CoE: Deep Coupled Embedding for Non-Rigid Point Cloud Correspondences

Huajian Zeng[1,2,*] Maolin Gao[1,2,*] Daniel Cremers[1,2]
[1] Technical University of Munich [2] Munich Center for Machine Learning [*] Equal contribution

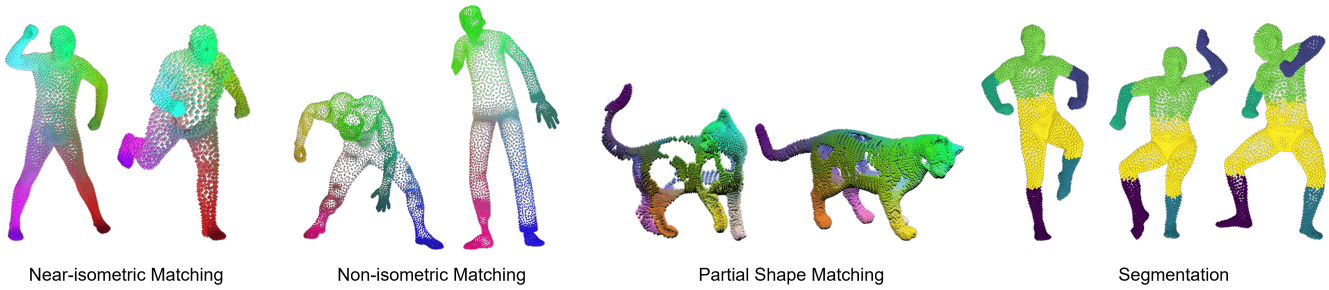

Figure 1. We propose a novel way to learn coupled embeddings of non-rigidly deformable shapes that are geometry-aware, robust and can be directly applied to retrieve accurate dense correspondences for near-isometric (*left*), non-isometric (*middle left*) and partial cases (*middle right*). Furthermore, it can also be employed for other shape analysis tasks such as shape segmentation (*right*).

Abstract

The interest in matching non-rigidly deformed shapes represented as raw point clouds is rising due to the proliferation of low-cost 3D sensors. Yet, the task is challenging since point clouds are irregular and there is a lack of intrinsic shape information. We propose to tackle these challenges by learning a new shape representation – a per-point high dimensional embedding, in an embedding space where semantically similar points share similar embeddings. The learned embedding has multiple beneficial properties: it is aware of the underlying shape geometry and is robust to shape deformations and various shape artefacts, such as noise and partiality. Consequently, this embedding can be directly employed to retrieve high-quality dense correspondences through a simple nearest neighbor search in the embedding space. Extensive experiments demonstrate new state-of-the-art results and robustness in numerous challenging non-rigid shape matching benchmarks and show its great potential in other shape analysis tasks, such as segmentation.

1. Introduction

Matching non-rigidly deformed 3D shapes is a long-standing and fundamental task in computer vision and graphics due to its ubiquitous role in many downstream tasks, such as shape editing, animation, medicine, statistical shape analysis, and robotics [17, 31, 42, 47]. Often, 3D shapes are represented as (triangular) meshes, which consist of both points and their (intrinsic) neighborhood connectivities. However, with the proliferation of low-cost sensors, the interest in methods that can directly deal with raw point clouds is expanding rapidly. Many (pointwise) shape descriptors have been proposed in the past decades, both hand-crafted [4, 41, 48] and learned [2, 8, 26]. Most of them are designed for shapes represented as triangular meshes and cannot be extended to point clouds without performance degradation [7, 21, 28]. A particularly interesting type of descriptor is a (high-dimensional) embedding of shapes, which is a shape representation that is ideally invariant under natural deformations and, at the same time, contains enough information to perform geometry processing tasks. Of particular interest is the global point signature (GPS) designed for triangular meshes [39], which transforms the extrinsic coordinates of each surface point into a higher (potentially infinite) dimensional space by exploiting the scale and isometric invariance of eigenfunctions of the Laplace-Beltrami Operator (LBO). While being effective, it turns out to be unstable due to sign ambiguity and complex spectrum in the LBO eigen-decomposition. On the other hand, the seminal work by Ovsjanikov et al. [30] proposes to align these LBO spectral embeddings before searching for the correspondence in this embedding space. The alignment of the high-dimensional spectral embeddings is called functional maps (fmaps), which can be represented compactly as a low-dimensional matrix. However, spectral em-

The order of equally contributed authors can be changed freely.
Code: https://github.com/zenghjian/coe

beddings are computed inefficiently by a non-differentiable eigen-decomposition of the LBO, which is sensitive to various practical artifacts, such as noise, partiality, and topological "short circuits".

Inspired by [16, 22], in which an (orthogonal) transformation of LBO eigenfunctions is estimated to obtain a consistent basis using manifold optimisation, we propose to leverage the power of deep learning to learn a coupled canonical embedding directly from raw point clouds, which can recover the LBO eigenbasis as a special case.

Due to insights gained from the classical geometry processing, we can obtain high-quality dense correspondences directly via a simple proximity search in the embedding space by training a *single* network, while all previous state-of-the-art methods have to train two networks [21, 28], underscoring the high practicability of our proposed method. Furthermore, our learned embedding is aware of the underlying geometry of the surface, efficient to compute, robust to various shape artefacts and applicable for various shape analysis tasks such as correspondences and segmentation (c.f. Fig. 1 & Sec. 5). Extensive experiments show that our proposed method can robustly map extrinsic coordinates of shapes, which undergo various non-rigid deformations, to a canonical embedding space where corresponding points share similar embeddings.

In summary, our contributions are:

- We propose a novel unsupervised way to learn per-point embeddings directly from raw point clouds under various non-rigid deformations. Inspired by classical geometry processing technique, our method is effective and simple that only requires to train a single network.
- In our learned embedding space, non-rigidly deformed shapes share similar and geometry-aware embeddings for corresponding points, which can be used for efficient matching by a simple nearest neighbor search.
- We show superior performance in a number of challenging non-rigid shape matching benchmarks, and unprecedented generalisation ability and robustness against different noise types, setting the new state-of-the-art.
- As a proof-of-concept, we show that our learned embeddings can be applied in other shape analysis tasks, such as partial shape matching and shape segmentation.

2. Related Work

The field of shape matching is vast and has rapidly developed over the past decades. Below we review the works which are most related to ours and can serve as baselines to our best knowledge. For a more comprehensive overview of the field, please refer to recent surveys [11, 40].

