for intrinsically symmetric pairs of points, e.g. points on the left/right thumb of a human shape. We adopt a combination of various unsupervised losses to guide the training of our intrinsic symmetry disentanglement network. In addition, to obtain a robust symmetry-informative descriptor, we also propose a symmetry-informative feature refinement technique based on a Markov Random Field (MRF) energy minimization formalism. Various experiments including intrinsic symmetry detection, left/right classification and matching have been conducted to show the effectiveness of our disentanglement framework compared to other state-of-the-art methods. We summarize our main contributions as follows:

- We propose an intrinsic symmetry-aware shape feature disentanglement framework, capable of decomposing per-vertex shape features into symmetry-informative and symmetry-agnostic pairs.
- We propose a generally applicable symmetry-aware feature refinement technique to obtain a robust symmetry-informative descriptor.
- We show the broad applicability of our method by applying it to various shape analysis tasks including symmetry detection, left/right classification, and shape matching.
- Through extensive empiric evaluation we demonstrate the superiority of our disentangled symmetry-aware descriptors over state-of-the-art methods.

Figure 2. χ [53] in some cases produces small patches of incorrectly assigned vertices. Red circles highlight these areas.

2. Related work

In the following we discuss works that are most relevant to our framework. We discuss symmetry in shape analysis and 2D foundation models used in 3D shape descriptors.

2.1. Symmetry in shape analysis

As a fundamental pattern observable in our world, symmetry attracts significant interest in shape analysis, because of its potential to enhance a wide range of tasks including matching [12, 58, 63], segmentation [39, 40], completion [27, 44, 49] and editing [30, 64]. Given the vast literature on symmetry detection, we introduce only the most relevant works and refer readers to surveys [26, 31].

Following a common convention [21, 31, 37, 38], symmetry can be categorized into extrinsic and intrinsic ones. Extrinsic symmetry can be characterized via rigid transformations, while intrinsic symmetry is shape-inherent and is invariant to isometric transformations (e.g., symmetry independently of poses of a human body). We focus on intrinsic symmetry in this work, since it is more general in the sense that it contains extrinsic symmetry as a subset.

Ovsjanikov et al. [34] show that the intrinsic symmetries of a shape could be transformed into the extrinsic symmetries in the signature space defined by the eigenfunctions of the Laplace-Beltrami operator, and devise an algorithm to detect and compute them. Its follow-up work Nagar and Raman [32] improves efficiency based on the hypothesis that if a shape is intrinsically symmetric, then so is the geodesic distance between two intrinsic symmetric vertices. Xu et al. [55] use a generalized voting scheme to find the partial intrinsic symmetry curve without explicitly finding the intrinsic symmetric counterpart for each vertex. Liu et al. [25] propose a method for detecting intrinsic symmetry on genus-zero mesh surfaces by extracting closed curves through the construction of a weighted blend of conformal maps derived from triplets of extremal points identified via the average geodesic distance (AGD) function. Similarly, Kim et al. [21] also detect symmetry invariant point sets using critical points of the AGD function, but get intrinsic symmetry from Möbius transformations computed by those selected points. Xu et al. [56] efficiently find pairs of intrinsically symmetric points using a voting based approach to detect intrinsic symmetries of shapes in different scales. Wang and Huang

[52] describe symmetries as linear transformations of the eigenfunctions of the Laplace-Beltrami operator on shapes, and they propose an efficient global intrinsic symmetry detection method based on this new representation. Qiao et al. [35] parametrize intrinsic symmetry using a functional map matrix, which could be easily computed given the signs of Laplacian eigenfunctions under the symmetric mapping.

2.2. 2D foundation models assisted 3D shape descriptors

Recently, with the emergence of various 2D foundation models [9, 33, 36, 43] and their demonstrated superiority over traditional features in various image-related tasks [10, 20, 50, 60, 61], an increasing number of works in 3D areas are considering using 2D foundation models to obtain 3D shape descriptors. 3D Highlighter [13] uses a pre-trained CLIP [36] encoder to localize semantic regions on a mesh using text as input. As a popular 3D scene representation, NeRF [29] has various following works that leverage 2D foundation models for 3D representations. Kobayashi et al. [23] distill the knowledge of 2D foundation models [9, 36] to a 3D feature field optimized in parallel to the radiance field for semantic scene decomposition task. FeatureNeRF [57] leverages 2D foundation models [9, 43] to 3D space via neural rendering, and then extract deep features for querying 3D points from NeRF layers. NeRF Analogies [18] performs transfer along semantic affinity driven by semantic features from some 2D foundation model [9] to achieve multi-view appearance consistency. Different from implicit NeRF representations, there is also a line of research aiming at building explicit per-vertex 3D shape descriptors from 2D foundation models. With a shape as input, Diff3F [15] constructs per-vertex shape descriptor by averaging correspondent pixel features of 2D foundation models [33, 43] from surrounding rendered multi-view images. Uzolas et al. [51] refines Diff3F [15] features on the shape matching task by introducing a constraint that the geodesic distance between refined features of each vertex pair should be consistent with geodesic distance of this vertex pair. DenseMatcher [65] uses a similar pipeline as Diff3F [15] to first extract per-vertex features, and further refine shape descriptors using a trainable DiffusionNet [46].

A recent work, χ [53], is most relevant to this work. It extracts intrinsic symmetry-informative features from input semantic-aware shape descriptors aggregated from 2D foundation models similarly to Diff3F [15]. In this work, our proposed framework disentangles input semantic-aware shape descriptors into symmetry-informative and symmetry-agnostic pairs, rather than only extracting one-dimensional feature as χ [53] does. Better performances on intrinsic symmetry detection and shape matching show the effectiveness and validate the necessity of this newly introduced symmetry-agnostic descriptor, thus our proposed framework.

3. Symmetry informative and agnostic feature disentanglement

In this section, we introduce our intrinsic symmetry-aware feature disentanglement framework (Fig. 3) in detail. We disentangle input shape descriptors into symmetry-informative and symmetry-agnostic pairs. We first explain the process of obtaining per-vertex semantic-aware feature (Sec. 3.1), which our lightweight disentanglement network (Sec. 3.2) uses as input. Next, the losses guiding the training of our network are elaborated in detail (Sec. 3.3). And finally, we propose a general symmetry-informative descriptor refinement technique to improve the robustness of descriptors (Sec. 3.4).

3.1. Shape descriptors extraction

We follow Diff3F [15] and χ [53] to extract per-vertex semantic-aware shape descriptors. We consider an untextured 3D surface shape in form of a mesh $\mathcal{X} = (\mathcal{V}, \mathcal{E})$ as input. Here, $\mathcal{V}$ are the vertices of the shape and $\mathcal{E}$ the edges. For that shape $\mathcal{X}$, we first generate N surrounding textured views $\{I_n \in \mathbb{R}^{H\times H\times 3}\}_{n=1}^{N}$ from N fixed camera poses using ControlNet [62]. Then features from 2D foundation models are extracted, concatenated and upsampled to form the feature map $F_n \in \mathbb{R}^{H\times W\times d}$ of each view I_n with feature dimension d.

