G-MSM: Unsupervised Multi-Shape Matching with Graph-based Affinity Priors

Marvin Eisenberger[1], Aysim Toker[1], Laura Leal-Taixé[1,†], Daniel Cremers[1,2,3]
[1]TU Munich, [2]Munich Center for Machine Learning, [3]University of Oxford

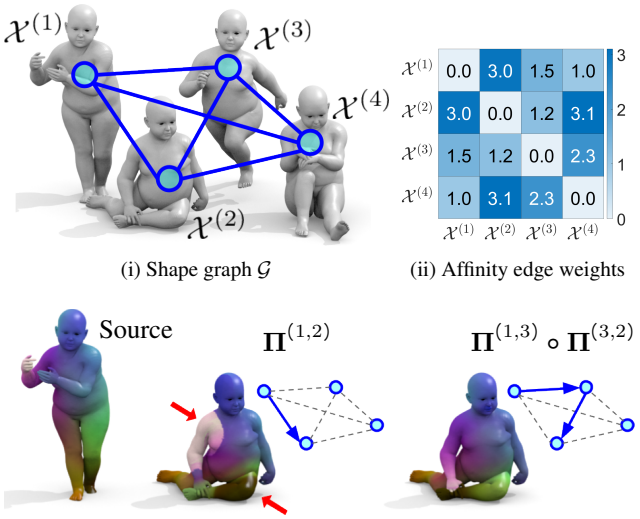
Abstract

*We present G-MSM (**G**raph-based **M**ulti-**S**hape **M**atching), a novel unsupervised learning approach for non-rigid shape correspondence. Rather than treating a collection of input poses as an unordered set of samples, we explicitly model the underlying shape data manifold. To this end, we propose an adaptive multi-shape matching architecture that constructs an affinity graph on a given set of training shapes in a self-supervised manner. The key idea is to combine putative, pairwise correspondences by propagating maps along shortest paths in the underlying shape graph. During training, we enforce cycle-consistency between such optimal paths and the pairwise matches which enables our model to learn topology-aware shape priors. We explore different classes of shape graphs and recover specific settings, like template-based matching (star graph) or learnable ranking/sorting (TSP graph), as special cases in our framework. Finally, we demonstrate state-of-the-art performance on several recent shape correspondence benchmarks, including real-world 3D scan meshes with topological noise and challenging inter-class pairs.*[1]

1. Introduction

Shape matching of non-rigid object categories is a central problem in 3D computer vision and graphics that has been studied extensively over the last few years. Especially in recent times, there is a growing demand for such algorithms as 3D reconstruction techniques and affordable scanning devices become increasingly powerful and broadly available. Classical shape correspondence approaches devise axiomatic algorithms that make specific assumptions about the resulting maps, such as near-isometry, area preservation, approximate rigidity, bounded distortion, or commutativity with the intrinsic Laplacian. In contrast, real-world scan meshes are often subject to various types of noise, including topological changes [16, 33], partial views [2], general non-isometric deformations [17, 65], objects in clutter [12], and varying data representations [57]. In this work, we address several of the aforementioned challenges and demonstrate that our proposed method achieves improved stability for a number of 3D scan mesh datasets.

The majority of existing deep learning methods for shape matching [2, 15, 20, 21, 24, 39, 51, 56] treat a given set of meshes as an unstructured collection of poses. During training, random pairs of shapes are sampled for which a neural network is queried and a pairwise matching loss is minimized. While this approach is straightforward, it often fails to recognize commonalities and context-dependent patterns



(i) Shape graph $\mathcal{G}$ (ii) Affinity edge weights

(iii) Putative correspondences $\mathbf{\Pi}^{(1,2)}$ vs multi-matching $\mathbf{\Pi}^{(1,3)} \circ \mathbf{\Pi}^{(3,2)}$

Figure 1. For a given collection of 3D meshes $\{\mathcal{X}^{(i)} | 1 \leq i \leq N\}$, (i) our method constructs, in a fully unsupervised manner, a shape graph $\mathcal{G}$ which approximates the underlying shape data manifold. (ii) Its edge weights (affinity scores) are derived from a putative pairwise correspondence loss signal. (iii) During training, we enforce cycle-consistency by propagating maps along shortest paths in the graph $\mathcal{G}$. As shown for the sample pair above $(\mathcal{X}^{(1)}, \mathcal{X}^{(2)})$, the resulting multi-matching $\mathbf{\Pi}^{(1,3)} \circ \mathbf{\Pi}^{(3,2)}$ is significantly more accurate than the pairwise map $\mathbf{\Pi}^{(1,2)}$.

[†] Currently at NVIDIA

[1]Our implementation is available under the following link: https://github.com/marvin-eisenberger/gmsm-matching

which only emerge from analyzing the shape collection as a whole. *Not all samples of a shape collection are created equal.* In most cases, some pairs of poses are much closer than others. Maps between similar geometries are inherently correlated and convey relevant clues to one another. This is particularly relevant for challenging real-world scenarios, where such redundancies can help disambiguate noisy geometries, non-isometric deformations, and topological changes. The most common approach of existing multi-matching methods is to learn a canonical embedding per pose, either in the spatial [10] or Laplace-Beltrami frequency domain [26,28]. This incentivizes the resulting matches to be consistent under concatenation. However, such approaches are in practice still trained in a fully pairwise manner for ease of training. Furthermore, relying on canonical embeddings can lead to limited generalization for unseen test poses. Concrete approaches often assume a specific mesh resolution and nearly-isometric poses [10], or require an additional fine-tuning optimization at test time [26, Sec. 5].

Rather than interpreting a given training set as a random, unstructured collection of shapes, our approach explicitly models the underlying shape manifold. To this end, we define an affinity graph $\mathcal{G}$ on the set of input shapes whose edge weights (i.e. affinity scores) are informed by the outputs of a pairwise matching module. We then devise a novel adaptive multi-matching architecture that propagates matches along shortest paths in the underlying shape graph $\mathcal{G}$. The resulting maps are topology-aware, i.e., informed by geometries from the whole shape collection. An example is shown in Figure 1, where the multi-matching $\mathbf{\Pi}^{(1,3)} \circ \mathbf{\Pi}^{(3,2)}$ obtained by our approach is significantly more accurate than the naive, pairwise map $\mathbf{\Pi}^{(1,2)}$. During training, we promote cycle-consistency of shortest paths in the shape graph. In summary, our contributions are as follows:

1. Introduce the notion of an edge-weighted, undirected shape graph $\mathcal{G}$ to approximate the underlying data manifold for an unordered collection of 3D meshes.

2. Propose a novel, adaptive multi-shape matching approach that enforces cycle-consistency for optimal paths in the shape graph $\mathcal{G}$ in a self-supervised manner.

3. Demonstrate state-of-the-art performance for a range of challenging non-rigid matching tasks, including non-isometric matching due to topological noise [16, 33] and inter-class pairs [17, 65].

2. Related work

Axiomatic correspondence methods Shape matching is an extensively studied topic with a variety of different approaches and methodologies. We summarize references relevant to our approach here and refer to recent surveys [54, 60] for a more complete picture. Classical methods for non-rigid matching often devise optimization-based approaches that minimize some type of distortion metric [7, 18, 49, 62, 64]. A common prerequisite is the extraction of hand-crafted local descriptors that are approximately preserved under non-rigid shape deformations. Common definitions include histogram-based statistics [59] or fully intrinsic features based on the eigenfunctions of the Laplace-Beltrami operator [3, 53, 58]. Over the last few years, functional maps [44] have become a central paradigm in shape matching. The core idea is to reframe the pairwise matching task from functions (points to points) to functionals (functions to functions). There are several extensions of the original framework to allow for partial matching [35, 36, 50], orientation preservation [48], iterative map upsampling [19, 42] and conformal maps [14]. Our approach utilizes functional maps as a fundamental building block within the differentiable matching layer.