2.1. Pose Invariant Shape Representation

The study of pose invariant shape representation dates back to the beautiful work by Torgerson in 1952 [49], where

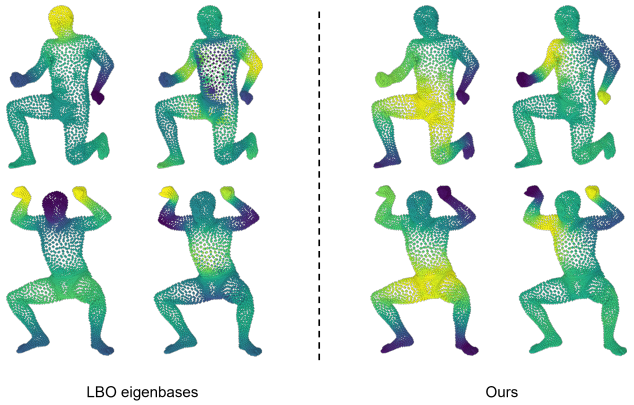

Figure 2. Examples of LBO eigenbases and our learned coupled embeddings on a pair of non-rigidly deformed shapes. Ours are consistent while LBO eigenbases suffer from sign flip (cf. Fig. 12 for more examples).

he introduced a lower-dimensional embedding, which preserves the pairwise (geodesic) distances between all graph nodes as much as possible and finds applications in many tasks, such as visualisation and clustering. This idea of dimensionality reduction has been further studied by [9, 37] based on the LBO. The work [9] utilises the LBO eigenfunctions and shows that these spectral properties can be employed to embed the data into a Euclidean space based on a diffusion process. Later approaches continue the exploration in the opposite direction, by embedding a surface into a higher dimensional space [36, 39, 54]. The GPS embedding [39] combines the LBO eigenvalues and eigenbases and in fact constructs a new surface in the infinite-dimensional space, which is invariant to (isometric) deformation. Impressive results on shape segmentation and clustering have demonstrated the effectiveness of the GPS embedding. From the standpoint of geodesic distance preservation, the authors of [54] design an embedding for the fast approximation of geodesics using a cascade strategy to gradually improve the accuracy of the approximation.

Most recently, deep-learning-based pose invariant embeddings are becoming prevalent. Most similar to our approach are the ones proposed in [21, 28]. DiffFmaps [28] proposes to learn a linearly invariant embedding from point clouds, which serves as a replacement for the pre-computed LBO eigenbasis in the fmaps framework. Improved robustness and better accuracy have been reported. Due to the employment of fmaps, it can efficiently regularise the maps in the functional space and incorporate structural regularisation. However, at the same time, it leads to the necessity of a *second* separate network dedicated to feature learning, hence a more complex pipeline in practice. Moreover, it requires ground truth correspondences to train, which is eliminated in NIE [21] by rigid pre-alignment (*i.e.* weak super-

vision), while retaining the dependency on a *second* feature network. In contrast, we propose to look at the correspondences problem through the classical geometry processing lens and learn a canonical embedding of the shape, which can be used directly for finding correspondences. As we show later, our embedding is motivated by the LBO spectral embedding while remaining coupled across different non-rigidly deformed shapes.

2.2. Basis Pursuit for Shape Analysis

In another line of work [16, 20, 22], researchers start to ask the question: is it possible to obtain a better set of basis suitable for shape correspondences? Huang et al. [20] proposes to learn a set of non-orthogonal bases and demonstrates its expressiveness and flexibility. However it internally converts the shape representation from point clouds to 3D voxel grid and requires careful engineering to obtain good results, such as post refinement and synchronisation. Moreover, it has to train a *second* network dedicated to feature learning due to the employment of fmaps. Kovnatsky et al. [22] tries to approximately diagonalise the LBOs of two shapes simultaneously. To reduce the dimensionality of the solution space, it makes use of subspace parametrisation to compute an (orthonormally) transformed version of the LBO eigenbases. However, ground truth dense correspondences are required, which we ultimately would like to estimate, rendering this method less practical for shape matching tasks. This requirement has been relaxed in [16] by requiring only sparse ground truth correspondences and dense descriptors. However its best performance still demands some ground truth labelling and both approaches involve complex manifold optimisation (cf. [16] & Tab. 2). Inspired by [16, 22] together with our insights on pose invariant representations, we propose to learn coupled embeddings directly from data. As we show in Sec. 4 & 5, our proposed method can learn high-quality coupled embeddings from low-quality shape descriptors. This attributes to the careful design of our geometry-aware unsupervised loss and network architecture, which enables *cross-communication* between shapes that is key for their coupling and consistency.

2.3. Learning on Point Clouds

Point cloud is arguably the most common representation for 3D shapes. However, due to its irregularity (compared to e.g. volumetric grids and triangular meshes), learning directly on point clouds has only become possible in recent years, enabled by specially designed architectures such as [33, 34, 45, 53]. While there are many works to learn deep features on meshes, tackling raw point clouds as the input 3D representation has been relatively less studied [14, 18, 20, 43]. This is partially due to the missing intrinsic proximity information in point clouds, which can be very helpful in many geometry processing tasks, such as computing geodesic distances. However, incorrect or even inconsistent topology often complicates algorithms and makes it very challenging to recover from it [8, 15]. In this work, we choose raw point clouds as our 3D shape representation. In this sense, our method is most similar to [14, 43], which employ fmaps in designing their losses. While GeomFmap [14] is a supervised approach which requires ground truth pointwise correspondences, a more recent work [43] shows that only an approximate pre-alignment of shapes can replace the costly demand in ground truth matches, which is often dubbed as weak supervision in the literature. We make use of the weak supervision same as in [43], since most datasets come already approximately rigidly aligned or can be easily aligned with very little manual intervention.

3. Background and Notation

In this section, we briefly review coupled diagonalisation for a pair of input shapes and introduce our notations (Tab. 1). See [16, 22] for a comprehensive discussion and the supplementary for an introduction of the LBO.

Given shapes $\mathcal{S}$ and $\mathcal{T}$ and their LBOs represented in stiffness matrices $\mathbf{L}_\mathcal{S}$, $\mathbf{L}_\mathcal{T}$ and mass matrices $\mathbf{M}_\mathcal{S}$, $\mathbf{M}_\mathcal{T}$, the coupled diagonalisation problem can be modelled as:

$$\min_{\{\mathbf{\Psi}_i\}} \sum_{i\in\{\mathcal{S},\mathcal{T}\}} \text{off}(\mathbf{\Psi}_i^\top \mathbf{L}_i \mathbf{\Psi}_i) + \mu_c \|\mathbf{D}_\mathcal{S}^\top \mathbf{M}_\mathcal{S} \mathbf{\Psi}_\mathcal{S} - \mathbf{D}_\mathcal{T}^\top \mathbf{M}_\mathcal{T} \mathbf{\Psi}_\mathcal{T}\| \tag{1}$$
$$\text{s.t.} \quad \mathbf{\Psi}_i^\top \mathbf{M}_i \mathbf{\Psi}_i = \mathbf{I}, \quad \text{for } i \in \{\mathcal{S},\mathcal{T}\}$$

where $\mathbf{D}_\mathcal{S}$ and $\mathbf{D}_\mathcal{T}$ are given descriptors (f.e. ground truth correspondences as indicator functions) and $\mathbf{I}$ is the identity matrix.

