



HAL
open science

Smoothness and effective regularizations in learned embeddings for shape matching

Riccardo Marin, Souhaib Attaiki, Simone Melzi, Emanuele Rodolà, Maks Ovsjanikov

► **To cite this version:**

Riccardo Marin, Souhaib Attaiki, Simone Melzi, Emanuele Rodolà, Maks Ovsjanikov. Smoothness and effective regularizations in learned embeddings for shape matching. 2023. hal-03938011

HAL Id: hal-03938011

<https://hal.science/hal-03938011>

Preprint submitted on 14 Jan 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Smoothness and effective regularizations in learned embeddings for shape matching

Riccardo Marin	Souhaib Attaiki	Simone Melzi
Sapienza	LIX	Sapienza
University of Rome	Ecole Polytechnique, IP Paris	University of Rome
marin@di.uniroma1.it	attaiki@lix.polytechnique.fr	melzi@di.uniroma1.it
Emanuele Rodolà	Maks Ovsjanikov	
Sapienza	LIX	
University of Rome	Ecole Polytechnique, IP Paris	
rodola@di.uniroma1.it	maks@lix.polytechnique.fr	

Abstract

Many innovative applications require establishing correspondences among 3D geometric objects. However, the countless possible deformations of smooth surfaces make shape matching a challenging task. Finding an embedding to represent the different shapes in high-dimensional space where the matching is easier to solve is a well-trodden path that has given many outstanding solutions. Recently, a new trend has shown advantages in learning such representations. This novel idea motivated us to investigate which properties differentiate these data-driven embeddings and which ones promote state-of-the-art results. In this study, we analyze, for the first time, properties that arise in data-driven learned embedding and their relation to the shape-matching task. Our discoveries highlight the close link between matching and smoothness, which naturally emerge from training. Also, we demonstrate the relation between the orthogonality of the embedding and the bijectivity of the correspondence. Our experiments show exciting results, overcoming well-established alternatives and shedding a different light on relevant contexts and properties for learned embeddings.

1 Introduction

Deep learning largely bases its success on converting input features into others useful to solve specific applications. From a more abstract perspective, training a neural network is no more than optimizing the parameters of a non-linear function which maps between two embeddings. Scholars largely explored this perspective, showing also how operations like interpolation [23], algebra [50], or analogies [64] difficult in the starting domain, may be simple in learned onece. Inferring and manipulating relations between different objects lies at the core of learning: deciding what is different, what is similar, and what are the relations in a population eases the learning process and enables powerful applications. Several fields uses alignment in learned embedding as a way to infer relations; few examples among others are Natural Language Processing [36, 53, 62], Biology [65, 32], and 3D Geometry [58, 6, 63]. This latter domain, given its nature, often paves the road with novel tools, providing geometrical insights that quickly spread across communities [60, 42, 9].

In 3D shape analysis, point-wise features and descriptors as high-dimensional embeddings are analyzed for decades [21, 28, 55]. One practical approach, namely the multidimensional scaling [58], consists in embedding the intrinsic geometry of a 3D shape into a higher-dimensional space where Euclidean distances approximate the proper metric of the object. The convenience of this procedure is

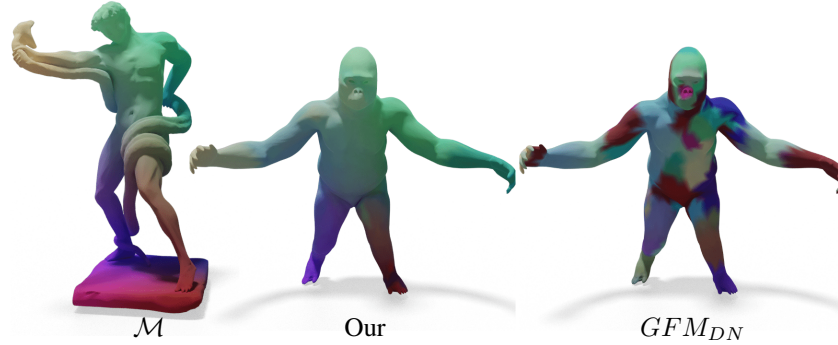


Figure 1: An example of extreme non-isometric matching. The left shape is a statue from [1], the second is a gorilla from [7].