Learning-based methods More recently, several approaches emerged that aim at extending the power of deep feature learning to deformable 3D shapes. Many such methods fall under the umbrella term 'geometric deep learning' [9], with analogous applications on different classes of non-Euclidean data like graphs or general manifold data. One class of approaches are charting-based methods [6,40,43,46,52] which imitate convolutions in Euclidean space with parameterized, intrinsic patch operators. Likewise, [57] proposed a learnable feature refinement module based on intrinsic heat diffusion.

The pioneering work of [34] proposes a differentiable matching layer based on functional maps [44], in combination with a deep feature extractor with several consecutive ResNet layers [25]. Numerous extensions of this paradigm were proposed over the last few years to allow for unsupervised loss functions [24, 51], learnable basis functions [39], point cloud feature extractors [15, 26, 56] or partial data [2]. Similarly, DeepShells [21] learns functional maps in an end-to-end trainable, hierarchical multi-scale pipeline. We adapt parts of its differentiable matching layer in our network. Other approaches [4, 23, 38] learn correspondences for a specific class of deformable objects by including additional domain knowledge, like a deformable human model [37]. Finally, [20] jointly learns to predict correspondences and a smooth interpolation between pairs of shapes.

Multi-shape matching Classical axiomatic multi-shape matching approaches devise optimization-based pipelines that enforce cycle-consistent maps. Specific solutions include semidefinite programming [27], convex relaxations of the corresponding quadratic assignment problem [30], graph cuts [55], as well as evolutionary game theory [11]. Such optimization-based approaches are computationally costly and therefore limited to matching sparse landmarks. Furthermore, there are a number of optimization frame-

works that compute synchronized, cycle-consistent functional maps [22, 28, 29]. Notably, such approaches are often limited to nearly-isometric poses [22, 29] or require high-quality initializations [28]. More recent learning-based approaches promote cycle-consistency by predicting a canonical embedding for each observed pose [10, 26]. However, obtaining stable embeddings is often difficult when generalizing to unseen test poses. Moreover, such approaches assume a specific mesh resolution and nearly-isometric poses [10] or require an additional fine-tuning optimization at test time to obtain canonical embeddings [26, Sec. 5].

3. Method

3.1. Problem formulation

In the following, we consider a collection of 3D shapes $\mathcal{S} = \{\mathcal{X}^{(1)}, \dots, \mathcal{X}^{(N)}\}$ from non-rigidly deformable shape categories. Each such shape $\mathcal{X}^{(i)}$ is a discretized approximation of a 2D Riemannian manifold, embedded in $\mathbb{R}^3$. Specifically, we define $\mathcal{X}^{(i)} = \big(\mathbf{V}^{(i)}, \mathbf{T}^{(i)}\big)$, where $\mathbf{V}^{(i)} \in \mathbb{R}^{m\times 3}$ and $\mathbf{T}^{(i)} \subset \mathbf{V}^{(i)}\times\mathbf{V}^{(i)}\times\mathbf{V}^{(i)}$ are sets of vertices and triangular faces, respectively. The goal is then to construct an algorithm that computes dense correspondence mappings $\mathbf{\Pi}^{(i,j)}$ between any two surfaces $\mathcal{X}^{(i)}$ and $\mathcal{X}^{(j)}$ from the shape collection $\mathcal{S}$. Specifically, such correspondences are represented by sparse assignment matrices $\mathbf{\Pi}^{(i,j)} \in \{0,1\}^{m\times n}$, where $\mathbf{\Pi}^{(i,j)}\mathbf{1}_n = \mathbf{1}_m$ and $\mathbf{\Pi}^{(i,j)}_{i',j'} = 1$ indicates a match between the i'-th vertex of $\mathcal{X}^{(i)}$ and the j'-th vertex of $\mathcal{X}^{(j)}$.

Scope Our method is unsupervised and thereby requires no additional inputs, like landmark annotations or ground-truth correspondences, beyond the raw input geometries $\mathcal{X}^{(i)}$. Following similar approaches in this line of work [4, 20, 39, 56], we assume that the shapes $\mathcal{X}^{(1)}, \dots, \mathcal{X}^{(N)}$ have an approximately canonical orientation. In the literature, this setting is commonly referred to as 'weakly supervised', see [56] and later [20]. Existing approaches often make additional assumptions about the input data $\mathcal{S}$, focusing on nearly-isometric correspondences [34, 51], maps with bounded distortion [20, 21] or partial views of the same non-rigid object [2, 50]. Others specialize in distinct classes of shapes like deformable human bodies [4, 23, 38]. In contrast, we demonstrate in our experiments that our proposed multi-matching approach excels at a broad range of challenging settings, including non-isometric pairs, poses with topological noise from self-intersections, and inter-class matching.

3.2. Network architecture

We now define the neural network architecture that forms the basis of our proposed approach. It consists of three separate components **I**-**III**, see also Figure 2 for an overview. The first two modules are standard components found in most learning-based shape matching approaches [2, 15, 20, 21, 24, 39, 51, 56], namely a learnable feature backbone **I** and a differentiable, pairwise matching layer **II**. We briefly outline these here and provide additional details in Appendix A.2. The multi-matching architecture **III** is introduced in Section 3.3.

Feature extractor The first component **I** of our model is a standard, learnable feature extraction backbone for representation learning, defined as

$$\Phi_{\text{feat}} : \mathcal{X}^{(i)} \;\mapsto\; \mathbf{F}^{(i)} \in \mathbb{R}^{m\times l}. \tag{1}$$

For a given input shape $\mathcal{X}^{(i)} = \big(\mathbf{V}^{(i)}, \mathbf{T}^{(i)}\big)$, the mapping Φ_{feat} produces an l-dimensional feature embedding $\mathbf{F}^{(i)}$ per vertex $\mathbf{V}^{(i)} \in \mathbb{R}^{m\times 3}$. While other choices are possible, we base Φ_{feat} on the off-the-shelf feature backbone DiffusionNet [57]. This network refines features via intrinsic heat-diffusion operators. Such operators are agnostic to the input discretization, thereby extremely robust to varying mesh resolutions and sampling densities. At the same time, it is computationally lightweight. For more details on the choice of backbone, see Appendix A.2.

Pairwise matching The second component **II** of our network is a differentiable, multi-scale matching scheme based on the recent pairwise shape matching method DeepShells [21]. The basis of this approach is the energy function

$$E_{\text{match}}\big(\mathbf{F}, \mathbf{G}; \tilde{\mathbf{\Pi}}\big) := \sum_{i'=1}^{m}\sum_{j'=1}^{n} \tilde{\mathbf{\Pi}}_{i',j'} \big\|\mathbf{F}_{i'} - \mathbf{G}_{j'}\big\|_2^2 \tag{2}$$

which has its roots in the theory of optimal transport. For a given transport plan $\tilde{\mathbf{\Pi}} \in [0,1]^{m\times n}$, the energy E_{match} specifies the distance between the discrete measures associated with two arbitrary l-dimensional feature embeddings $\mathbf{F} = \big(\mathbf{F}_1, \dots, \mathbf{F}_m\big) \in \mathbb{R}^{m\times l}$ and $\mathbf{G} = \big(\mathbf{G}_1, \dots, \mathbf{G}_n\big) \in \mathbb{R}^{n\times l}$. Taking the minimum over all possible transport plans $\arg\min_{\tilde{\mathbf{\Pi}}} E_{\text{match}}\big(\mathbf{F}, \mathbf{G}; \tilde{\mathbf{\Pi}}\big)$ results in the Kantorovich formulation of optimal transport [45, 63]. Following the approach described in [21], we obtain a multi-scale shape matching scheme that minimizes Equation (2) in an iterative optimization. For a given pair of shapes $\mathcal{X}^{(i)}$ and $\mathcal{X}^{(j)}$, this scheme defines a mapping

$$\Phi_{\text{match}} : \big(\mathbf{F}^{(i)}, \mathbf{F}^{(j)}\big) \;\mapsto\; \big(\mathbf{\Pi}^{(i,j)}, \mathbf{V}^{(i,j)}, \ell^{(i,j)}_{\text{match}}\big). \tag{3}$$