The off$(\cdot)$ term ensures that the coupled bases behave as approximate LBO eigenbases by penalising the off-diagonal entries and we chose $\text{off}(\mathbf{\Psi}^\top \mathbf{L} \mathbf{\Psi}) = \text{off}(\mathbf{\Psi}^\top \mathbf{L} \mathbf{\Psi}; \mathbf{\Lambda}) = \|\mathbf{\Psi}^\top \mathbf{L} \mathbf{\Psi} - \mathbf{\Lambda}\|$ throughout our experiments consistently, where $\mathbf{\Lambda}$ is a diagonal matrix of eigenvalues of the respective LBO. This choice helps to select leading bases (bases corresponding to small eigenvalues) with increasing frequency, which are the most informative ones in shape matching [30]. The second term is a coupling term that encourages the corresponding descriptors to behave similarly in the respective bases, which amounts to coupling the bases and making them to "speak the same language". Note that the basis of a 2-manifold is a generalisation of the (1D/2D) Fourier basis in the Euclidean space, which is fixed and always consistent.

The corresponding descriptors can be indicator (delta) functions representing (dense/sparse) ground truth pointwise correspondences, blobs or stable regions, distance functions and dense descriptors as discussed in [16, 22], and in practice some amount of ground truth information

Symbol	Description
$\mathcal{S}$	Source shape (point cloud)
$\mathbf{V}_{\mathcal{S}} \in \mathbb{R}^{n_{\mathcal{S}} \times 3}$	All $n_{\mathcal{S}}$ points of shape $\mathcal{S}$
$\mathbf{L}_{\mathcal{S}} \in \mathbb{R}^{n_{\mathcal{S}} \times n_{\mathcal{S}}}$	Stiffness matrix of shape $\mathcal{S}$
$\mathbf{M}_{\mathcal{S}} \in \mathbb{R}^{n_{\mathcal{S}} \times n_{\mathcal{S}}}$	Mass matrix of shape $\mathcal{S}$
$\mathbf{D}_{\mathcal{S}} \in \mathbb{R}^{n_{\mathcal{S}} \times d}$	Pointwise descriptors of shape $\mathcal{S}$
$\mathbf{\Phi}_{\mathcal{S}} \in \mathbb{R}^{n_{\mathcal{S}} \times k}$	LBO Eigenfunctions of shape $\mathcal{S}$
$\mathbf{\Lambda}_{\mathcal{S}} \in \mathbb{R}^{k \times k}$	LBO Eigenvalues of shape $\mathcal{S}$
$\hat{\mathbf{\Psi}}_{\mathcal{S}} \in \mathbb{R}^{n_{\mathcal{S}} \times k}$	Intermediate embedding of shape $\mathcal{S}$
$\mathbf{\Psi}_{\mathcal{S}} \in \mathbb{R}^{n_{\mathcal{S}} \times k}$	Predicted embedding of shape $\mathcal{S}$
$\mathcal{T}$	Target shape (point cloud)
$\vdots$	(analogous as above for $\mathcal{T}$)
f_θ	Embedding extractor with learnable θ
h_φ	Cross attention block with learnable φ
$\Pi_{\mathcal{ST}} \in \mathbb{R}^{n_{\mathcal{S}} \times n_{\mathcal{T}}}$	Binary matching matrix from $\mathcal{S}$ to $\mathcal{T}$

Table 1. Summary of our notation used in the paper.

is required for good performance [16, 22] (also see Sec. 5, Tab. 2), since the quality of the estimated coupled bases is strongly tied to the quality of the corresponding descriptors. Note that when $\mu_c \to 0$, problem (1) becomes separable and amounts to solving the LBO eigen-decomposition of $\mathcal{S}$ and $\mathcal{T}$ separately.

Since Eq. (1) does not scale well with the size of the shape, it makes the optimisation problem very challenging or even intractable for high resolution shapes. Therefore, the authors propose to solve a surrogate problem by subspace parameterisation, namely representing the coupled basis $\mathbf{\Psi}$ as a linear combination of the LBO eigenbasis $\mathbf{\Phi}$, *i.e.* $\mathbf{\Psi} = \mathbf{\Phi}\mathbf{R}$, where $\mathbf{R}$ is a Stiefel matrix. Compared to the original problem in Eq. (1), this modification greatly reduced the computational complexity, however it still involves difficult manifold optimisation for only approximately solving the original one. Furthermore, it demands at least a sparse set of ground truth correspondences to obtain good coupled bases, which, unfortunately, makes it dependent on either sparse shape matching methods or manual labeling (to produce the sparse correspondences).

To overcome these issues, we propose to directly learn coupled embeddings without any ground truth correspondences and without any subspace parameterisation. As demonstrated below, we only require noisy easy-to-obtain pointwise feature descriptors, f.e. heat kernel signature (HKS) [48], out of which our network can learn high-quality embeddings which are coupled and can be used directly for shape correspondence tasks.

4. Deep Coupled Embeddings

Real world shapes such as human and animals are intrinsically 2-dimensional compact manifolds and often embeded into the 3-dimensional Euclidean space and discretised as point clouds. It is of great interest to study the structure of the 2-manifold, rather than its Euclidean 3D embedding.

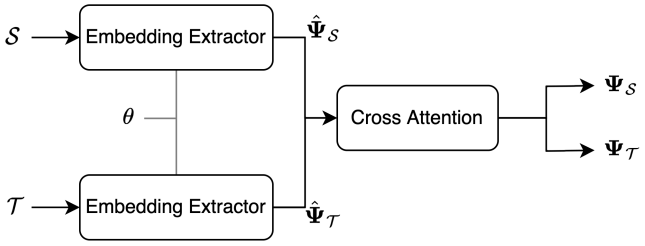

Figure 3. Pipeline overview. Given a pair of shapes $\mathcal{S}$ and $\mathcal{T}$ represented in point clouds, Our embedding extractor – ASAP DiffusionNet with shared weights θ (not to be confused with generative diffusion models [19, 46]), extracts the intermediate per-point embeddings $\hat{\mathbf{\Psi}}_{\mathcal{S}}$ and $\hat{\mathbf{\Psi}}_{\mathcal{T}}$, which are further refined by the subsequent cross attention block to output the final coupled embeddings $\mathbf{\Psi}_{\mathcal{S}}$ and $\mathbf{\Psi}_{\mathcal{T}}$. The cross attention block constructs a complete bipartite graph that connects every point on the shape $\mathcal{S}$ with every points on the shape $\mathcal{T}$ to enable their cross-communication. Our unsupervised loss encourages the predicted embeddings of both shapes to be coupled while closely resembling the LBO eigenbases.

One reason is that the intrinsic information of a shape (f.e. proximity) is "hidden" in point clouds, despite its convenience to store and render, limiting its usage in shape geometry and analysis tasks. Our network is designed primarily to recover the intrinsic proximity information, which can be used for direct retrieval of dense shape correspondences.