clear, as one commonly prefers to work with extrinsic distances (lengths of straight lines in Euclidean space) instead of intrinsic ones (lengths of curved paths on a surface). Another possibility arises from the eigenfunctions of the Laplace-Beltrami Operator (LBO) that have played a prominent role in deformable shape representation. Their theoretical properties (i.e. orthogonality and smoothness) make them convenient for several applications, such as surface fairing, function analysis, segmentation, and shape matching [15, 45, 34]. Among the variations on this theme [10, 29, 25, 24], recently, some learned embeddings specific for shape matching have been proposed [61, 54]. Along this line, [22] proposed to learn an alternative embedding to address non-rigid shape matching for point clouds. It offers a novel data-driven direction and provides promising results. Despite the wide attention devoted to these techniques, the research mainly follows general intuitions without analyzing how such properties may influence the relations between elements in the learned embedding.

We aim to fill this knowledge gap by analyzing some properties and their effectiveness in the learned embedding, which kind of relations they promote, and their applicative contexts.

We instantiate the method from [22] to work on meshes adopting a state-of-the-art feature extractor [52]; this setting lets us define interesting properties clearly (e.g., smoothness). Moreover, the point cloud scenario of [22] limits the possible analysis given the absence of reliable axiomatic competitors. For the first time, we study the effect of smoothness and orthogonality on features for matching, providing both theoretical derivations and empirical evidence of our results. Furthermore, previous methods mainly focus on objects in bijection or with weak remeshing. Instead, we consider both bijective and non-bijective cases, showing more realistic scenarios. Our analysis not only pursues theoretical considerations but also has an applicative impact. Designing the embedding properties carefully improve performances, achieving a precision level otherwise impossible with previous representations (even when they enjoy the best transformation possible). Figure 1 is a qualitative example of this. We believe that our results significantly impact any field where embedding alignment is a common practice, starting from the Geometry Processing one where this problem is prominent.

To summarize, our work provides the following novelties and contributions:

1. We suggest *novel effective regularizations* to learn embedding for the matching task. We introduce new simple losses, providing theoretical motivations and showing their practical utility.
2. We offer an *extensive analysis of learned embeddings* in the shape matching application. We question some of the most established beliefs from the 3D shape matching community for the first time. We shed new light on prominent properties that characterize existing solutions disclosing new research direction.
3. We *outperform state-of-the-art alternatives*. We exceed strong shape-matching competitors in multiple scenarios. To our best knowledge, this is the first study that produces an embedding that can represent shape correspondence better than LBO eigenfunctions, even for the near-isometric case.

All the data and code will be made publicly available for research purposes.

2 Related work

Shape matching has a long history [59, 49]. We cover the more closely related methods to our work.

Embeddings for shape matching Multidimensional scaling and its variants [58, 6, 8] are seminal approaches for computing shape embeddings for shape matching. The main idea is to realize the geodesic metric of a given 3D object as the Euclidean (also called restricted) metric of some high-dimensional space. Such approaches are variational since, in general, a solution does not exist even for simple cases [19]. Several works have recently proposed extensions of the Laplacian eigenbasis to deal with difficult settings. In [29], it was proposed to augment the LBO set with a locally supported basis, defined only in a specific region, to increase its local representation capability; [25] proposes to add extrinsic information by appending the orthonormalized version of the three coordinates vectors; [18] suggested finding a basis by jointly diagonalizing the Laplacians of the input pair of shapes. Interestingly, all the previous methods base their representation on the standard LBO basis.

Another popular approach in shape matching is finding a set of features (also called descriptors) that are able to identify each point uniquely, such as the Global Point Signature [48], Wave Kernel Signature [4], Heat Kernel Signature [55], AWFT [28], SHOT [57] and its robust version proposed in [30]. However, while they are widely used, such features alone are insufficient to provide accurate point-to-point matching, especially in non-rigid scenarios.