We provide further details on the exact update steps of the optimization scheme Φ_{match} in Appendix A.2. From a high-level perspective, Φ_{match} is defined as a deterministic, differentiable function that takes local feature encodings

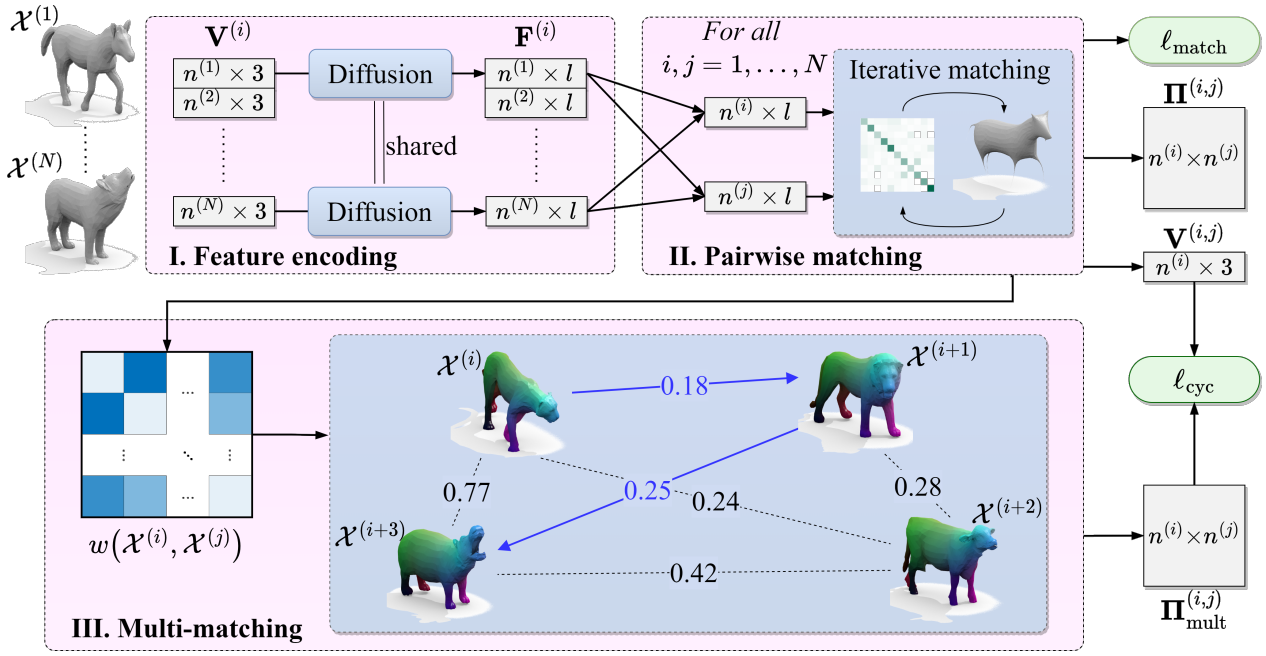


Figure 2. **Pipeline overview.** For a collection of shapes $\mathcal{S} = \{\mathcal{X}^{(1)}, \dots, \mathcal{X}^{(N)}\}$, **I.** feature embeddings are extracted with DiffusionNet [57] and **II.** pairwise correspondences $\mathbf{\Pi}^{(i,j)}$ are predicted via an iterative, differentiable matching layer [21]. **III.** The pairwise matches are utilized to construct a shape graph $\mathcal{G} = (\mathcal{S}, w)$ with affinity edge weights $w(\mathcal{X}^{(i)}, \mathcal{X}^{(j)}) \geq 0$. During training, we minimize the pairwise matching loss $\ell_{\text{match}}^{(i,j)}$, as well as the cycle consistency loss $\ell_{\text{cyc}}^{(i,j)}$ between the pairwise registrations $\mathbf{V}^{(i,j)}$ and multi-matches $\mathbf{\Pi}_{\text{mult}}^{(i,j)}$.

$\mathbf{F}^{(i)} \in \mathbb{R}^{m \times l}$ and $\mathbf{F}^{(j)} \in \mathbb{R}^{n \times l}$ as input and predicts a set of correspondences $\mathbf{\Pi}^{(i,j)} \in \{0,1\}^{m \times n}$. Additionally, Φ_{match} outputs a deformed embedding $\mathbf{V}^{(i,j)} \in \mathbb{R}^{m \times 3}$ of the vertices of $\mathcal{X}^{(i)}$. These coordinates specify a registered version of the first input shape $\mathcal{X}^{(i)}$ that closely aligns with the pose of the second input shape $\mathcal{X}^{(j)}$. The third output $\ell_{\text{match}}^{(i,j)} > 0$ is a training loss signal.

3.3. Graph-based multi-shape matching

Shape graph We now provide details on our multi-shape matching architecture **III**. To this end, we start by defining an affinity graph

$$\mathcal{G} := (\mathcal{S}, w), \quad \text{with} \quad w : \mathcal{S} \times \mathcal{S} \to [0, \infty] \tag{4}$$

on the set of training shapes $\mathcal{S} = \{\mathcal{X}^{(1)}, \dots, \mathcal{X}^{(N)}\}$, see **III** in Figure 2 for a visualization. W.l.o.g., we construct $\mathcal{G}$ as a complete graph (i.e. undirected, fully connected), where a missing edge between $\mathcal{X}^{(i)}$ and $\mathcal{X}^{(j)}$ can be specified equivalently by setting the corresponding edge weight to $w(\mathcal{X}^{(i)}, \mathcal{X}^{(j)}) = \infty$.

We define the pairwise edge weights $w(\mathcal{X}^{(i)}, \mathcal{X}^{(j)}) \in [0, \infty]$ such that they represent affinity scores between pairs of shapes $\mathcal{X}^{(i)}$ and $\mathcal{X}^{(j)}$. By convention, small values $w(\mathcal{X}^{(i)}, \mathcal{X}^{(j)}) \approx 0$ reflect that $\mathcal{X}^{(i)}$ and $\mathcal{X}^{(j)}$ have a comparably similar geometric structure. Since our method is fully unsupervised, we have no a priori knowledge of such affinities and thereby have to infer them directly from the geometries $\mathcal{X}^{(i)}$ and $\mathcal{X}^{(j)}$. To this effect, we propose a simple heuristic for a given pair of shapes $\mathcal{X}^{(i)} = (\mathbf{V}^{(i)}, \mathbf{T}^{(i)})$ and $\mathcal{X}^{(j)} = (\mathbf{V}^{(j)}, \mathbf{T}^{(j)})$ and define the (symmetric) affinity score w as

$$w(\mathcal{X}^{(i)}, \mathcal{X}^{(j)}) := \min\Big\{ E_{\text{match}}(\mathbf{V}^{(i,j)}, \mathbf{V}^{(j)}; \mathbf{\Pi}^{(i,j)}), \\ E_{\text{match}}(\mathbf{V}^{(j,i)}, \mathbf{V}^{(i)}; \mathbf{\Pi}^{(j,i)}) \Big\}. \tag{5}$$

In this context, E_{match} is the (self-supervised) matching energy defined in Equation (2), while $\mathbf{\Pi}^{(i,j)}$ and $\mathbf{V}^{(i,j)}$ are the putative correspondences and registrations produced by Equation (3), respectively. The intuition behind this choice of edge weights w is that a small matching energy E_{match} implies a high correspondence accuracy, which is in turn indicative of a high geometric similarity between the input poses $\mathcal{X}^{(i)}$ and $\mathcal{X}^{(j)}$.