Given shapes $\mathcal{S}$ and $\mathcal{T}$ represented as point clouds $\mathbf{V}_{\mathcal{S}}$ and $\mathbf{V}_{\mathcal{T}}$, our proposed method learns their high-dimensional deep coupled embeddings $\mathbf{\Psi}_{\mathcal{S}}$ and $\mathbf{\Psi}_{\mathcal{T}}$, based on which accurate dense correspondences can be obtained via a simple proximity search in the embedding space. In this section, We first introduce our network architecture design in Sec. 4.1, which combines a recent variant of DiffusionNet [2, 45] with cross attention to encourage information exchange during the learning process. Subsequently, we introduce our loss in Sec. 4.2. Note that our loss does not require any ground truth, hence enabling 3D representation learning in a fully data driven fashion. In Sec. 4.3, we explain our dense correspondence retrieval based on the learned embeddings. Throughout this section, we discuss the key design insight to achieve the coupling of learned embeddings and shed light to their geometric properties that are valuable for shape analysis tasks.

4.1. Network Architecture

Our network architecture is simple, efficient and comprises two main building blocks: an embedding extractor f_θ and a cross attention module h_φ with learnable parameters θ and φ, which we will elaborate next. An illustration of our pipeline can be found in Fig. 3.

Embedding Extractor Module computes per point intermediate embedding $\hat{\mathbf{\Psi}}_{(\cdot)}$, which is a non-linear mapping:

$$f_\theta : \mathbf{V}_{(\cdot)} \rightarrow \hat{\mathbf{\Psi}}_{(\cdot)} \tag{2}$$

where $\cdot$ can either be shape $\mathcal{S}$ or $\mathcal{T}$ and $\hat{\mathbf{\Psi}}_{(\cdot)}$ will be further refined in the up-coming cross attention block.

Note that many point cloud learning methods [33, 34, 45, 53] discussed in Sec. 2.3 can be employed here. However, careful design choice is required due to our special learning objective, namely the learned embedding must retain close to the LBO eigenbasis (cf. Sec. 4.2), indicating that the learned (intermediate) embedding must be smooth. This relates to the fact that the smallest eigenfunctions (low frequency) of LBO vary smoothly on the manifold.

This naturally leads to the choice of As-Smooth-As-Possible (ASAP) DiffusionNet, a variant of DiffusionNet architecture proposed by Attaiki et al. [2] as the default backbone of our embedding extractor. It captures the local geometric information of different scales on the manifold by modelling a heat diffusion process with different timesteps and constrains the learned embedding to live in the space spanned by the LBO eigenbasis. Both aspects encourage smoothness while retaining expressiveness of the learned embedding, which we found particularly suitable for our task. Note that the realm of point cloud learning is still very active and yet our pipeline is flexible, that advances in the field can be directly incorporated by a drop-in replacement of the ASAP DiffusionNet.

Cross Attention Block refines the independently predicted intermediate embeddings $\hat{\mathbf{\Psi}}_\mathcal{S}$, $\hat{\mathbf{\Psi}}_\mathcal{T}$ by encouraging the communication between them. It follows the Transformer architecture [51] and learns a non-linear mapping:

$$h_\varphi : \{\hat{\mathbf{\Psi}}_\mathcal{S}, \hat{\mathbf{\Psi}}_\mathcal{T}\} \rightarrow \{\mathbf{\Psi}_\mathcal{S}, \mathbf{\Psi}_\mathcal{T}\} \tag{3}$$

The output $\mathbf{\Psi}_\mathcal{S}$ and $\mathbf{\Psi}_\mathcal{T}$ are directly used to form our unsupervised loss (cf. Sec. 4.2), which will be minimised and update the learnable network parameters through back-propagation. Specifically, we construct a fully connected, bipartite graph that connects every point on the shape $\mathcal{S}$ with every points on the shape $\mathcal{T}$. Each node in the graph is assigned with the corresponding intermediate embedding learned by the embedding extractor. The core concept of cross attention is that it computes a similarity matrix between the key and query (transformed version of $\hat{\mathbf{\Psi}}_\mathcal{S}$, $\hat{\mathbf{\Psi}}_\mathcal{T}$), and makes use of it to weight the value (again a transformed version of $\hat{\mathbf{\Psi}}_\mathcal{S}$ or $\hat{\mathbf{\Psi}}_\mathcal{T}$) to produce the final output (please refer to [3, 51] for details).

The key in this process is that it enables the cross-talk of the intermediate embeddings $\hat{\mathbf{\Psi}}_\mathcal{S}$ and $\hat{\mathbf{\Psi}}_\mathcal{T}$, which is essential for a coupled and consistent shape embedding. This is akin to the idea of joint diagonalisation [22] and image co-segmentation [52], where the information of the other object (shape/image) has to be made available in some way to achieve consistency. As a result, our final embedding is aware of the other shape and hence coupled and consistent as shown in Fig. 2 and Fig. 12.

4.2. Unsupervised Loss

Our unsupervised loss is inspired by the work of classical geometry processing [16, 22] and consists of three terms. Among them the off-diagonal loss and the orthogonal loss together encourage the learned embeddings to behave similarly as the classical LBO eigenbases, and the contrastive loss penalises their inconsistency. We will introduce them one by one in the following.

Off-diagonal Loss: Similar as in Eq. (1), the learned embedding $\mathbf{\Psi}_{(\cdot)}$ should approximately diagonalise the respective Laplacian $\mathbf{L}_{(\cdot)}$.

$$L_{\text{off}} = \sum_{i \in \{\mathcal{S},\mathcal{T}\}} \left\| \mathbf{\Psi}_i^T \mathbf{L}_i \mathbf{\Psi}_i - \mathbf{\Lambda}_i \right\|_F \tag{4}$$

Note that $\mathbf{\Lambda}_{(\cdot)}$ is a diagonal matrix of increasing eigenvalues of the respective LBO sitting on the diagonal. This term also encourages the learned embedding to be frequency-aligned, namely the smoother (lower frequency) an embedding is, the earlier it is positioned in the full set of embeddings.

Orthogonal Loss: The orthogonal constraint in Eq. (1) is relaxed to a soft penalty in our training objective. It encourages the learned embedding to possess a basis structure and prevent undesired rank deficiency, hence maximising the embedding space spanned by the learned embeddings.

$$L_{\text{o}} = \sum_{i \in \{\mathcal{S},\mathcal{T}\}} \left\| \mathbf{\Psi}_i^\top \mathbf{M}_i \mathbf{\Psi}_i - \mathbf{I} \right\|_F \tag{5}$$

In fact, the optimal embedding to minimise both the orthogonal loss and the off-diagonal loss is the individual LBO eigenbasis of shape $\mathcal{S}$ and $\mathcal{T}$, which is a special case of our formulation. Moreover, we circumvent the intractable complexity of high-dimensional manifold optimisation in Eq. (1) by leveraging a data-drive learning technique, which enables a direct prediction of per-point embedding without any subspace parameterisation required in [16, 22].