Functional maps [34, 35] introduce and describe the paradigm considered in this paper. They move the correspondence problem to the functional domain. The advantages come from the low-dimensional approximation of these spaces, which arises from a set of fixed basis functions (mainly the LBO eigenvectors). This framework has been extended to take into account partiality [47], to provide precise matching within triangle faces [14], and improved with several regularizations [33, 44] or iterative refinement [27, 43, 37].

Learning-based pipelines Several data-driven pipelines for shape matching only address the rigid case [38, 17, 51, 61], while 3DCoded [16] addresses the non-rigid setting at the cost of an expensive optimization at test time. In the latter setting, an entire line of research has been devoted to learning the descriptors used by functional maps. This has been done using random forests [46], and more recently via deep learning models such as FMNet [20]. Recently, [12] proposed to learn the features directly from the point cloud coordinates while keeping the LBO basis. The main inspiration of our work is [22], which proposes to learn both the bases and descriptors. This method only processes 3D point clouds since it uses the PointNet architecture [40] without exploiting the possible input connectivity. Furthermore, it does not investigate the properties and applicability of the learned basis. Among the other contributions, our work also aims to fill this gap.

3 Notation and background

3D shapes. We model 3D shapes as 2-dimensional Riemannian manifolds embedded in \mathbb{R}^3 . In the discrete setting, we encode these objects as triangular meshes composed of vertices and faces defined by the oriented triplet of vertices that belong to the same triangular face. Each face is glued to a maximum of three other faces, one for each edge. We denote with \mathcal{M} and \mathcal{N} a pair of shapes and with $X_{\mathcal{M}} \in \mathbb{R}^{n_{\mathcal{M}} \times 3}$ and $X_{\mathcal{N}} \in \mathbb{R}^{n_{\mathcal{N}} \times 3}$ the list of the 3D coordinates of their $n_{\mathcal{M}}$ and $n_{\mathcal{N}} \in \mathbb{N}$ vertices. To each of these shapes, we associated the Laplace-Beltrami Operator (LBO), the second-order partial differential operator extending the standard Laplacian to non-Euclidean domains and denoted respectively as $\Delta_{\mathcal{M}}$ and $\Delta_{\mathcal{N}}$. We adopt the same notation referring to the square matrices, with size $n_{\mathcal{M}}$ and $n_{\mathcal{N}}$ respectively, which encode these operators in the discrete setting and that can be estimated through the cotangent weight formula [39, 31].

The LBO is a symmetric, positive semi-definite operator which admits an eigendecomposition with non-negative real eigenvalues, sorted in non-descending order $0 = \lambda_1 \leq \lambda_2 \leq \dots$. The eigenfunctions $\phi_1^\Delta, \phi_2^\Delta, \dots$ associated with each eigenvalue compose a basis for the space of square-integrable functions defined over the surface, in analogy with the Fourier basis on Euclidean domains. In the discrete setting, each eigenfunction corresponds to a vector with a length equal to the number of vertices. We store the set of the k eigenfunctions associated to the first k eigenvalues with smallest absolute values, as columns of a matrix $\Phi^\Delta = [\phi_1^\Delta, \dots, \phi_k^\Delta]$. Each row of this matrix is a vector in

\mathbb{R}^k and is referred to as the *spectral embedding* of the corresponding vertex of the mesh. The matrix Φ^Δ thus encodes the spectral embedding of the entire shape.