Using the known camera poses, we can recover correspondences between each vertex v and pixels in each view. Then we can compute the per-vertex feature $\mathcal{F}_v \in \mathbb{R}^d$ as the average of all correspondent pixel features over all N views. We flip each view from $\{I_n\}_{n=1}^{N}$ horizontally (i.e. along x-axis) to obtain $\{\bar{I}_n\}_{n=1}^{N}$. Then we compute the feature map $\hat{F}_n \in \mathbb{R}^{H\times W\times d}$ of $\bar{I}_n$ similarly to F_n. Finally we flip $\hat{F}_n$ horizontally to obtain $\bar{F}_n$. The per-vertex semantic flipped feature $\bar{\mathcal{F}}_v$ is computed from $\{\bar{F}_n\}_{n=1}^{N}$ similarly to $\mathcal{F}_v$.

Finally, we apply vertex-wise normalization to each feature in the feature set $\{\mathcal{F}_v\}_{v=1}^{|V|}$, i.e. $\mathcal{F}_v = \mathcal{F}_v / \|\mathcal{F}_v\|_2$. We denote the stack of all features along the vertex dimension as $\mathcal{F} \in \mathbb{R}^{|V|\times d}$. $\bar{\mathcal{F}}$ can be obtained analogously.

3.2. Symmetry-aware feature disentanglement

We use a lightweight auto-encoder as our disentanglement network. Specifically, we use an encoder E to obtain an intermediate feature $\mathcal{F}^{\text{mid}} \in \mathbb{R}^{|V|\times d}$ from the input $\mathcal{F}$. Further, we feed $\mathcal{F}^{\text{mid}}$ into a decoder D to reconstruct the input.

With intermediate features $\mathcal{F}^{\text{mid}}$ as input, we use the disentanglement steps depicted below:

- First, we perform per-vertex normalization on $\mathcal{F}^{\text{mid}}$, i.e. $\mathcal{F}_v^{\text{mid}} = \mathcal{F}_v^{\text{mid}} / \|\mathcal{F}_v^{\text{mid}}\|_2$, where $v \in \mathcal{V}$.
- To facilitate feature disentanglement, we project each normalized vertex descriptor $\mathcal{F}_v^{\text{mid}} \in \mathbb{R}^d$ using a global trainable orthonormal square matrix A as: $\mathcal{F}_v^{\text{proj}} = \mathcal{F}_v^{\text{mid}} A$.

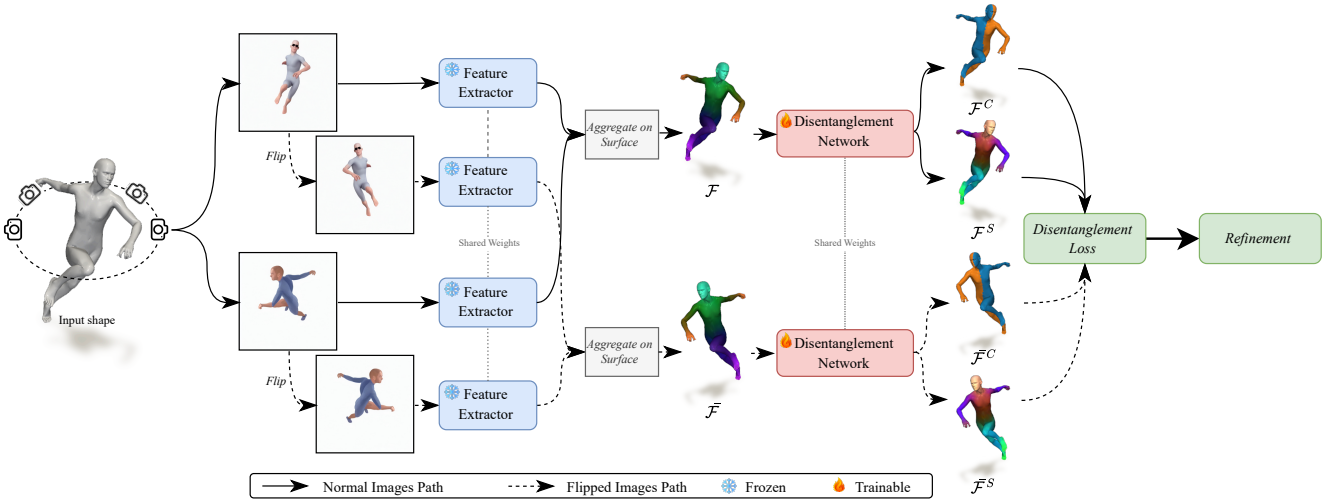

Figure 3. Overview of our method. As in Diff3F [15] and χ [53], we render different views from a mesh. For each view, 2D foundation models are used to extract features from it and its horizontally flipped counterpart. Shape descriptors are obtained by back-projecting those image features. The disentanglement network then learns to disentangle symmetry from symmetry-agnostic information using a combination of disentanglement losses. A refinement technique is applied on $\mathcal{F}^{\mathrm{C}}$ to increase robustness.

- After that, we separate the projected descriptor $\mathcal{F}_v^{\text{proj}}$ into two: $\mathcal{F}_v^{\text{proj}} := [\mathcal{F}_v^{\mathrm{C}}, \mathcal{F}_v^{\mathrm{S}}]$, where $\mathcal{F}_v^{\mathrm{C}} \in \mathbb{R}$ represents the symmetry-informative (chirality [53]) descriptor, and $\mathcal{F}_v^{\mathrm{S}} \in \mathbb{R}^{d-1}$ represents the symmetry-agnostic descriptor.
- Additionally, we apply normalization on the symmetry-agnostic descriptor $\mathcal{F}_v^S$ as: $\mathcal{F}_v^{\mathrm{S}} = \mathcal{F}_v^{\mathrm{S}} / \|\mathcal{F}_v^{\mathrm{S}}\|_2$.

Similarly, we follow the same steps to arrive at respective flipped symmetry informative and agnostic descriptors $\bar{\mathcal{F}}_v^{\mathrm{C}}, \bar{\mathcal{F}}_v^{\mathrm{S}}$ using flipped $\bar{\mathcal{F}}$ as input.

In order to make the disentanglement framework intrinsic symmetry aware, we use a combination of semantic and geometric losses to guide the training of the network (the trainable components are encoder E, decoder D, and the orthonormal matrix A), which we discuss next.

3.3. Unsupervised losses

With stacked per-vertex feature $\mathcal{F}$, we collect the disentangled symmetry-informative and symmetry-agnostic features from the network, and stack them along the vertex dimension to obtain $\mathcal{F}^{\mathrm{C}} \in [-1,1]^{|V|}$ and $\mathcal{F}^{\mathrm{S}} \in [-1,1]^{|V| \times (d-1)}$, respectively. Similarly, we compute $\bar{\mathcal{F}}^{\mathrm{C}}$ and $\bar{\mathcal{F}}^{\mathrm{S}}$ using $\bar{\mathcal{F}}$ as input.