Contrastive Loss: This term couples the learned embeddings $\mathbf{\Psi}_\mathcal{S}$ and $\mathbf{\Psi}_\mathcal{T}$ and encourages their mutual consistency.

$$L_{\text{c}} = \left\| \mathbf{D}_\mathcal{S}^T \mathbf{M}_\mathcal{S} \mathbf{\Psi}_\mathcal{S} - \mathbf{D}_\mathcal{T}^T \mathbf{M}_\mathcal{T} \mathbf{\Psi}_\mathcal{T} \right\|_F \tag{6}$$

Similar as in Eq. (1), the coupling is achieved by driving the Fourier coefficients of corresponding descriptor functions $\mathbf{D}_\mathcal{S}$ and $\mathbf{D}_\mathcal{T}$ to be as close as possible. Different to Eq. (1), we can learn highly accurate coupled embeddings from low-quality descriptor functions (f.e. HKS), fully eliminating the need of ground truth correspondences required in [16, 22] (see Sec. 5, Tab. 2). To our best knowledge, this enables, for the first time, the practical application

Geo. error (×100)	Train	FAUST		FAUST			SCAPE		SCAPE		
	Test	FAUST	SCAPE	SHREC19	TOPKIDS	DT4D-M	SCAPE	FAUST	SHREC19	TOPKIDS	DT4D-M
NN Spectral Embedding		67.1	-	-	-	-	62.3	-	-	-	-
HKS [48]		43.0	-	-	-	-	40.5	-	-	-	-
CQHB-HKS [22]		37.2	-	-	-	-	31.6	-	-	-	-
CQHB-GT [22]		10.5	-	-	-	-	10.8	-	-	-	-
SyNoRiM(S) [20]		7.9	21.9	25.5	-	-	9.5	24.6	26.8	-	-
GeomFMaps(S) [14]		6.1	11.2	10.8	26.2	38.5	7.7	9.0	12.4	21.7	28.6
WSupFMNet(W) [43]		6.0	12.5	13.8	28.9	40.2	11.3	7.5	12.6	24.5	30.1
DiffFMaps(S) [28]		4.3	18.7	14.6	20.5	18.5	14.4	10.8	14.2	18.0	15.9
NIE(W) [21]		5.9	16.7	15.1	18.9	13.3	11.6	8.6	13.2	16.2	12.1
NIE(W) [21](with ASAP)		5.6	15.0	20.7	19.7	13.5	12.6	5.9	23.5	15.3	12.0
SSMSM(W) [7]		**2.4**	**6.8**	**9.0**	14.2	**11.8**	4.1	4.1	**5.2**	12.3	8.0
Ours w/o ASAP(W)		3.9	8.8	16.2	15.3	14.0	4.3	3.9	13.1	14.6	10.9
Ours(W)		3.7	8.7	9.5	**13.7**	13.1	**3.2**	**3.7**	8.1	**11.0**	**7.8**

Table 2. Quantitative results on FAUST, SCAPE, SHREC19, TOPKIDS and DT4D-M. The **best** results are highlighted, and the second best results are indicated in blue. All methods only take point clouds as input except the multimodal method SSMSM [7], which requires meshes. Ours outperforms all baselines, both classical and learning-based methods, and is comparable (if not superior) to SSMSM. Letters S,W in parentheses stand for supervised and weakly supervised respectively.

of dense shape correspondence estimation based on coupled embeddings.
Finally, our full unsupervised loss is written as:

$$L_{\text{total}} = \mu_{\text{off}} L_{\text{off}} + \mu_{\text{o}} L_{\text{o}} + \mu_{\text{c}} L_{\text{c}} \tag{7}$$

where $\mu_{\text{off}} = 1, \mu_{\text{o}} = 5e1$ and $\mu_{\text{c}} = 1e3$ are the corresponding weights. Please see supplementary for implementation details.

4.3. Dense Correspondences

After the network is trained, we can directly obtain coupled embeddings $\mathbf{\Psi}_{\mathcal{S}}, \mathbf{\Psi}_{\mathcal{T}}$ from two input point clouds $\mathbf{V}_{\mathcal{S}}, \mathbf{V}_{\mathcal{T}}$ at inference time. Since both the coupled embeddings are predicted by the same network and live in the same embedding space, they are directly comparable. To retrieve dense pointwise correspondences, we employ the simple nearest neighbor search.

$$\text{NN} : \{\mathbf{\Psi}_{\mathcal{S}}, \mathbf{\Psi}_{\mathcal{T}}\} \rightarrow \Pi_{\mathcal{ST}} \tag{8}$$

Namely for the i-th source point in shape $\mathcal{S}$, we search for a target j-th point in $\mathcal{T}$, whose l_2 distance to the source point is smallest in the embedding space and assign $\Pi_{\mathcal{ST}}(i, j) = 1$, indicating a match. Note that $\Pi_{\mathcal{ST}}$ is a binary matrix, but not (always) a permutation matrix, since the correspondences are not guaranteed to be bijective.

5. Experiments

We start this section by introducing the most relevant baselines in Sec. 5.1 before reporting experiment results on near-isometric and non-isometric matching in Sec. 5.2 & 5.3. Then we study the generalisation ability (Sec. 5.4) and robustness (Sec. 5.5) of our method due to their high practical relevance. Lastly as proof-of-concepts, we show that our learned embedding can be used for challenging partial shape matching (Sec. 5.6) and segmentation (Sec. 5.7).

5.1. Baselines

We compare our method with relevant baselines, including both axiomatic and learning-based methods.

CQHB is an inspiring work by Kovnatsky et al. [22] using classical optimisation. We evaluate it in two different settings: namely with HKS or ground truth correspondences as indicator functions and report their results as *CQHB-HKS* and *CQHB-GT* respectively.

DiffFMaps [28], *NIE* [21] and *SyNoRiM* [20] are the SOTA methods to learn shape embeddings (or bases as in *SyNoRim*) from point clouds. They are the most related to ours since all methods aim to learn a good shape embedding under challenging non-rigid deformations. Additionally, we report the results of the most competitive method *NIE* using ASAP DiffusionNet as feature extractor.

GeomFMaps [14] and *WSupFMNet* [43] are SOTA fmaps-based learning methods, they are relevant to ours since they only take point clouds as input and produce dense shape correspondences.

Lastly, the SOTA multimodal learning method *SSMSM* [7] is also a fmaps-based shape matching method. In addition to the input point cloud, it also requires the face information contained in meshes, where as ours only needs raw point clouds. We showcase in the experiments that our proposed method, despite only having access to point clouds, performs on par (if not superior) with *SSMSM*.

5.2. Near-isometric Shape Matching

Datasets We choose FAUST [6], SCAPE [1] and SHREC19 [29] as testbeds for the task of near-isometric shape matching, specifically the more recent remeshed version [13, 35] of them. The FAUST dataset encompasses 100 human shapes, representing 10 individuals in 10 distinct poses. We split them as $80/20$ for train and test. The SCAPE dataset comprises 71 shapes of a single person in different poses. We split them as $51/20$ for train and test.