Multi-matching Since we define the edge weights w according to the self-supervised matching score E_{match}, small weights $w(\mathcal{X}^{(i)}, \mathcal{X}^{(j)})$ generally correlate with a high correspondence accuracy of $\mathbf{\Pi}^{(i,j)}$. Based on this assumption,

we obtain multi-shape matches from the putative correspondences $\mathbf{\Pi}^{(i,j)}$ via the following expression

$$\big(i, s_1, \dots, s_{M-1}, j\big) \;:=\; \text{Dijkstra}\big(\mathcal{X}^{(i)}, \mathcal{X}^{(j)}; \mathcal{G}\big) \quad (6a)$$

$$\mathbf{\Pi}^{(i,j)}_{\text{mult}} := \mathbf{\Pi}^{(i,s_1)} \circ \mathbf{\Pi}^{(s_1,s_2)} \circ \dots \circ \mathbf{\Pi}^{(s_{M-1},j)}. \quad (6b)$$

Rather than matching a pair of shapes $\mathcal{X}^{(i)}$ and $\mathcal{X}^{(j)}$ directly, the multi-shape correspondence maps are passed along shortest paths $\mathcal{X}^{(i)}, \mathcal{X}^{(s_1)}, \dots, \mathcal{X}^{(s_{M-1})}, \mathcal{X}^{(j)}$ in the graph $\mathcal{G}$. The approach thereby favors edges with a close affinity, i.e., a small pairwise matching cost $w\big(\mathcal{X}^{(s_k)}, \mathcal{X}^{(s_{k+1})}\big)$. In our experiments, we demonstrate that this simple heuristic yields significant empirical improvements for a broad range of non-rigid matching tasks.

In practice, we utilize the multi-matching from Equation (6) for two distinct use-cases: For once, we can directly query the improved maps $\mathbf{\Pi}^{(i,j)}_{\text{mult}}$ at test time. Additionally, we promote cycle-consistency during training via the following loss term

$$\ell^{(i,j)}_{\text{cyc}} := E_{\text{match}}\big(\mathbf{V}^{(i,j)}, \mathbf{V}^{(j)}; \mathbf{\Pi}^{(i,j)}_{\text{mult}}\big). \quad (7)$$

This loss $\ell^{(i,j)}_{\text{cyc}}$ imposes a soft penalty on inconsistencies between the registration $\mathbf{V}^{(i,j)}$ produced by the pairwise matching module Φ_{match}, and the multi-shape correspondences $\mathbf{\Pi}^{(i,j)}_{\text{mult}}$ from Equation (6). As before, E_{match} is the matching energy defined in Equation (2).

3.4. Training protocol

The overall loss function that we minimize during training consists of two individual components

$$\ell := \mathbb{E}_{\mathcal{X}^{(i)}, \mathcal{X}^{(j)} \sim \mathcal{S}} \Big[\ell^{(i,j)}_{\text{match}} + \lambda_{\text{cyc}} \ell^{(i,j)}_{\text{cyc}} \Big]. \quad (8)$$

Our complete pipeline is depicted in Figure 2. The whole network is trained end-to-end. In each training iteration, the backbone **I** and pairwise matching module **II** are queried in sequence to produce a pairwise matching for a pair of shapes $\mathcal{X}^{(i)}$ and $\mathcal{X}^{(j)}$. The shape graph module **III** then produces the cycle-consistency loss $\ell^{(i,j)}_{\text{cyc}}$. The shape graph $\mathcal{G}$ is updated regularly after a fixed number of epochs, taking into account the pairwise matches $\mathbf{\Pi}^{(i,j)}$ for all $i, j = 1, \dots, N$. For more details on the training schedule and choices of hyperparameters, see Appendix A.1.

4. Experiments

We provide various benchmark evaluations for non-rigid shape matching. We consider classical nearly-isometric datasets in Section 4.1, as well as more specialized benchmarks for matching with topological changes in Section 4.2 and inter-class pairs in Section 4.3. In Section 4.4, we compare different types of shape graph topologies. In Section 4.5, we provide an ablation study of our model, assessing the significance of individual network components.

Method	FAUST	SCAPE	F on S	S on F	SUR	SH'19
UFM [24]	5.7	10.0	12.0	9.3	9.2	15.5
SURFM [51]	7.4	6.1	19.0	23.0	38.9	37.7
WFM [56]	1.9	4.9	8.0	4.3	38.5	15.0
DiffNet [57]	1.9	2.6	2.7	1.9	8.8	11.0
DS [21]	1.7	2.5	5.4	2.7	2.7	12.1
NM [20]	**1.5**	4.0	6.7	2.0	9.7	2.8
IsoMuSh [22]	4.4	5.6	–	–	4.8	24.6
CZO [28]	2.2	2.5	–	–	2.2	6.3
UDM [10]	**1.5**	2.0	3.2	3.2	3.1	22.8
SyNoRiM [26]	7.9	9.5	21.9	24.6	12.7	7.5
Ours w/o **III**	1.7	3.3	4.2	1.7	8.1	6.2
Ours	**1.5**	**1.8**	**2.1**	**1.5**	**2.1**	**2.7**

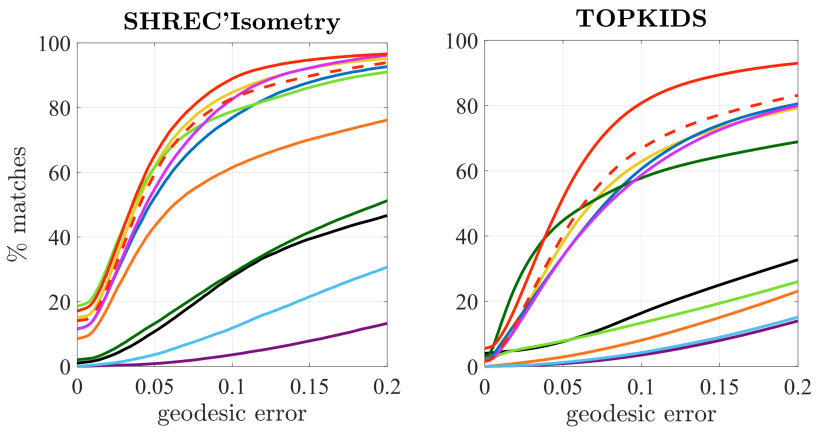
Table 1. **Nearly isometric matching.** A quantitative comparison on four nearly-isometric human shape benchmarks, FAUST [5], SCAPE [1], SURREAL [61] and SHREC'19 [41]. Following prior work [15, 56, 57], we additionally show generalization results when training on FAUST and testing on SCAPE (F on S), and vice versa. We consider both standard, pairwise baselines [20, 21, 24, 51, 56, 57] and multi-matching approaches [10, 22, 26, 28].

Baselines We compare G-MSM to existing deep learning approaches for unsupervised, deformable 3D shape correspondence. To this end, we consider both standard pairwise matching [20, 21, 24, 51, 56, 57] and multi-matching approaches [10, 26]. Since there are, to date, only very few learning-based multi-matching approaches, we additionally include Consistent ZoomOut [28] and IsoMuSh [22] as recent axiomatic multi-matching approaches.