Dissimilarity loss $\mathcal{L}_{\text{dis}}$ [53]. We apply a dissimilarity loss $\mathcal{L}_{\text{dis}}$ to push symmetry-informative descriptors of each vertex and its intrinsic symmetric counterpart away from each other:

$$\mathcal{L}_{\text{dis}} = -\frac{1}{\sqrt{|V|}} \|\mathcal{F}^{\mathrm{C}} - \bar{\mathcal{F}}^{\mathrm{C}}\|_2. \tag{1}$$

Similarity loss $\mathcal{L}_{\text{sim}}$. Further, we apply a similarity loss $\mathcal{L}_{\text{sim}}$ to make sure the symmetry-agnostic descriptors of each vertex and its intrinsic symmetric counterpart are close:

$$\mathcal{L}_{\text{sim}} = \frac{1}{\sqrt{|V|}} \|\mathcal{F}^{\mathrm{S}} - \bar{\mathcal{F}}^{\mathrm{S}}\|_F. \tag{2}$$

Reconstruction loss $\mathcal{L}_{\text{rec}}$ [53]. We want to avoid information loss of the intermediate feature $\mathcal{F}^{\text{mid}}$ given by the encoder E. Thus, we apply a reconstruction loss

$$\mathcal{L}_{\text{rec}} = \frac{1}{\sqrt{|V|}} \|[\mathcal{F}, \bar{\mathcal{F}}] - \mathrm{D}(\mathrm{E}([\mathcal{F}, \bar{\mathcal{F}}]))\|_F. \tag{3}$$

Here, $[\mathcal{F}, \bar{\mathcal{F}}]$ corresponds to the column stack of $\mathcal{F}$ and $\bar{\mathcal{F}}$.

Boundary loss $\mathcal{L}_{\text{bou}}$. To regularize the boundary to follow an approximately straight path along the surface manifold, we introduce a boundary loss $\mathcal{L}_{\text{bou}}$. This loss is built upon two components. First, we define the tangential cosine similarity $C_v(u,w)$ between the incoming edge (u,v) and outgoing edge (v,w) for given vertices $u,v,w \in V$. Second, we define $L(u,v,w)$ as the sum of squared differences of symmetry-informative features across these edges:

$$L(u,v,w) = \|\mathcal{F}_u^{\mathrm{C}} - \mathcal{F}_v^{\mathrm{C}}\|_2^2 + \|\mathcal{F}_v^{\mathrm{C}} - \mathcal{F}_w^{\mathrm{C}}\|_2^2, \tag{4}$$

where $\bar{L}(u,v,w)$ is defined analogously using $\bar{\mathcal{F}}$. To encourage colinear neighboring pairs $u,w \in V$ around each vertex $v \in V$ with similar symmetry-informative features $\mathcal{F}_v^{\mathrm{C}}$, we formulate the loss as

$$\mathcal{L}_{\text{bou}} = \frac{1}{|V|} \sum_{v \in V} \Big(\min_{(u,v,w) \in \mathcal{S}_v} \{L(u,v,w) - C_v(u,w)\} + \min_{(u,v,w) \in \mathcal{S}_v} \{\bar{L}(u,v,w) - C_v(u,w)\} \Big). \tag{5}$$

Here, $\mathcal{S}_v$ denotes the set of vertex tuples (u, v, w) such that $(u, v), (v, w) \in \bar{\mathcal{E}}$, where $\bar{\mathcal{E}}$ is the set of directed edges which is obtained from $\mathcal{E}$ by interpreting every undirected edge as two opposing directed edges. A more thorough explanation and visualization can be found in Sec. 9 in the supplementary material.

Consistency loss $\mathcal{L}_{\text{con}}$. To regularize the features, we enforce self-consistency by matching a shape to itself through its mirrored version. Let $W \in (\mathbb{R}^+)^{|V|\times|V|}$ with $W_{i,j} = \left(\mathcal{F}^{\text{C}}_{v_i} - \mathcal{F}^{\text{C}}_{v_j}\right)^2$ be the squared difference between $\mathcal{F}^{\text{C}}$ for different vertices $v_i, v_j \in V$, and let $C_c = \mathcal{F}^{\text{S}}(\mathcal{F}^{\text{S}})^\top \in [-1, 1]^{|V|\times|V|}$ be the similarity matrix between the $\mathcal{F}^{\text{S}}$ features. We then compute $\Pi \in [0, 1]^{|V|\times|V|}$ as the element-wise product after the min-max normalization to the $[0, 1]$ interval , i.e.

$$\Pi = \text{norm}(W) \odot \text{norm}(C_c). \tag{6}$$

Intuitively, Π represents a soft assignment of each vertex to its intrinsically symmetric counterpart. Consequently, Π^2 should ideally approximate the identity matrix, representing a mapping of the shape to itself via the symmetric reflection. $\bar{\Pi}$ is defined analogously with $\bar{\mathcal{F}}$.

This allows us to introduce the loss $\mathcal{L}_{\text{con}}$ as

$$\mathcal{L}_{\text{con}} = \frac{1}{|V|}\|[\mathbb{I}_{|V|}, \mathbb{I}_{|V|}] - [\Pi^2, \bar{\Pi}^2]\|_F, \tag{7}$$

where $\mathbb{I}_{|V|}$ is an identity matrix of size $|V| \times |V|$.

Overall loss. The overall loss is a weighted combination of those unsupervised losses:

$$\mathcal{L} = \mathcal{L}_{\text{dis}} + \lambda_1\mathcal{L}_{\text{sim}} + \lambda_2\mathcal{L}_{\text{rec}} + \lambda_3\mathcal{L}_{\text{bou}} + \lambda_4\mathcal{L}_{\text{con}}. \tag{8}$$

For choices of $\lambda_1, \lambda_2, \lambda_3$ and λ_4, we refer to the implementation details in Sec. 8 in the supplementary materials.

3.4. Symmetry aware feature refinement

As shown in Fig. 2, χ [53] produces incorrect small patches in symmetry-informative descriptors. To circumvent such undesirable solutions, we propose a simple yet effective refinement technique based upon a Markov Random Field (MRF) energy minimization formalism. In particular, we compute refined symmetry-informative features $\tilde{\mathcal{F}}^C$ by solving the following MRF problem

$$\min_{\tilde{\mathcal{F}}^C \in \{0,1\}^{|\mathcal{V}|}} \sum_{v\in\mathcal{V}} \theta_v(\tilde{\mathcal{F}}^C_v) + \sum_{(v,\bar{v})\in\mathcal{E}} \theta_{v\bar{v}}(\tilde{\mathcal{F}}^C_v, \tilde{\mathcal{F}}^C_{\bar{v}}). \tag{9}$$

We first do min-max normalization of $\mathcal{F}^C$ to transform it into $\mathcal{F}^C \in [0, 1]^{|\mathcal{V}|}$. Then we choose for the unary potentials $\theta_v(0) = \mathcal{F}^C_v$ and $\theta_v(1) = 1 - \mathcal{F}^C_v$ for all $v \in \mathcal{V}$. Furthermore, for the pairwise potentials, we use $\theta_{v\bar{v}}(0, 0) = \theta_{v\bar{v}}(1, 1) = 0$ and $\theta_{v\bar{v}}(0, 1) = \theta_{v\bar{v}}(1, 0) = 1$ for all edges $(v, \bar{v}) \in \mathcal{E}$. As an intuition, using these potentials essentially favors a refined symmetry-informative feature $\tilde{\mathcal{F}}^C$ which is as close as possible to the predicted feature $\mathcal{F}^C$ while having as little boundary as possible (which effectively reduces the patches). We efficiently solve (9) using graph cuts [5, 6, 14, 24].