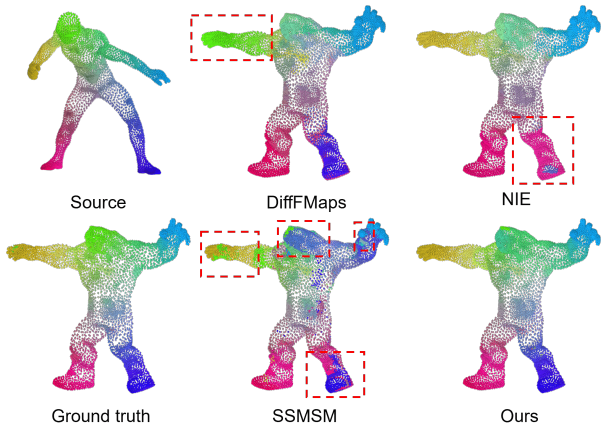

Figure 4. Qualitative result on DT4D-M. Ours produces the most accurate and smooth correspondences, despite highly non-isometric deformation (errors highlighted in red).

The SHREC19 dataset includes 44 human shapes and is exclusively used as a test set. Note that due to remeshing, the distribution of each point cloud is totally different, rendering the matching task more realistic and challenging.

Results We train on FAUST and SCAPE respectively and evaluate on FAUST, SCAPE and SHREC19.

As shown in Tab. 2, our proposed method can produce significantly more accurate correspondences than its classical counterpart CQHB, even under CQHB-GT in which dense ground truth correspondences are used. Moreover, we conduct a simple experiment to retrieve dense shape correspondences by nearest neighbor search directly using HKS (cf. Tab. 2). The quality of estimated correspondences is inferior. However, our method can fully exploit the information available in the low-quality HKS descriptor and predict highly accurate correspondences.

Our method also outperforms SOTA learning methods, even the ones with ground truth supervision such as SyNoRiM and GeomFMaps. Remarkably, ours is capable to compete (it not superior) with the multimodal learning method SSMSM, which requires meshes as input. This highlights the importance of the careful design of our network architecture and our unsupervised loss inspired by the classical geometry processing technique. As an ablative study we disable the ASAP component hence employ the vanilla DiffusionNet as feature extractor and report its quantitative results in Tab. 2 as Ours w/o ASAP. Note that the mean geodesic error deteriorates in all cases, underlining the importance of smoothness of learned embeddings. Please refer to the supplementary for qualitative results and additional ablation experiments.

5.3. Non-isometric Shape Matching

Datasets We employ the recent non-isometric benchmark DT4D-M [27] as the testbed for this task. This dataset includes shapes from the large-scale animation dataset DT4D [25] and consists of 293 humanoid shapes from 9 different classes. We split it as $198/95$ for train and test. Following the train/test split proposed in [24], we conduct experiments with all 9 classes of humanoid shapes, which undergo significant non-isometric deformation (cf. Fig. 4).

Results We test on DT4D-M using our model trained on FAUST and SCAPE respectively. Note that this is a harder case than training and testing on the same dataset, since all methods are only trained with near-isometric shapes. However, our proposed method performs favorably than all baselines and achieves comparable result with mesh-dependent SSMSM method. We report the quantitative and qualitative results in Tab. 2 (column DT4D-M) and Fig. 4 respectively.

Geo. error ($\times$100)	Train / Test	FAUST	**SURREAL** SCAPE	SHREC19
GeomFMaps(S) [13]		10.4 (-)	8.7 (-)	14.1 (-)
WSupFMNet(W) [43]		16.0 (-)	14.7 (-)	27.8 (-)
DiffFMaps(S) [28]		7.8 (22.8)	18.9 (26.9)	27.8 (34.2)
NIE(W) [21]		6.9 (11.3)	11.0 (17.2)	11.3 (18.2)
NIE(W) [21](with ASAP)		5.5 (8.8)	9.5 (13.6)	11.0 (16.3)
SSMSM(W) [7]		3.5 (6.8)	3.8 (6.4)	6.6 (9.8)
Ours(W)(w/o ASAP)		**3.3 (5.1)**	4.3 (5.8)	8.8 (13.2)
Ours(W)		3.4 (5.2)	**3.3 (5.2)**	**4.6 (9.4)**

Table 3. Generalisation ability. The **best** results in each column are highlighted. Our method outperforms all learning based baselines. Letters S,W in parentheses stand for supervised and weakly supervised respectively.

5.4. Generalisation

Datasets To further study the generalisability of our proposed method, we employ the SURREAL dataset [50], which is a synthetic dataset of human shapes. We train our model and baselines on a randomly sampled subset of the $230K$ synthetic shapes and test on FAUST, SCAPE, and SHREC19.

Results Quantitative and qualitative results are reported in Tab. 3 and Fig. 5 respectively. Remarkably, ours outperforms all baselines including the multimodal mesh-dependent method SSMSM under this setting. A possible reason is that our learned embeddings are driven by the geometry-aware supervision, and are further coupled via both the network architecture (cross attention block) and the (constrastive) loss. This geometry-aware supervision and strong coupling foster the generalisation ability, leading to the superior performance of our proposed method.

5.5. Robustness

We evaluate robustness from two perspectives: (1) random additive Gaussian noise to point clouds, (2) changes and inconsistency in shape topology. Both scenarios are common in real-world raw point clouds, hence are highly relevant for the practicability of investigated methods.

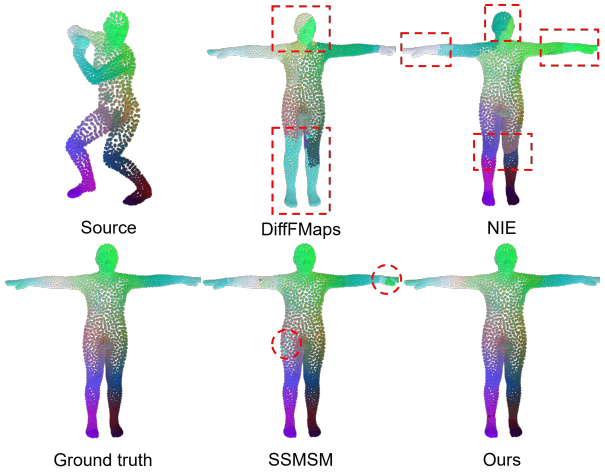

Figure 5. Generalisation from the training set SURREAL to the test set SHREC19. Our method generalises better compared to baselines (errors highlighted in red).

Additive Gaussian Noise We make use of the trained model in SURREAL (Sec. 5.4) and test on *noisy* point clouds from FAUST, SCAPE, SHREC19. Every point in the test point clouds is perturbed by a Gaussian with $\mu = 0$ and $\sigma = 0.01$ and within a range of $[-0.05, 0.05]$. Quantitative results are shown in Tab. 3 (numbers in parentheses). Under this noisy setting, the quality of our correspondences retains the best among all competing methods. Compared to the noise-free case, we also have the least overall performance degradation. An illustration is shown in Fig. 17 in the supplementary.

Topology changes We employ models pre-trained on FAUST and SCAPE respectively and test on the TOPKIDS dataset [23], which contains 26 shapes of kids with non-rigid deformation and topological changes for this task.