Evaluation For each experimental setting, we report the mean geodesic correspondence error over all pairs of a given test set category. All evaluations are performed in accordance with the standard Princeton benchmark protocol [31].

4.1. Nearly isometric matching

Datasets We evaluate our method on four classical, nearly-isometric datasets. FAUST [5] contains 10 humans in 10 different poses each and SCAPE [1] contains 71 diverse poses of the same individual. We follow the standard benchmark protocol from existing work [15, 56, 57]. Specifically, we consider the more challenging remeshed geometries from [48] to avoid overfitting to a particular triangulation. SURREAL [61] consists of synthetic SMPL [37] meshes fitted to raw 3D motion capture data. The last benchmark, which is the most challenging among the four, is SHREC'19 Connectivity [41]. It contains human shapes in different poses with significantly varying sampling density and quality, as well as a small number of non-isometric poses.



Method	SH'Iso	KIDS
UFM [24]	13.4	38.5
SURFM [51]	45.6	48.6
WFM [56]	38.0	47.9
DiffNet [57]	26.5	35.7
DS [21]	6.3	13.7
NM [20]	7.7	13.8
CZO [28]	7.6	39.3
UDM [10]	23.6	18.2
SyNoRiM [26]	6.2	13.8
Ours w/o **III**	6.3	12.0
Ours	**5.2**	**7.9**

Figure 3. **Matching with topological noise.** A summary of our quantitative comparisons on the topology benchmarks SHREC'Isometry [16] and TOPKIDS [33]. For both benchmarks, we show the cumulative error curves of our approach (red) and all considered baselines. Additionally, we provide the mean geodesic errors, averaged over all pairs of shapes, respectively (table, right).

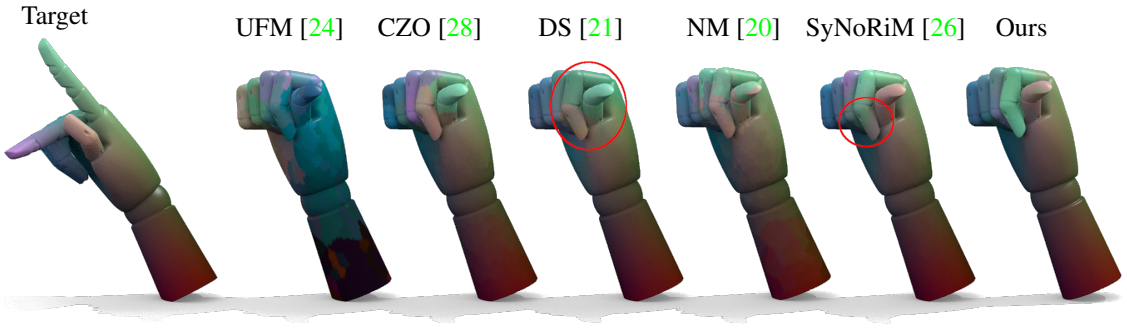


Figure 4. **Qualitative baseline comparison.** We consider a pair of real 3D scan meshes from SHREC'Iso [16], corresponding to the quantitative experiments shown in Figure 3. The two geometries are subject to topological merging due to self-contact of different fingers and parts of the hand. Correspondences are shown via a colormap for our method, as well as several baseline approaches. For this example, the best results are achieved by [21], [26] and our method. However, for both [21] and [26], the front part of the index and middle fingers are erroneous (tip of index finger should be bright green). See our supplementary material for additional qualitative examples.

Discussion The results on these four benchmarks are summarized in Table 1. Our method obtains state-of-the-art performance in all considered settings. Remarkably, these results were achieved directly through querying our network, whereas many baselines require correspondence postprocessing [10, 20, 24, 51, 56, 57]. Furthermore, the results underline that the shape graph module **III** plays a critical role in our pipeline for optimal performance.

4.2. Matching with topological changes

Datasets The benchmarks SHREC'Isometry [16] and TOPKIDS [33] focus on matching with topological noise. This is a common phenomenon when working with real scans, where the mesh topology is often corrupted by self-contact of separate parts of the scanned objects. Such topological merging severely affects the correspondence estimation since it distorts the intrinsic shape geometry non-isometrically. The first benchmark SHREC'Isometry [16] contains real scans of different humanoid puppets and hand models. A majority of poses in their 'heteromorphic' test set are subject to topological changes, see also Figure 4 for an example. The TOPKIDS [33] dataset contains synthetic shapes of human children where topological merging is emulated by computing the outer hull of intersecting geometries, see Figure 1 for an example.

Discussion Quantitative results are shown in Figure 3. We observe that topological merging commonly leads to unstable behavior for methods that rely on intrinsic priors like preservation of the Laplace-Beltrami operator [10, 28, 51, 56, 57] or pairwise geodesic distances [24]. It further inhibits approaches that learn to morph input geometries [20, 26] with explicit deformation priors, since merged regions tend to adhere to each other, see, e.g., the discussion on failure cases in [26, Sec. 7]. Our pipeline decreases the correspondence error by a decisive margin of 19% for SHREC'Iso and 73% for TOPKIDS. We provide a qualitative comparison in Figure 4, as well as additional examples in Appendix D.

	SH'20	SH'20 on SMAL	SH'20 on TOSCA					
			Cat	Centaur	Dog	Horse	Human	Wolf
UFM [24]	39.8	32.9	39.4	39.2	37.5	34.1	49.6	4.4
SURFM [51]	53.4	37.7	54.0	57.7	57.9	57.0	65.8	55.3
WFM [56]	31.4	20.2	20.6	21.9	16.7	22.4	38.1	5.7
DiffNet [57]	40.5	18.2	14.2	8.3	13.6	9.1	24.5	2.6
DS [21]	35.0	10.8	7.6	9.1	5.5	2.5	10.1	2.1
NM [20]	**10.0**	9.9	16.8	12.7	14.6	11.2	29.7	1.5
CZO [28]	21.7	–	–	–	–	–	–	–
UDM [10]	52.6	25.5	40.7	34.3	43.6	43.0	45.8	34.3
SyNoRiM [26]	10.4	5.7	12.8	11.6	10.6	7.1	28.2	2.0
Ours w/o **III**	11.1	3.4	6.3	6.0	4.9	2.6	20.1	2.2
Ours	10.6	**2.6**	**5.2**	**2.0**	**3.0**	**2.2**	**8.3**	**1.4**

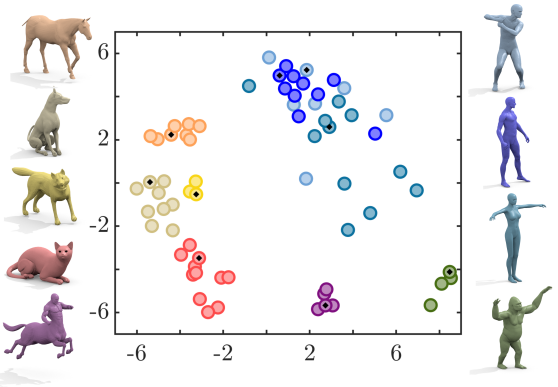


Figure 5. **Inter-class matching.** (left) A comparison of our approach to the considered baselines on SHREC'20 [17], as well as additional, synthetic test sets from SMAL [65] and TOSCA [8]. (right) Additionally, we visualize the shape graph node embeddings of our approach on TOSCA [8] through 2D multi-dimensional scaling. Since the learned edge weights express affinity scores, shapes with similar geometries tend to cluster together. Shapes with four legs (orange ≙ horse, red ≙ cat, . . .) and two legs (blue ≙ human, green ≙ gorilla) are linearly separable in the 2D MDS space. Interestingly, the centaur classes' embedding (purple) lies exactly between these two categories.