Since this post processing technique does not rely on the design of our disentanglement framework, it could be applied on any coarse symmetry-informative features. We validate the general usage of this technique also on χ [53].

4. Experiments

We conduct different shape analysis experiments on our disentangled symmetry-informative and symmetry-agnostic descriptors pair to show the effectiveness of our proposed framework. In Sec. 4.1, we first introduce all datasets we use. In Sec. 4.2, intrinsic symmetry detection shows the superiority of our disentangled descriptors pair compared to several other shape descriptors. Then, in Sec. 4.3, the left/right classification experiment validates our symmetry-informative descriptor alone. We also introduce a new metric to quantify the robustness of symmetry-informative descriptors and the results prove the effectiveness of our refinement technique. The application of our features in shape matching is explored in Sec. 4.4. Lastly, to verify the necessity of our losses, an ablation study is done in Sec. 4.5.

4.1. Datasets

We use a recently proposed benchmark BeCoS [17] to generate a large scale dataset with rich annotations including inter-shape correspondences, per-shape self-correspondences (i.e. intrinsic symmetric pairs annotations), and left/right annotations. Additionally, we also experiment on several popular datasets: FAUST [4], SCAPE [1], SMAL [66] and TOSCA [7]. More details can be found below.

- The BeCoS [17] dataset consists of humanoid and four-legged animals with 20370/284/274 train/test/validation split, generated from 7 remeshed shape datasets, namely TOSCA [7], FAUST [4], SCAPE [1], KIDS [41], DT4D [28], SMAL [66] and SHREC'20 [16].
- BeCoS$_{\text{-h}}$ [17] consists of only humanoid shapes of BeCoS [17] with 9697/64/58 train/test/validation split.
- BeCoS$_{\text{-a}}$ [17] consists of only four-legged animals of BeCoS [17] with 10263/220/216 train/test/validation split.
- FAUST [4], SCAPE [1], SMAL [66] and TOSCA [7] datasets. We follow the original train/validation/test splits where available, and otherwise use the BeCoS [17] splits. Since these datasets lack left/right and self-correspondence annotations, we obtain them via BeCoS.

Train	BeCoS	BeCoS$_{\text{-h}}$		BeCoS$_{\text{-a}}$		FAUST		SMAL	
Test	BeCoS	BeCoS$_{\text{-h}}$	BeCoS$_{\text{-a}}$	BeCoS$_{\text{-h}}$	BeCoS$_{\text{-a}}$	FAUST	SCAPE	SMAL	TOSCA
DINO+SD + GT	0.098	0.091	0.100	0.091	0.100	0.037	0.061	0.125	0.059
Ours + GT	**0.059**	**0.058**	**0.093**	**0.073**	**0.061**	**0.025**	**0.032**	**0.070**	**0.050**
DINO+SD + χ	0.145	0.164	0.182	0.257	0.132	0.054	0.128	0.129	0.154
Ours	<u>0.101</u>	<u>0.114</u>	0.162	0.197	<u>0.087</u>	**0.042**	<u>0.073</u>	**0.075**	<u>0.115</u>
DINO+SD + (χ + refine)	0.120	0.122	<u>0.161</u>	<u>0.144</u>	0.120	0.049	0.103	0.135	0.124
Ours + refine	**0.078**	**0.077**	**0.146**	**0.108**	**0.078**	**0.042**	**0.049**	<u>0.076</u>	**0.084**

Table 1. Geodesic error of intrinsic symmetry detection using different features across datasets. The best and second-best results for each case are written in bold and underlined, respectively. Both our sym-agno feature and the refinement step lead to consistently lower errors.

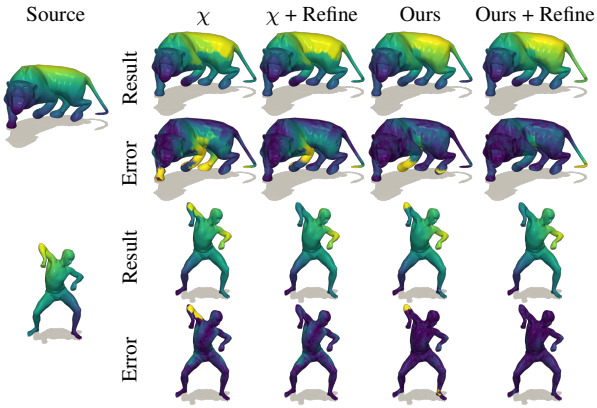

Figure 4. Results of intrinsic symmetry detection visualized as a mirrored matching of a shape to itself. Our method produces noticeably better quality than χ [53] both with and without refinement. Examples are chosen to reflect χ's failure modes.

4.2. Intrinsic symmetry detection

Metrics. For a given shape $\mathcal{X}$ represented with a vertex set $V \in \mathbb{R}^{|V|\times 3}$, intrinsic symmetry detection aims to match each vertex $v \in V$ to its intrinsic symmetric correspondent vertex $\bar{v} \in V$. Thus we use the average matching error as our evaluation metrics:

$$\text{err}_{\text{int}} = \frac{1}{|V|} \sum_{v \in V} \text{dist}(f(v), v_{\text{gt}}^{\text{sym}}), \tag{10}$$

where $f(v)$ is the predicted intrinsic symmetric vertex of v, $v_{\text{gt}}^{\text{sym}}$ is the ground truth intrinsic symmetry vertex of v, and $\text{dist}(\cdot,\cdot)$ is the geodesic distance normalized by the square root of the area of a shape following [22].

Baselines. We compare performance of our disentangled descriptors with other shape descriptors. For a detailed evaluation of our framework, we include three different baselines. All methods first cluster the vertices into two sets; the clustering strategy differs across methods as follows:

- DINO+SD+GT: We cluster vertices using ground-truth left/right annotations.
- DINO+SD+χ: We use χ [53] features for clustering.
- DINO+SD+(χ+refine): We use χ [53] features refined with Eq. 9 for clustering.

Then we use descriptors constructed following Diff3F [15] pipeline with concatenated DINO-V2 [33] and StableDiffusion [43] features to match between these two sets using cosine similarity.