Quantitative results are shown in Tab. 2 (column TOPKIDS). Note that all investigated methods suffer from the challenging topological changes, however ours outperforms by achieving the lowest mean geodesic error. Qualitative illustration (Fig. 6 & 15) also shows that our predicted correspondences are the closest to the ground truth.

5.6. Partial Shape Matching

As a proof-of-concept, we show that our proposed method can be applied to the challenging partial shape matching. For this we train our network on SHREC16 Partiality [10]. During training we take a full and partial pair and employ an extended loss (see supplementary for details). Once the network is trained, it can be used to directly match two partial shapes by proximity search in the embedding space. Examples can be found in Fig. 1 and in the supplementary.

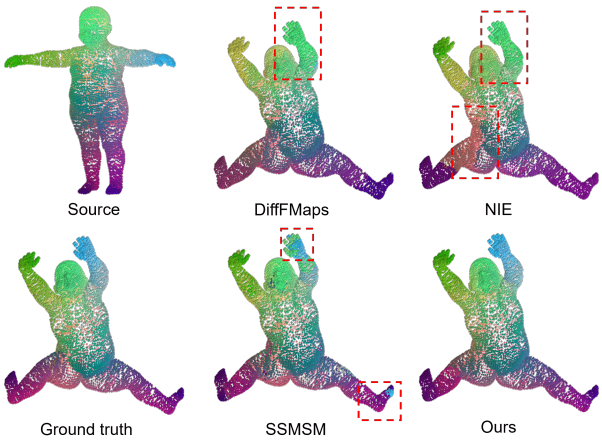

Figure 6. Robustness against topological changes (the left shoulder and face of the kid are glued together). Ours is least sensitive to this noise among all competing methods (errors highlighted in red).

5.7. Shape Segmentation

As a proof-of-concept, we show that our learned embedding can also be employed for shape segmentation tasks. Specifically we conduct a k-mean clustering on the learned embedding of each shape. An example is shown in Fig. 1 and the segmentation is meaningful and even consistent across different shapes, despite independently segmented.

6. Limitations, Future Work and Conclusion

In this paper, we proposed an unsupervised method to learn high-quality, well-generalised embeddings directly from raw point clouds. The embedding is aware of the underlying shape geometry and robust to various shape artefacts and non-rigid (both isometric and non-isometric) deformations and can be used to obtain dense correspondences via a simple proximity search in the canonical embedding space. Extensive experiments showcase that our proposed method achieves superior results in a number of non-rigid matching benchmarks and is promising in other shape analysis challenges, such as partial shape matching and segmentation, hence setting the new state-of-the-art.

Our method also has limitations. First, it requires shapes to be pre-aligned. An interesting direction is to incorporate the advancement in SO(3)/SE(3) invariant architecture [12] to eliminate the necessity of pre-alignment. Second, it is interesting to explore the possibility for a fully descriptor-free approach. Lastly, an extension of our method to shape collections would be a promising avenue for future research.

7. Acknowledgment

The project is supported by the ERC Advanced Grant SIMULACRON and 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. 6

[2] Souhaib Attaiki and Maks Ovsjanikov. Understanding and improving features learned in deep functional maps. In *CVPR*, 2023. 1, 4, 5

[3] Souhaib Attaiki, Gautam Pai, and Maks Ovsjanikov. DPFM: Deep Partial Functional Maps. In *3DV*, 2021. 5, 1

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

[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, 2014. IEEE. 3

[6] Federica Bogo, Javier Romero, Matthew Loper, and Michael J Black. Faust: Dataset and evaluation for 3d mesh registration. In *CVPR*, 2014. 6

[7] Dongliang Cao and Florian Bernard. Self-supervised learning for multimodal non-rigid 3d shape matching. In *CVPR*, 2023. 1, 6, 7, 3

[8] Dongliang Cao, Paul Roetzer, and Florian Bernard. Unsupervised learning of robust spectral shape matching. *ACM Transactions on Graphics (TOG)*, 2023. 1, 3

[9] Ronald R Coifman, Stephane Lafon, Ann B Lee, Mauro Maggioni, Boaz Nadler, Frederick Warner, and Steven W Zucker. Geometric diffusions as a tool for harmonic analysis and structure definition of data: Diffusion maps. *Proceedings of the national academy of sciences*, 102(21):7426–7431, 2005. 2

[10] Luca Cosmo, Emanuele Rodola, Michael M Bronstein, Andrea Torsello, Daniel Cremers, Y Sahillioğlu, et al. Shrec'16: Partial matching of deformable shapes. In *Eurographics Workshop on 3D Object Retrieval, EG 3DOR*, pages 61–67. Eurographics Association, 2016. 8, 2, 3

[11] Bailin Deng, Yuxin Yao, Roberto M. Dyke, and Juyong Zhang. A Survey of Non-Rigid 3D Registration. *CGF*, 2022. 2

[12] Congyue Deng, Or Litany, Yueqi Duan, Adrien Poulenard, Andrea Tagliasacchi, and Leonidas Guibas. Vector neurons: a general framework for so(3)-equivariant networks. *arXiv preprint arXiv:2104.12229*, 2021. 8

[13] Nicolas Donati, Abhishek Sharma, and Maks Ovsjanikov. Deep geometric functional maps: Robust feature learning for shape correspondence. In *CVPR*, 2020. 6, 7

[14] Nicolas Donati, Abhishek Sharma, and Maks Ovsjanikov. Deep geometric functional maps: Robust feature learning for shape correspondence. In *CVPR*, 2020. 3, 6

[15] Marvin Eisenberger, Aysim Toker, Laura Leal-Taixé, and Daniel Cremers. G-msm: Unsupervised multi-shape matching with graph-based affinity priors. In *CVPR*, 2023. 3

[16] Davide Eynard, Artiom Kovnatsky, Michael M. Bronstein, Klaus Glashoff, and Alexander M. Bronstein. Multimodal manifold analysis by simultaneous diagonalization of laplacians. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2015. 2, 3, 4, 5

[17] Pablo Gainza, Freyr Sverrisson, Frederico Monti, Emanuele Rodola, Davide Boscaini, Michael M Bronstein, and Bruno E Correia. Deciphering interaction fingerprints from protein molecular surfaces using geometric deep learning. *Nature Methods*, 17(2):184–192, 2020. 1

[18] Thibault Groueix, Matthew Fisher, Vladimir G Kim, Bryan C Russell, and Mathieu Aubry. 3d-coded: 3d correspondences by deep deformation. In *ECCV*, 2018. 3

[19] Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising diffusion probabilistic models. In *Neurips*, 2020. 4

[20] 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 Transactions on Pattern Analysis and Machine Intelligence*, pages 1–1, 2022. 3, 6

[21] Puhua Jiang, Mingze Sun, and Ruqi Huang. Neural intrinsic embedding for non-rigid point cloud matching. In *CVPR*, pages 21835–21845, 2023. 1, 2, 6, 7