4.3. Inter-class matching

Datasets The SHREC'20 [17] challenge contains real scans of various four-legged animal models, including: elephant, giraffe, bear, and many more. These geometries were obtained from inhomogeneous acquisition sources, i.e., different types of scanners and 3D reconstruction pipelines. Sparse ground-truth correspondences were obtained through manual annotation. We further assess the generalization to additional test sets from the synthetic SMAL [65] and TOSCA [8] datasets. SMAL contains inter-class pairs between different animal classes, whereas TOSCA contains nearly-isometric pairs with both animal and human classes.

Discussion Our approach yields the most stable results overall, see Figure 5. Several baselines suffer from unstable behavior for animals from SHREC'20, because they depend on either noisy SHOT [59] input features [21], intrinsic priors that favor near isometries [28, 56, 57], or both [10, 24, 51]. While methods with an explicit deformation prior [20, 26] perform well on SHREC'20, they do not generalize well to unseen test poses from SMAL and TOSCA. Our method learns a topology-aware shape graph prior and thereby gets the best out of both worlds, i.e., robustness to inter-class pairs and strong generalization to unseen test pairs.

4.4. Sparse graph topologies

Throughout our experiments, we use complete shape graphs $\mathcal{G}$ with a full set of $\frac{N(N-1)}{2}$ edges, as specified in Section 3.3. Here, we explore a few alternative graph topologies with sparse connectivity patterns, i.e., $\mathcal{O}(N)$ edges: (ii) Minimal spanning trees (MST), (iii) minimal paths solving the traveling salesman problem (TSP) and (iv) star graphs, where all nodes are connected to one center node. For a detailed discussion and visualizations of these graph types, see Appendix B. We compare (ii)-(iv) to (i) the standard full graph and (v) our pipeline without the graph module **III**. A given graph type can be employed either for the cycle consistency loss in Equation (7) during training or for map refinement at test time. For a complete picture, we report results for all $5 \times 5 = 25$ possible combinations.

Our results in Table 2 indicate that, while the full graph is generally the most accurate, the sparse topologies often perform comparably, especially MST. This makes them a viable alternative to the full graph in certain scenarios with limited resources, both in terms of the required memory and query time. We provide a comprehensive cost analysis in Appendix C. In most cases, using the shape graph during training is beneficial, even when no graph is available at test time (Table 2, bottom row). This makes it relevant for online applications where not all test pairs are available at once. Regardless of the graph type, it is generally preferable to include some version of our graph module **III** rather than directly using the pairwise correspondences 'w/o'.

4.5. Ablation study

Our proposed architecture consists of several basic building blocks **I**-**III**, as defined in Section 3.2 and Section 3.3. While the shape-graph module **III** is unique to our approach, the feature backbone **I** and matching module **II** can, in principle, be replaced by any analogous off-the-shelf architectures. We compare several popular alternatives: For the feature backbone, we consider I.a the spectral convolution architecture from [21], I.b our network with SHOT [59] input features, I.c the ResNet architecture from [34], I.d PointNet [47], as well as I.e the message passing architecture from [20]. For the differentiable matching module, we compare II.a a functional map layer [44], II.b a single Sinkhorn

SH'Iso		Train on Full	MST	TSP	Star	w/o
Test on	Full	5.16	**5.09**	5.20	5.90	5.17
	MST	5.68	5.49	5.50	6.36	5.49
	TSP	6.08	5.62	5.80	6.94	6.51
	Star	5.54	5.26	5.44	6.71	6.02
	w/o	5.32	5.27	5.42	6.33	6.27

TOPKIDS		Train on Full	MST	TSP	Star	w/o
Test on	Full	**7.92**	8.13	8.44	11.03	9.13
	MST	8.56	8.62	9.39	10.57	9.98
	TSP	13.18	12.33	13.10	19.72	15.07
	Star	8.61	8.84	8.34	11.92	9.63
	w/o	10.62	10.64	11.61	13.62	12.02

Table 2. **Graph topology comparison.** We compare the quantitative performance of our model for different graph topologies $\mathcal{G}$. Specifically, we revisit the experiment from Figure 3 and report the mean geodesic error on SHREC'Iso [16] and TOPKIDS [33]. The standard 'full' graph is compared to three sparse topologies 'MST', 'TSP', 'star' graph, as well as the 'w/o **III**' variant of our pipeline.

		TOPKIDS		SHREC'20	
	with **III**	✓	✗	✓	✗
Feature **I**	**I.a** SpecConv [21]	8.53	13.68	28.54	34.99
	I.b SHOT [59]	7.93	13.66	22.74	29.81
	I.c ResNet [34]	7.94	13.14	39.69	40.66
	I.d PointNet [47]	8.78	14.10	11.01	11.54
	I.e GraphNN [20]	14.18	25.57	14.53	18.33
Match **II**	**II.a** FM [44]	39.12	40.66	50.58	51.37
	II.b Sinkhorn [13]	12.25	14.81	11.58	12.66
	II.c Softmax [20]	12.78	13.46	11.49	13.47
	G-MSM (ours)	**7.92**	**12.02**	**10.65**	**11.06**

Table 3. **Ablation network architecture.** We compare several off-the-shelf network architectures for the feature backbone **I** and matching module **II** to our full model, as defined in Section 3.2. For each setting, we contrast the results obtained with (✓) and without (✗) the graph-based multi-matching module **III**.

layer [13] and II.c a standard per-point Softmax [20].

We then replace either the feature backbone I.a-I.e or matching module II.a-II.c in our method and observe how it affects the accuracy on TOPKIDS [33] and SHREC'20 [17] from Section 4.2 and Section 4.3, respectively. The results are summarized in Table 3. Replacing either module **I** or **II** in our approach leads to a drop in performance. Moreover, we see that, regardless of the concrete architecture, our multi-matching approach **III** (✓in Table 3) improves the performance over the pairwise matches (✗).

5. Conclusion

We propose G-MSM, a novel multi-matching approach for non-rigid shape correspondence. For a given collection of 3D meshes, we define a shape graph $\mathcal{G}$ which approximates the underlying shape data manifold. Its edge weights w are extracted from putative pairwise correspondence signals in a self-supervised manner. Our network promotes cycle-consistency of optimal paths in $\mathcal{G}$. Thus, it produces context-aware multi-matches that are informed by commonalities and salient geometric features across all training poses. In our experiments, we demonstrate that this simple strategy yields significant improvements in correspondence accuracy on a wide range of challenging, real-world 3D mesh benchmarks.

Limitations & future work Our method can effectively learn the underlying canonical shape topology from a collection of 3D meshes. On the other hand, it relies on at least some of the poses to convey this latent topology. In the extreme case of $N = 2$ input poses, our multi-shape pipeline does not yield an improvement over the naive pairwise maps. The multi-matching **III** shows consistent improvements in our empirical evaluations. On the other hand, it is difficult to provide theoretical guarantees for our approach, since it is based on a self-supervised graph heuristic. We would like to further explore this direction in future work. While our approach learns correspondences in a self-supervised manner, it also assumes rigidly aligned input shapes which might be limiting for certain applications.

A promising avenue for future research is applying our multi-matching formulation to related but distinct settings to extend them beyond the pairwise training paradigm. One potential direction is extending our framework to allow for partial views, e.g., by leveraging recent advances on learnable partial functional maps [2].

Societal impact Advancing the robustness and accuracy of shape correspondence methods has the potential to open up new avenues for future applications based on 3D scan data. Our algorithm constitutes one small advancement in this effort of extending computer vision algorithms to the 3D domain. Since our algorithm is fully unsupervised, it can directly reduce deployment costs as no manual correspondence annotations are required to train our model. Shape correspondence is a fundamental building block at the heart of many 3D vision algorithms and we do not anticipate any immediate risk of misuse associated with this work.