Results. From Tab. 1, superiority of our disentangled descriptors is three-fold. First, comparison of DINO+SD+GT with Ours+GT (We use ground-truth for clustering and then $\mathcal{F}^{\text{S}}$ for matching) shows that our disentangled symmetry-agnostic descriptor $\mathcal{F}^{\text{S}}$ filters out intrinsic symmetry information and leads to better detection results, which validates the necessity of $\mathcal{F}^{\text{S}}$. Secondly, comparison of DINO+SD+χ with Ours (We use $\mathcal{F}^{\text{C}}$ for clustering and then $\mathcal{F}^{\text{S}}$ for matching) shows that without ground truth left/right annotations, our disentangled descriptors surpasses previous work χ [53], validating the usefulness of our framework in an unsupervised setting. Finally, the comparison of DINO+SD+χ with DINO+SD+(χ+refine) or the comparison of Ours with Ours + refine validates the effectiveness of our refinement technique onto symmetry-informative descriptors. Fig. 4 shows qualitative results consistent with our conclusions.

4.3. Left/right classification

Metrics. Given a set of shapes $X = \{\mathcal{X}_1, ..., \mathcal{X}_N\}$, where each shape $\mathcal{X}_n$ has a vertex set $V_{\mathcal{X}_n}$, we use the same metrics defined in χ [53] to compute the left/right classification accuracy:

$$\text{acc}_{\text{L/R}} = \max\{\text{hit}, 1 - \text{hit}\}, \tag{11}$$

where

$$\text{hit} = \frac{1}{N} \sum_{n=1}^{N} \frac{1}{|V_{\mathcal{X}_n}|} \sum_{v \in V_{\mathcal{X}_n}} \mathbb{1}(\text{sign}(\mathcal{F}_v^{\text{C}}) = \mathcal{F}_v^{gt}). \tag{12}$$

Train	BeCoS	BeCoS-h		BeCoS-a		FAUST		SMAL	
Test	BeCoS	BeCoS-h	BeCoS-a	BeCoS-h	BeCoS-a	FAUST	SCAPE	SMAL	TOSCA
Liu et al. [25]	79.98	79.83	80.46	79.83	80.46	90.45	80.84	75.71	72.88
χ [53]	92.51	94.25	84.97	92.16	91.63	95.56	96.14	96.49	88.14
Ours	91.65	94.13	85.09	92.18	90.96	**96.28**	96.90	96.72	91.12
χ [53] + refine	**93.04**	**94.69**	86.24	93.27	**92.05**	95.73	96.53	96.58	89.91
Ours + refine	92.03	94.49	**87.27**	**93.64**	91.17	96.26	**97.29**	**96.83**	**91.74**

Table 2. Left/right classification accuracy ($acc_{\text{L/R}}$) results. The best and second-best results for each case are written in bold and underlined, respectively. The refinement step consistently leads to higher accuracies, while our features result in similar accuracies to χ [53].

Train	Test	χ [53]	Ours	χ [53] + refine	Ours + refine
BeCoS	**BeCoS**	10.96	11.07	2.58	2.22
BeCoS-h	**BeCoS-h**	9.19	9.80	2.53	2.34
	BeCoS-a	21.51	21.94	3.86	2.90
BeCoS-a	**BeCoS-h**	15.25	15.33	3.06	2.73
	BeCoS-a	10.04	9.42	2.48	2.23
FAUST	**FAUST**	4.30	3.60	2.60	2.20
	SCAPE	6.88	6.88	2.63	2.38
SMAL	**SMAL**	4.63	4.50	2.00	2.00
	TOSCA	51.02	42.28	3.91	2.40

Table 3. Average number of connected components for different datasets and methods. The refinement step is able to effectively reduce the number of connected components of our solutions.

$\mathcal{F}_v^{\text{C}}$ is symmetry-informative descriptor for vertex $v \in V_{\mathcal{X}_n}$ of each shape $\mathcal{X}_n$, $\mathbb{1}$ is the indicator function, and $\mathcal{F}_v^{gt}$ is the ground truth left/right annotation of vertex $v \in V_{\mathcal{X}_n}$.

Baselines. We use disentangled descriptor $\mathcal{F}^{\text{C}}$ to perform left/right classification, and compare with descriptors given by χ [53], and refined descriptors from χ [53] using our proposed refinement technique. An axiomatic method Liu et al. [25] that extracts closed symmetric curves based on self-correspondence map prediction is also included.

Results. Tab. 2 summarizes left/right classification results on different datasets. Although our disentangled symmetry-informative feature $\mathcal{F}^{\text{C}}$ achieves comparable results with χ [53], we can conclude that our proposed refinement technique indeed improves symmetry-informative descriptors. In addition, for each shape in every dataset, we first do a 2-center clustering using symmetry-informative descriptor, and then compute the number of connected components of the shape, and finally average this number over the whole dataset. Results shown in Tab. 3 indeed validate the effectiveness of our proposed refinement technique again. Fig. 5 also indicates consistent conclusions qualitatively.

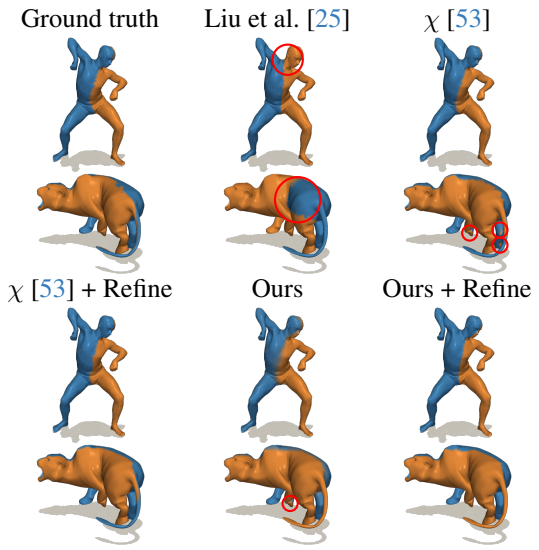

Figure 5. Qualitative results of left/right classification. Both χ [53] and our method achieve high accuracy. The refining step is able to reduce the number of incorrectly classified patches significantly.

4.4. Shape matching

Metrics. Following Diff3F [15] and χ [53], we use average matching error as metrics. For a source shape $\mathcal{X}$ and a target shape $\mathcal{Y}$, represented with vertex sets $V_{\mathcal{X}} \in \mathbb{R}^{|V_{\mathcal{X}}| \times 3}$ and $V_{\mathcal{Y}} \in \mathbb{R}^{|V_{\mathcal{Y}}| \times 3}$, respectively, the matching error is

$$err_{\text{mat}} = \frac{1}{|V_{\mathcal{X}}|} \sum_{v \in V_{\mathcal{X}}} \text{dist}(f(v) - y_{gt}), \tag{13}$$

where $f(v)$ is the predicted matching point of $v \in V_{\mathcal{X}}$ in $\mathcal{Y}$, $y_{gt} \in V_{\mathcal{Y}}$ is the ground truth corresponding point of v, and $\text{dist}(\cdot, \cdot)$ is the geodesic distance normalized by the square root of the area of $\mathcal{Y}$, following [22].