[22] Artiom Kovnatsky, Michael M. Bronstein, Alexander M. Bronstein, Klaus Glashoff, and Ron Kimmel. Coupled Quasi-harmonic Bases. *CGF*, 2013. 2, 3, 4, 5, 6

[23] Zorah Lähner, Emanuele Rodolà, Michael M Bronstein, Daniel Cremers, Oliver Burghard, Luca Cosmo, Alexander Dieckmann, Reinhard Klein, Y Sahillioğlu, et al. Shrec'16: Matching of deformable shapes with topological noise. In *Eurographics Workshop on 3D Object Retrieval, EG 3DOR*, pages 55–60. Eurographics Association, 2016. 8

[24] Lei Li, Nicolas Donati, and Maks Ovsjanikov. Learning multi-resolution functional maps with spectral attention for robust shape matching. In *NeurIPS*, 2022. 7

[25] Yang Li, Hikari Takehara, Takafumi Taketomi, Bo Zheng, and Matthias Nießner. 4dcomplete: Non-rigid motion estimation beyond the observable surface. In *ICCV*, 2021. 7

[26] Or Litany, Tal Remez, Emanuele Rodola, Alex Bronstein, and Michael Bronstein. Deep functional maps: Structured prediction for dense shape correspondence. In *ICCV*, 2017. 1

[27] Robin Magnet, Jing Ren, Olga Sorkine-Hornung, and Maks Ovsjanikov. Smooth non-rigid shape matching via effective dirichlet energy optimization. In *3DV*, 2022. 7

[28] Riccardo Marin, Marie-Julie Rakotosaona, Simone Melzi, and Maks Ovsjanikov. Correspondence learning via linearly-invariant embedding. *Neurips*, 2020. 1, 2, 6, 7

[29] Simone Melzi, Riccardo Marin, Emanuele Rodolà, Umberto Castellani, Jing Ren, Adrien Poulenard, Peter Wonka, and Maks Ovsjanikov. Shrec 2019: Matching humans with different connectivity. In *Eurographics Workshop on 3D Object Retrieval*, 2019. 6

[30] 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):1–11, 2012. 1, 3

[31] Gianluca Paravati, Fabrizio Lamberti, Valentina Gatteschi, Claudio Demartini, and Paolo Montuschi. Point cloud-based

automatic assessment of 3d computer animation courseworks. *IEEE Transactions on Learning Technologies*, 2017. 1

[32] Adam Paszke, Sam Gross, Francisco Massa, Adam Lerer, James Bradbury, Gregory Chanan, Trevor Killeen, Zeming Lin, Natalia Gimelshein, Luca Antiga, et al. Pytorch: An imperative style, high-performance deep learning library. *Neurips*, 2019. 1

[33] Charles Ruizhongtai Qi, Hao Su, Kaichun Mo, and Leonidas J Guibas. Pointnet: Deep learning on point sets for 3d classification and segmentation. In *CVPR*, 2017. 3, 5

[34] Charles Ruizhongtai Qi, Li Yi, Hao Su, and Leonidas J Guibas. Pointnet++: Deep hierarchical feature learning on point sets in a metric space. *Neurips*, 2017. 3, 5

[35] Jing Ren, Adrien Poulenard, Peter Wonka, and Maks Ovsjanikov. Continuous and orientation-preserving correspondences via functional maps. *ACM Transactions on Graphics (TOG)*, 37(6):1–16, 2018. 6

[36] Martin Reuter. Hierarchical shape segmentation and registration via topological features of laplace-beltrami eigenfunctions. *Int. J. Comput. Vision*, 2010. 2

[37] Martin Reuter, Franz-Erich Wolter, and Niklas Peinecke. Laplace-beltrami spectra as 'shape-dna' of surfaces and solids. *Comput. Aided Des.*, 2006. 2

[38] Emanuele Rodolà, Luca Cosmo, Michael M Bronstein, Andrea Torsello, and Daniel Cremers. Partial functional correspondence. In *CGF*, 2017. 3

[39] Raif M. Rustamov. Laplace-Beltrami Eigenfunctions for Deformation Invariant Shape Representation. In *Geometry Processing*. The Eurographics Association, 2007. 1, 2

[40] Yusuf Sahillioğlu. Recent advances in shape correspondence. In *The Visual Computer*, 2020. 2

[41] Samuele Salti, Federico Tombari, and Luigi Di Stefano. Shot: Unique signatures of histograms for surface and texture description. *Computer Vision and Image Understanding*, 125:251–264, 2014. 1

[42] Carlos Sánchez-Belenguer, Simone Ceriani, Pierluigi Taddei, Erik Wolfart, and Vítor Sequeira. Global matching of point clouds for scan registration and loop detection. *Robot. Auton. Syst.*, 2020. 1

[43] Abhishek Sharma and Maks Ovsjanikov. Weakly supervised deep functional maps for shape matching. In *Neurips*, 2020. 3, 6, 7

[44] Nicholas Sharp and Keenan Crane. A Laplacian for Nonmanifold Triangle Meshes. *CGF*, 2020. 1

[45] Nicholas Sharp, Souhaib Attaiki, Keenan Crane, and Maks Ovsjanikov. Diffusionnet: Discretization agnostic learning on surfaces. *ACM Trans. Graph.*, 2022. 3, 4, 5

[46] Yang Song, Jascha Sohl-Dickstein, Diederik P Kingma, Abhishek Kumar, Stefano Ermon, and Ben Poole. Score-based generative modeling through stochastic differential equations. In *ICLR*, 2021. 4

[47] Mikkel B Stegmann and David Delgado Gomez. A brief introduction to statistical shape analysis. *Informatics and mathematical modelling, Technical University of Denmark, DTU*, 15(11), 2002. 1

[48] Jian Sun, Maks Ovsjanikov, and Leonidas Guibas. A concise and provably informative multi-scale signature based on heat diffusion. In *CGF*, 2009. 1, 4, 6, 3

[49] Warren S Torgerson. Multidimensional scaling: I. theory and method. *Psychometrika*, 17(4):401–419, 1952. 2

[50] Gul Varol, Javier Romero, Xavier Martin, Naureen Mahmood, Michael J Black, Ivan Laptev, and Cordelia Schmid. Learning from synthetic humans. In *CVPR*, 2017. 7

[51] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is all you need. *Neurips*, 2017. 5

[52] Fan Wang, Qixing Huang, and Leonidas J. Guibas. Image co-segmentation via consistent functional maps. In *ICCV*, 2013. 5

[53] Yue Wang, Yongbin Sun, Ziwei Liu, Sanjay E. Sarma, Michael M. Bronstein, and Justin M. Solomon. Dynamic graph cnn for learning on point clouds. *ACM Trans. Graph.*, 2019. 3, 5

[54] Qianwei Xia, Juyong Zhang, Zheng Fang, Jin Li, Mingyue Zhang, Bailin Deng, and Ying He. Geodesicembedding (ge): A high-dimensional embedding approach for fast geodesic distance queries. *IEEE Transactions on Visualization and Computer Graphics*, 2022. 2