Acknowledgements

We acknowledge support by the ERC Advanced Grant SIMULACRON, the Munich School for Data Science and the Munich Center for Machine Learning.

References

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

[2] Souhaib Attaiki, Gautam Pai, and Maks Ovsjanikov. Dpfm: Deep partial functional maps. In *2021 International Conference on 3D Vision (3DV)*, pages 175–185. IEEE, 2021. 1, 2, 3, 8

[3] Matthieu Aubry, Ulrich Schlickewei, and Daniel Cremers. The wave kernel signature: A quantum mechanical approach to shape analysis. *IEEE International Conference on Computer Vision (ICCV) - Workshop on Dynamic Shape Capture and Analysis*, 2011. 2

[4] Bharat Lal Bhatnagar, Cristian Sminchisescu, Christian Theobalt, and Gerard Pons-Moll. Loopreg: Self-supervised learning of implicit surface correspondences, pose and shape for 3d human mesh registration. *Advances in Neural Information Processing Systems*, 33:12909–12922, 2020. 2, 3, 14

[5] Federica Bogo, Javier Romero, Matthew Loper, and Michael J. Black. FAUST: Dataset and evaluation for 3D mesh registration. In *Proceedings IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, Piscataway, NJ, USA, June 2014. IEEE. 5

[6] Davide Boscaini, Jonathan Masci, Emanuele Rodolà, and Michael Bronstein. Learning shape correspondence with anisotropic convolutional neural networks. In *Advances in neural information processing systems*, pages 3189–3197, 2016. 2

[7] Alexander M Bronstein, Michael M Bronstein, and Ron Kimmel. Generalized multidimensional scaling: a framework for isometry-invariant partial surface matching. *PNAS*, 103(5):1168–1172, 2006. 2

[8] Alexander M Bronstein, Michael M Bronstein, and Ron Kimmel. *Numerical geometry of non-rigid shapes*. Springer, 2008. http://tosca.cs.technion.ac.il/book/resources_data.html. 7

[9] Michael M Bronstein, Joan Bruna, Yann LeCun, Arthur Szlam, and Pierre Vandergheynst. Geometric deep learning: going beyond euclidean data. *IEEE Signal Processing Magazine*, 34(4):18–42, 2017. 2

[10] D. Cao and F. Bernard. Unsupervised deep multi-shape matching. In *European Conference on Computer Vision (ECCV)*, 2022. 2, 3, 5, 6, 7, 16

[11] Luca Cosmo, Emanuele Rodola, Andrea Albarelli, Facundo Mémoli, and Daniel Cremers. Consistent partial matching of shape collections via sparse modeling. In *Computer Graphics Forum*, volume 36, pages 209–221. Wiley Online Library, 2017. 2

[12] Luca Cosmo, Emanuele Rodola, Jonathan Masci, Andrea Torsello, and Michael M Bronstein. Matching deformable objects in clutter. In *2016 Fourth International Conference on 3D Vision (3DV)*, pages 1–10. IEEE, 2016. 1

[13] Marco Cuturi. Sinkhorn distances: Lightspeed computation of optimal transport. In *Advances in neural information processing systems*, pages 2292–2300, 2013. 8, 13

[14] Nicolas Donati, Etienne Corman, Simone Melzi, and Maks Ovsjanikov. Complex functional maps: a conformal link between tangent bundles. *arXiv preprint arXiv:2112.09546*, 2021. 2

[15] Nicolas Donati, Abhishek Sharma, and Maks Ovsjanikov. Deep geometric functional maps: Robust feature learning for shape correspondence. *arXiv preprint arXiv:2003.14286*, 2020. 1, 2, 3, 5

[16] Roberto Dyke, Caleb Stride, Yukun Lai, and Paul Rosin. Shrec-19: Shape correspondence with isometric and non-isometric deformations. *Eurographics Workshop on 3D Object Retrieval*, 2019. 1, 2, 6, 8

[17] Roberto M Dyke, Yu-Kun Lai, Paul L Rosin, Stefano Zappalà, 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. 1, 2, 7, 8

[18] Marvin Eisenberger, Zorah Lähner, and Daniel Cremers. Divergence-free shape correspondence by deformation. In *Computer Graphics Forum*, volume 38, pages 1–12. Wiley Online Library, 2019. 2

[19] Marvin Eisenberger, Zorah Lahner, and Daniel Cremers. Smooth shells: Multi-scale shape registration with functional maps. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 12265–12274, 2020. 2, 13

[20] Marvin Eisenberger, David Novotny, Gael Kerchenbaum, Patrick Labatut, Natalia Neverova, Daniel Cremers, and Andrea Vedaldi. Neuromorph: Unsupervised shape interpolation and correspondence in one go. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 7473–7483, 2021. 1, 2, 3, 5, 6, 7, 8, 16, 17

[21] Marvin Eisenberger, Aysim Toker, Laura Leal-Taixe, and Daniel Cremers. Deep shells: Unsupervised shape correspondence with optimal transport. *arXiv preprint*, 2020. 1, 2, 3, 4, 5, 6, 7, 8, 12, 13, 15, 16, 17

[22] Maolin Gao, Zorah Lahner, Johan Thunberg, Daniel Cremers, and Florian Bernard. Isometric multi-shape matching. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 14183–14193, 2021. 3, 5

[23] Thibault Groueix, Matthew Fisher, Vladimir G. Kim, Bryan C. Russell, and Mathieu Aubry. 3d-coded: 3d correspondences by deep deformation. In *The European Conference on Computer Vision (ECCV)*, September 2018. 2, 3, 14

[24] Oshri Halimi, Or Litany, Emanuele Rodola, Alex M Bronstein, and Ron Kimmel. Unsupervised learning of dense shape correspondence. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 4370–4379, 2019. 1, 2, 3, 5, 6, 7

[25] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016. 2

[26] Jiahui Huang, Tolga Birdal, Zan Gojcic, Leonidas J Guibas, and Shi-Min Hu. Multiway non-rigid point cloud registration via learned functional map synchronization. *IEEE Transac-*

tions on Pattern Analysis and Machine Intelligence, 2022. 2, 3, 5, 6, 7, 16, 17

[27] Qi-Xing Huang and Leonidas Guibas. Consistent shape maps via semidefinite programming. In *Computer graphics forum*, volume 32, pages 177–186. Wiley Online Library, 2013. 2

[28] Ruqi Huang, Jing Ren, Peter Wonka, and Maks Ovsjanikov. Consistent zoomout: Efficient spectral map synchronization. In *Computer Graphics Forum*, volume 39, pages 265–278. Wiley Online Library, 2020. 2, 3, 5, 6, 7

[29] Faria Huq, Adrish Dey, Sahra Yusuf, Dena Bazazian, Tolga Birdal, and Nina Miolane. Riemannian functional map synchronization for probabilistic partial correspondence in shape networks. *arXiv preprint arXiv:2111.14762*, 2021. 3

[30] Itay Kezurer, Shahar Z Kovalsky, Ronen Basri, and Yaron Lipman. Tight relaxation of quadratic matching. In *Computer graphics forum*, volume 34, pages 115–128. Wiley Online Library, 2015. 2

[31] Vladimir G Kim, Yaron Lipman, and Thomas A Funkhouser. Blended intrinsic maps. *Transactions on Graphics (TOG)*, 30(4), 2011. 5

[32] Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014. 12

[33] Zorah Lähner, Emanuele Rodolà, Michael M Bronstein, Daniel Cremers, Oliver Burghard, Luca Cosmo, Andreas Dieckmann, Reinhard Klein, and Yusuf Sahillioglu. Shrec'16: Matching of deformable shapes with topological noise. *Proceedings of Eurographics Workshop on 3D Object Retrieval (3DOR)*, 2:11, 2016. 1, 2, 6, 8, 16