Baselines. As a matching baseline, we use pairwise cosine similarity of DINO+SD, DINO+SD concatenated to χ [53] and a refined version of χ, respectively.

Train	BeCoS	BeCoS$_{\text{-h}}$		BeCoS$_{\text{-a}}$		FAUST		SMAL	
Test	BeCoS	BeCoS$_{\text{-h}}$	BeCoS$_{\text{-a}}$	BeCoS$_{\text{-h}}$	BeCoS$_{\text{-a}}$	FAUST	SCAPE	SMAL	TOSCA
DINO+SD	0.163	0.146	0.222	0.146	0.222	0.163	0.167	0.131	0.146
DINO+SD + χ [53]	0.104	0.096	0.123	0.117	0.106	0.050	0.063	0.060	0.102
Ours	0.079	0.091	0.159	0.134	0.074	0.056	0.054	0.046	0.095
DINO+SD + χ [53] + refine	0.100	0.087	**0.122**	**0.094**	0.104	**0.048**	**0.048**	0.062	0.108
Ours + refine	**0.075**	**0.082**	0.157	0.110	**0.072**	0.058	0.049	**0.045**	**0.094**

Table 4. Geodesic error of shape matching using different features across datasets. Refinement generally enhances matching performance, with our complete framework achieving state-of-the-art results on several major datasets.

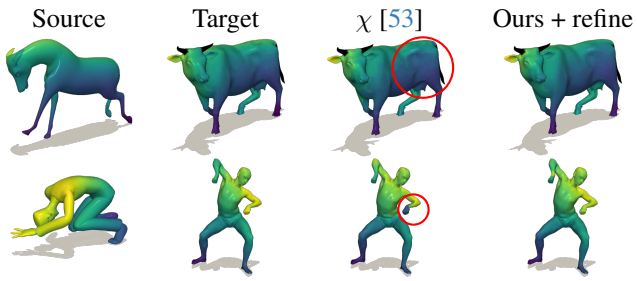

Figure 6. Shape matching results of two shape pairs from different categories, using χ [53] and our descriptors. Our method is able to create a smoother and more accurate matching.

Results. Tab. 4 summarizes shape matching results across multiple datasets and Fig. 6 visualizes qualitative results across multiple example shape pairs. We observe that refining the features improves matching quality for both χ [53] and our method across most datasets. When combining our new method with our proposed refinement, we are able to achieve significantly better matching performance than the baseline χ method on most datasets, as visible in Fig. 6.

4.5. Ablation study

We run ablation studies on FAUST [4] using left/right accuracy, intrinsic symmetry error, and shape matching error. We also compare with a modified χ [53], denoted $\bar{\chi}$, by adding orthogonality constraints to its MLP as in our method, trained with χ's losses. Tab. 5 shows that our loss combination yields a good balance of quality, informativeness, and symmetry invariance.

w/o	$\mathcal{L}_{\text{dis}}$	$\mathcal{L}_{\text{sim}}$	$\mathcal{L}_{\text{rec}}$	$\mathcal{L}_{\text{bou}}$	$\mathcal{L}_{\text{con}}$	$\bar{\chi}$	Full
$\text{acc}_{\text{L/R}}$	60.87	95.83	95.15	75.42	**97.14**	90.48	96.28
err_{sym}	0.292	0.420	0.052	0.116	0.071	0.132	**0.042**
err_{mat}	0.224	0.118	0.056	0.096	0.143	0.080	**0.056**

Table 5. The ablation study on FAUST validates our overall loss combination across all metrics.

5. Limitations

Although our proposed disentanglement framework shows remarkable performance in various shape analysis tasks on different datasets, it nevertheless has certain drawbacks. Similar to [53], our method requires mesh connectivity information to regularize the training, restricting its application to certain representations of 3D shapes. In addition, we observe inaccuracies in both symmetry-informative and symmetry-agnostic descriptors for vertices for which two different parts of a shape are in close proximity. Furthermore, our proposed refinement method depends on the predicted, continuous symmetry-informative features and essentially discretizes these into binary values. Consequently, slight deviations of the continuous features could lead to large deviations of the discretized features. Moreover, although our method is theoretically independent of a shape's genus, our current experimental evaluation is restricted to genus-zero surfaces.

6. Conclusion

In this work, we propose an intrinsic symmetry aware feature disentanglement network, which takes per vertex semantic shape descriptors as input and outputs robust disentangled symmetry-informative and symmetry-agnostic descriptors. To this end, we propose a selection of unsupervised losses, combined with a symmetry-informative feature refinement technique based on Markov Random Field energy minimization, to guide our framework to successfully disentangle symmetry-informative and symmetry-agnostic information. Various experiments, including intrinsic symmetry detection, left/right classification and shape matching, validate the effectiveness of the 3D shape descriptors obtained with our approach.

Acknowledgments

This work is supported by the ERC starting grant no. 101160648 (Harmony).

References

[1] Dragomir Anguelov, Praveen Srinivasan, Daphne Koller, Sebastian Thrun, Jim Rodgers, and James Davis. Scape: shape completion and animation of people. *SIGGRAPH*, pages 408–416, 2005. 5

[2] Dimitrije Antić, Georgios Paschalidis, Shashank Tripathi, Theo Gevers, Sai Kumar Dwivedi, and Dimitrios Tzionas. SDFit: 3D object pose and shape by fitting a morphable SDF to a single image. In *ICCV*, 2025. 1

[3] Mathieu Aubry, Ulrich Schlickewei, and Daniel Cremers. The wave kernel signature: A quantum mechanical approach to shape analysis. In *ICCV workshops*, pages 1626–1633. IEEE, 2011. 1

[4] Federica Bogo, Javier Romero, Matthew Loper, and Michael J Black. Faust: Dataset and evaluation for 3d mesh registration. In *CVPR*, pages 3794–3801, 2014. 5, 8

[5] Yuri Boykov and Vladimir Kolmogorov. An experimental comparison of min-cut/max-flow algorithms for energy minimization in vision. *TPAMI*, 2004. 5

[6] Yuri Boykov, Olga Veksler, and Ramin Zabih. Fast approximate energy minimization via graph cuts. *TPAMI*, 2002. 5

[7] Alexander Bronstein, Michael Bronstein, and Ron Kimmel. *Numerical Geometry of Non-Rigid Shapes*. Springer Publishing Company, Incorporated, 1 edition, 2008. 5

[8] Dongliang Cao, Paul Roetzer, and Florian Bernard. Unsupervised learning of robust spectral shape matching. *TOG*, 2023. 1

[9] Mathilde Caron, Hugo Touvron, Ishan Misra, Hervé Jégou, Julien Mairal, Piotr Bojanowski, and Armand Joulin. Emerging properties in self-supervised vision transformers. In *ICCV*, pages 9650–9660, 2021. 3

[10] Shoufa Chen, Peize Sun, Yibing Song, and Ping Luo. Diffusiondet: Diffusion model for object detection. In *ICCV*, pages 19830–19843, 2023. 1, 3

[11] Zhangquan Chen, Puhua Jiang, and Ruqi Huang. Dv-matcher: Deformation-based non-rigid point cloud matching guided by pre-trained visual features. In *CVPR*, pages 27264–27274, 2025. 1

[12] Luca Cosmo, Emanuele Rodola, Andrea Albarelli, Facundo Mémoli, and Daniel Cremers. Consistent partial matching of shape collections via sparse modeling. In *CGF*, pages 209–221. Wiley Online Library, 2017. 2

[13] Dale Decatur, Itai Lang, and Rana Hanocka. 3d highlighter: Localizing regions on 3d shapes via text descriptions. In *CVPR*, pages 20930–20939, 2023. 3

[14] Andrew Delong, Anton Osokin, Hossam N Isack, and Yuri Boykov. Fast approximate energy minimization with label costs. *IJCV*, 2012. 5

[15] Niladri Shekhar Dutt, Sanjeev Muralikrishnan, and Niloy J Mitra. Diffusion 3d features (diff3f): Decorating untextured shapes with distilled semantic features. In *CVPR*, pages 4494–4504, 2024. 1, 2, 3, 4, 6, 7

[16] Roberto M Dyke, Yu-Kun Lai, Paul L Rosin, Stefano Zappala, Seana Dykes, Daoliang Guo, Kun Li, Riccardo Marin, Simone Melzi, and Jingyu Yang. Shrec'20: Shape correspondence with non-isometric deformations. *Computers & Graphics*, 92:28–43, 2020. 5

[17] Viktoria Ehm, Nafie El Amrani, Yizheng Xie, Lennart Bastian, Maolin Gao, Weikang Wang, Lu Sang, Dongliang Cao, Tobias Weißberg, Zorah Lähner, Daniel Cremers, and Florian Bernard. Beyond Complete Shapes: A Benchmark for Quantitative Evaluation of 3D Shape Matching Algorithms. *CGF*, 2025. 5, 1

[18] Michael Fischer, Zhengqin Li, Thu Nguyen-Phuoc, Aljaz Bozic, Zhao Dong, Carl Marshall, and Tobias Ritschel. Nerf analogies: Example-based visual attribute transfer for nerfs. In *CVPR*, pages 4640–4650, 2024. 3

[19] Zhoujie Fu, Jiacheng Wei, Wenhao Shen, Chaoyue Song, Xiaofeng Yang, Fayao Liu, Xulei Yang, and Guosheng Lin. Sync4d: Video guided controllable dynamics for physics-based 4d generation. *arXiv preprint arXiv:2405.16849*, 2024. 1

[20] Eric Hedlin, Gopal Sharma, Shweta Mahajan, Xingzhe He, Hossam Isack, Abhishek Kar, Helge Rhodin, Andrea Tagliasacchi, and Kwang Moo Yi. Unsupervised keypoints from pretrained diffusion models. In *CVPR*, pages 22820–22830, 2024. 1, 3

[21] Vladimir G Kim, Yaron Lipman, Xiaobai Chen, and Thomas Funkhouser. Möbius transformations for global intrinsic symmetry analysis. In *CGF*, pages 1689–1700. Wiley Online Library, 2010. 2

[22] Vladimir G Kim, Yaron Lipman, and Thomas Funkhouser. Blended intrinsic maps. *TOG*, 30(4):1–12, 2011. 6, 7

[23] Sosuke Kobayashi, Eiichi Matsumoto, and Vincent Sitzmann. Decomposing nerf for editing via feature field distillation. *NeurIPS*, 35:23311–23330, 2022. 3

[24] Vladimir Kolmogorov and Ramin Zabin. What energy functions can be minimized via graph cuts? *TPAMI*, 2004. 5

[25] Tianqiang Liu, Vladimir G Kim, and Thomas Funkhouser. Finding surface correspondences using symmetry axis curves. In *CGF*, pages 1607–1616. Wiley Online Library, 2012. 2, 7, 3

[26] Yanxi Liu, Hagit Hel-Or, Craig S Kaplan, Luc Van Gool, et al. Computational symmetry in computer vision and computer graphics. *FTCGV*, 5(1–2):1–195, 2010. 2

[27] Changfeng Ma, Yinuo Chen, Pengxiao Guo, Jie Guo, Chongjun Wang, and Yanwen Guo. Symmetric shape-preserving autoencoder for unsupervised real scene point cloud completion. In *CVPR*, pages 13560–13569, 2023. 2

[28] Robin Magnet, Jing Ren, Olga Sorkine-Hornung, and Maks Ovsjanikov. Smooth non-rigid shape matching via effective dirichlet energy optimization. In *3DV*, pages 495–504. IEEE, 2022. 5

[29] Ben Mildenhall, Pratul P Srinivasan, Matthew Tancik, Jonathan T Barron, Ravi Ramamoorthi, and Ren Ng. Nerf: Representing scenes as neural radiance fields for view synthesis. *ACM*, 65(1):99–106, 2021. 3

[30] Niloy J Mitra, Leonidas J Guibas, and Mark Pauly. Symmetrization. *TOG*, 26(3):63–es, 2007. 2

[31] Niloy J Mitra, Mark Pauly, Michael Wand, and Duygu Ceylan. Symmetry in 3d geometry: Extraction and applications. In *CGF*, pages 1–23. Wiley Online Library, 2013. 2

[32] Rajendra Nagar and Shanmuganathan Raman. Fast and accurate intrinsic symmetry detection. In *ECCV*, pages 417–434, 2018. 2

[33] Maxime Oquab, Timothée Darcet, Théo Moutakanni, Huy Vo, Marc Szafraniec, Vasil Khalidov, Pierre Fernandez, Daniel Haziza, Francisco Massa, Alaaeldin El-Nouby, et al. Dinov2: Learning robust visual features without supervision. *arXiv:2304.07193*, 2023. 1, 3, 6

[34] Maks Ovsjanikov, Jian Sun, and Leonidas Guibas. Global intrinsic symmetries of shapes. In *CGF*, pages 1341–1348. Wiley Online Library, 2008. 2

[35] Yi-Ling Qiao, Lin Gao, Shu-Zhi Liu, Ligang Liu, Yu-Kun Lai, and Xilin Chen. Learning-based intrinsic reflectional symmetry detection. *TVCG*, 29(9):3799–3808, 2022. 3

[36] Alec Radford, Jong Wook Kim, Chris Hallacy, Aditya Ramesh, Gabriel Goh, Sandhini Agarwal, Girish Sastry, Amanda Askell, Pamela Mishkin, Jack Clark, et al. Learning transferable visual models from natural language supervision. In *ICML*, pages 8748–8763. PmLR, 2021. 1, 3