[34] Or Litany, Tal Remez, Emanuele Rodolà, Alex Bronstein, and Michael Bronstein. Deep functional maps: Structured prediction for dense shape correspondence. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 5659–5667, 2017. 2, 3, 7, 8

[35] Or Litany, Emanuele Rodolà, Alex Bronstein, and Michael Bronstein. Fully spectral partial shape matching. *Computer Graphics Forum*, 36(2):1681–1707, 2017. 2

[36] Or Litany, Emanuele Rodolà, Alex M Bronstein, Michael M Bronstein, and Daniel Cremers. Non-rigid puzzles. *Computer Graphics Forum (CGF), Proceedings of Symposium on Geometry Processing (SGP)*, 35(5), 2016. 2

[37] Matthew Loper, Naureen Mahmood, Javier Romero, Gerard Pons-Moll, and Michael J. Black. SMPL: A skinned multi-person linear model. *ACM Trans. Graphics (Proc. SIGGRAPH Asia)*, 34(6):248:1–248:16, Oct. 2015. 2, 5, 14

[38] Riccardo Marin, Simone Melzi, Emanuele Rodolà, and Umberto Castellani. FARM: functional automatic registration method for 3d human bodies. *CoRR*, abs/1807.10517, 2018. 2, 3

[39] Riccardo Marin, Marie-Julie Rakotosaona, Simone Melzi, and Maks Ovsjanikov. Correspondence learning via linearly-invariant embedding. *Advances in Neural Information Processing Systems*, 33:1608–1620, 2020. 1, 2, 3

[40] Jonathan Masci, Davide Boscaini, Michael Bronstein, and Pierre Vandergheynst. Geodesic convolutional neural networks on riemannian manifolds. In *Proceedings of the IEEE international conference on computer vision workshops*, pages 37–45, 2015. 2

[41] Simone Melzi, Riccardo Marin, Emanuele Rodolà, and Umberto Castellani. Matching humans with different connectivity. *Proceedings of Eurographics Workshop on 3D Object Retrieval (3DOR)*, 2019. 5

[42] Simone Melzi, Jing Ren, Emanuele Rodolà, Abhishek Sharma, Peter Wonka, and Maks Ovsjanikov. Zoomout: Spectral upsampling for efficient shape correspondence. *ACM Transactions on Graphics (TOG)*, 38(6):155, 2019. 2

[43] Federico Monti, Davide Boscaini, Jonathan Masci, Emanuele Rodola, Jan Svoboda, and Michael M Bronstein. Geometric deep learning on graphs and manifolds using mixture model cnns. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 5115–5124, 2017. 2

[44] Maks Ovsjanikov, Mirela Ben-Chen, Justin Solomon, Adrian Butscher, and Leonidas Guibas. Functional maps: a flexible representation of maps between shapes. *ACM Transactions on Graphics (TOG)*, 31(4):30, 2012. 2, 7, 8, 13

[45] Gabriel Peyré, Marco Cuturi, et al. Computational optimal transport: With applications to data science. *Foundations and Trends® in Machine Learning*, 11(5-6):355–607, 2019. 3

[46] Adrien Poulenard and Maks Ovsjanikov. Multi-directional geodesic neural networks via equivariant convolution. *ACM Transactions on Graphics (TOG)*, 37(6):1–14, 2018. 2

[47] Charles R Qi, Hao Su, Kaichun Mo, and Leonidas J Guibas. Pointnet: Deep learning on point sets for 3d classification and segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 652–660, 2017. 7, 8

[48] Jing Ren, Adrien Poulenard, Peter Wonka, and Maks Ovsjanikov. Continuous and orientation-preserving correspondences via functional maps. *ACM Trans. Graph.*, 37(6):248:1–248:16, Dec. 2018. 2, 5

[49] Emanuele Rodola, Samuel Rota Bulo, Thomas Windheuser, Matthias Vestner, and Daniel Cremers. Dense non-rigid shape correspondence using random forests. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 4177–4184, 2014. 2

[50] Emanuele Rodolà, Luca Cosmo, Michael Bronstein, Andrea Torsello, and Daniel Cremers. Partial functional correspondence. *Computer Graphics Forum (CGF)*, 2016. 2, 3

[51] Jean-Michel Roufosse, Abhishek Sharma, and Maks Ovsjanikov. Unsupervised deep learning for structured shape matching. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 1617–1627, 2019. 1, 2, 3, 5, 6, 7

[52] Klaus Hildebrandt Ruben Wiersma, Elmar Eisemann. Cnns on surfaces using rotation-equivariant features. *Transactions on Graphics*, 39(4), July 2020. 2

[53] Raif M Rustamov et al. Laplace-beltrami eigenfunctions for deformation invariant shape representation. In *Symposium on geometry processing*, volume 257, pages 225–233, 2007. 2

[54] Yusuf Sahillioğlu. Recent advances in shape correspondence. *The Visual Computer*, pages 1–17, 2019. 2

[55] Frank R Schmidt, Eno Töppe, Daniel Cremers, and Yuri Boykov. Intrinsic mean for semi-metrical shape retrieval via graph cuts. In *Joint Pattern Recognition Symposium*, pages 446–455. Springer, 2007. 2

[56] Abhishek Sharma and Maks Ovsjanikov. Weakly supervised deep functional map for shape matching. *arXiv preprint arXiv:2009.13339*, 2020. 1, 2, 3, 5, 6, 7

[57] Nicholas Sharp, Souhaib Attaiki, Keenan Crane, and Maks Ovsjanikov. Diffusionnet: Discretization agnostic learning on surfaces. *ACM Transactions on Graphics (TOG)*, 41(3):1–16, 2022. 1, 2, 3, 4, 5, 6, 7, 12, 13

[58] Jian Sun, Maks Ovsjanikov, and Leonidas Guibas. A concise and provably informative multi-scale signature based on heat diffusion. In *Computer graphics forum*, volume 28, pages 1383–1392. Wiley Online Library, 2009. 2

[59] Federico Tombari, Samuele Salti, and Luigi Di Stefano. Unique signatures of histograms for local surface description. *In Proceedings of European Conference on Computer Vision (ECCV)*, 16(9):356–369, 2010. 2, 7, 8

[60] Oliver van Kaick, Hao Zhang, Ghassan Hamarneh, and Daniel Cohen-Or. A survey on shape correspondence. *Computer Graphics Forum*, 30(6):1681–1707, 2011. 2

[61] Gul Varol, Javier Romero, Xavier Martin, Naureen Mahmood, Michael J Black, Ivan Laptev, and Cordelia Schmid. Learning from synthetic humans. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 109–117, 2017. 5, 14, 15

[62] Matthias Vestner, Zorah Lähner, Amit Boyarski, Or Litany, Ron Slossberg, Tal Remez, Emanuele Rodolà, Alex M. Bronstein, Michael M. Bronstein, Ron Kimmel, and Daniel Cremers. Efficient deformable shape correspondence via kernel matching. In *International Conference on 3D Vision (3DV)*, October 2017. 2

[63] Cédric Villani. *Topics in optimal transportation*. American Mathematical Soc., 2003. 3

[64] Thomas Windheuser, Ulrich Schlickwei, Frank R Schimdt, and Daniel Cremers. Large-scale integer linear programming for orientation preserving 3d shape matching. In *Computer Graphics Forum*, volume 30, pages 1471–1480. Wiley Online Library, 2011. 2

[65] Silvia Zuffi, Angjoo Kanazawa, David W Jacobs, and Michael J Black. 3d menagerie: Modeling the 3d shape and pose of animals. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 6365–6373, 2017. 1, 2, 7, 17