[37] Dan Raviv, Alexander M Bronstein, Michael M Bronstein, and Ron Kimmel. Symmetries of non-rigid shapes. In *ICCV*, pages 1–7. IEEE, 2007. 2

[38] Dan Raviv, Alexander M Bronstein, Michael M Bronstein, and Ron Kimmel. Full and partial symmetries of non-rigid shapes. *IJCV*, 89(1):18–39, 2010. 2

[39] Tammy Riklin-Raviv, Nir Sochen, and Nahum Kiryati. On symmetry, perspectivity, and level-set-based segmentation. *TPAMI*, 31(8):1458–1471, 2009. 2

[40] Sochen N. Riklin-Raviv T, Kiryati N. Segmentation by level sets and symmetry. In *CVPR*, pages 1015–1022. IEEE, 2006. 2

[41] Emanuele Rodola, Samuel Rota Bulo, Thomas Windheuser, Matthias Vestner, and Daniel Cremers. Dense non-rigid shape correspondence using random forests. In *CVPR*, pages 4177–4184, 2014. 5

[42] Paul Roetzer and Florian Bernard. Fast globally optimal and geometrically consistent 3d shape matching. In *ICCV*, 2025. 1

[43] Robin Rombach, Andreas Blattmann, Dominik Lorenz, Patrick Esser, and Björn Ommer. High-resolution image synthesis with latent diffusion models. In *CVPR*, pages 10684–10695, 2022. 1, 3, 6

[44] Taras Rumezhak, Oles Dobosevych, Rostyslav Hryniv, Vladyslav Selotkin, Volodymyr Karpiv, and Mykola Maksymenko. Towards realistic symmetry-based completion of previously unseen point clouds. In *ICCV*, pages 2542–2550, 2021. 2

[45] Samuele Salti, Federico Tombari, Luigi DI STEFANO, et al. Shot: Unique signatures of histograms for surface and texture description. *CVIU*, 125:251–264, 2014. 1

[46] Nicholas Sharp, Souhaib Attaiki, Keenan Crane, and Maks Ovsjanikov. Diffusionnet: Discretization agnostic learning on surfaces. *TOG*, 41(3):1–16, 2022. 1, 3

[47] Aleksandar Shtedritski, Christian Rupprecht, and Andrea Vedaldi. Shic: Shape-image correspondences with no keypoint supervision. In *ECCV*, pages 129–145. Springer, 2024. 1

[48] Jian Sun, Maks Ovsjanikov, and Leonidas Guibas. A concise and provably informative multi-scale signature based on heat diffusion. In *CGF*, pages 1383–1392. Wiley Online Library, 2009. 1

[49] Minhyuk Sung, Vladimir G Kim, Roland Angst, and Leonidas Guibas. Data-driven structural priors for shape completion. *TOG*, 34(6):1–11, 2015. 2

[50] Fabio Tosi, Pierluigi Zama Ramirez, and Matteo Poggi. Diffusion models for monocular depth estimation: Overcoming challenging conditions. In *ECCV*, pages 236–257. Springer, 2024. 1, 3

[51] Lukas Uzolas, Elmar Eisemann, and Petr Kellnhofer. Surface-aware distilled 3d semantic features. In *SIGGRAPH Asia*, 2025. 1, 3

[52] Hui Wang and Hui Huang. Group representation of global intrinsic symmetries. In *CGF*, pages 51–61. Wiley Online Library, 2017. 3

[53] Weikang Wang, Tobias Weißberg, Nafie El Amrani, and Florian Bernard. Symmetry understanding of 3d shapes via chirality disentanglement. In *ICCV*, 2025. 1, 2, 3, 4, 5, 6, 7, 8

[54] Yizheng Xie, Viktoria Ehm, Paul Roetzer, Nafie El Amrani, Maolin Gao, Florian Bernard, and Daniel Cremers. Echomatch: Partial-to-partial shape matching via correspondence reflection. In *CVPR*, 2025. 1

[55] Kai Xu, Hao Zhang, Andrea Tagliasacchi, Ligang Liu, Guo Li, Min Meng, and Yueshan Xiong. Partial intrinsic reflectional symmetry of 3d shapes. In *SIGGRAPH Asia*, pages 1–10, 2009. 2

[56] Kai Xu, Hao Zhang, Wei Jiang, Ramsay Dyer, Zhiquan Cheng, Ligang Liu, and Baoquan Chen. Multi-scale partial intrinsic symmetry detection. *TOG*, 31(6):1–11, 2012. 2

[57] Jianglong Ye, Naiyan Wang, and Xiaolong Wang. Featurenerf: Learning generalizable nerfs by distilling foundation models. In *ICCV*, pages 8962–8973, 2023. 3

[58] Yusuke Yoshiyasu, Eiichi Yoshida, Kazuhito Yokoi, and Ryusuke Sagawa. Symmetry-aware nonrigid matching of incomplete 3d surfaces. In *CVPR*, pages 4193–4200, 2014. 2

[59] Haocheng Yuan, Jing Xu, Hao Pan, Adrien Bousseau, Niloy J Mitra, and Changjian Li. Cadtalk: An algorithm and benchmark for semantic commenting of cad programs. In *CVPR*, pages 3753–3762, 2024. 1

[60] Junyi Zhang, Charles Herrmann, Junhwa Hur, Luisa Polania Cabrera, Varun Jampani, Deqing Sun, and Ming-Hsuan Yang. A tale of two features: Stable diffusion complements dino for zero-shot semantic correspondence. *NeurIPS*, 36: 45533–45547, 2023. 1, 3

[61] Junyi Zhang, Charles Herrmann, Junhwa Hur, Eric Chen, Varun Jampani, Deqing Sun, and Ming-Hsuan Yang. Telling left from right: Identifying geometry-aware semantic correspondence. In *CVPR*, pages 3076–3085, 2024. 1, 3

[62] Lvmin Zhang, Anyi Rao, and Maneesh Agrawala. Adding conditional control to text-to-image diffusion models. In *ICCV*, 2023. 3

[63] Zhiyuan Zhang, KangKang Yin, and Kelvin WC Foong. Symmetry robust descriptor for non-rigid surface matching. In *CGF*, pages 355–362. Wiley Online Library, 2013. 2

[64] Qian Zheng, Zhuming Hao, Hui Huang, Kai Xu, Hao Zhang, Daniel Cohen-Or, and Baoquan Chen. Skeleton-intrinsic symmetrization of shapes. In *CGF*, pages 275–286. Wiley Online Library, 2015. 2

[65] Junzhe Zhu, Yuanchen Ju, Junyi Zhang, Muhan Wang, Zhecheng Yuan, Kaizhe Hu, and Huazhe Xu. Densematcher: Learning 3d semantic correspondence for category-level manipulation from a single demo. *ICLR*, 2025. 1, 3

[66] Silvia Zuffi, Angjoo Kanazawa, David Jacobs, and Michael J. Black. 3D menagerie: Modeling the 3D shape and pose of animals. In *CVPR*, 2017. 5