Functional maps. Given the two shapes \mathcal{M} and \mathcal{N} , together with their truncated set of LBO eigenfunctions $\Phi^\Delta_{\mathcal{M}}$ and $\Phi^\Delta_{\mathcal{N}}$ respectively, we denote with $\mathcal{F}(\mathcal{M})$ and $\mathcal{F}(\mathcal{N})$ the space of real-valued functions defined on \mathcal{M} and \mathcal{N} . Any point-to-point correspondence $T_{\mathcal{M}\mathcal{N}} : \mathcal{M} \rightarrow \mathcal{N}$ induces a functional mapping (with opposite direction) $T_{\mathcal{N}\mathcal{M}}^{\mathcal{F}} : \mathcal{F}(\mathcal{N}) \rightarrow \mathcal{F}(\mathcal{M})$ via pull-back. Exploiting the Fourier analogy, we can approximate the spaces $\mathcal{F}(\mathcal{M})$ and $\mathcal{F}(\mathcal{N})$ in the given bases $\Phi^\Delta_{\mathcal{M}}$ and $\Phi^\Delta_{\mathcal{N}}$ of size k . Thanks to this approximation, we can compactly encode the mapping $T_{\mathcal{N}\mathcal{M}}^{\mathcal{F}}$ in a matrix $C_{\mathcal{N}\mathcal{M}} \in \mathbb{R}^{k \times k}$ which corresponds to the linear transformation that maps the coefficients of functions approximated by $\Phi^\Delta_{\mathcal{N}}$ to the coefficients of their images through $T_{\mathcal{N}\mathcal{M}}^{\mathcal{F}}$ represented by $\Phi^\Delta_{\mathcal{M}}$. In matrix notation, if we encode the point-to-point map $T_{\mathcal{M}\mathcal{N}}$ in a binary matrix $\Pi_{\mathcal{M}\mathcal{N}} \in \mathbb{R}^{n_{\mathcal{M}} \times n_{\mathcal{N}}}$, such that its entries $\Pi_{\mathcal{M}\mathcal{N}}(i, j) = 1$ if and only if the correspondence $T_{\mathcal{M}\mathcal{N}}$ associates, to the i -th vertex of \mathcal{M} , the vertex of index j on \mathcal{N} , then we can explicitly compute the functional map $C_{\mathcal{N}\mathcal{M}} = \Phi^\Delta_{\mathcal{M}}^\dagger \Pi_{\mathcal{M}\mathcal{N}} \Phi^\Delta_{\mathcal{N}}$, where we denote with \dagger the Moore–Penrose pseudoinverse. In this framework, due to the analogy with Fourier, the matrices $\Phi^\Delta_{\mathcal{M}}$ and $\Phi^\Delta_{\mathcal{N}}$ play the role of bases for the functional spaces. For this reason, we will refer to them both as spectral embedding and as basis.

Linearly invariant training. Our work follows the architecture introduced in [22] but modifies it accordingly to work with triangular meshes. Given a set of shape pairs $(\mathcal{M}, \mathcal{N})$, equipped with a ground truth correspondence $\Pi_{\mathcal{M}\mathcal{N}}^{gt}$, we train an embedding network \mathcal{E} . This network takes as input the coordinates $X_{\mathcal{M}}$ of a shape \mathcal{M} and outputs a high-dimensional embedding $\Phi^{\mathcal{E}}_{\mathcal{M}}$. The network is trained by considering the optimal linear transformation between the two shape embeddings:

$$C_{\mathcal{N}\mathcal{M}}^{opt} = (\Phi^{\mathcal{E}}_{\mathcal{M}})^\dagger \Pi_{\mathcal{M}\mathcal{N}}^{gt} \Phi^{\mathcal{E}}_{\mathcal{N}}, \quad (1)$$

which is converted to a penalty measuring how well the embedded points are aligned; since nearest-neighbor is not differentiable, we cast this problem as follows:

$$\begin{aligned} D &= dist(\Phi^{\mathcal{E}}_{\mathcal{M}} C_{\mathcal{N}\mathcal{M}}^{opt}, \Phi^{\mathcal{E}}_{\mathcal{N}}) \\ S_{\mathcal{M}\mathcal{N}} &= softmax(-D), \end{aligned} \quad (2)$$

where $dist$ computes the matrix of Euclidean distances in the embedding space, and $S_{\mathcal{M}\mathcal{N}}$ acts as a score of similarity between points. Finally, the network loss is formulated as:

$$L_{\mathcal{E}}(\Phi^{\mathcal{E}}_{\mathcal{M}}, \Phi^{\mathcal{E}}_{\mathcal{N}}) = \sum \|S_{\mathcal{M}\mathcal{N}} X_{\mathcal{N}} - \Pi_{\mathcal{M}\mathcal{N}}^{gt} X_{\mathcal{M}}\|_2^2. \quad (4)$$

In experiments which also require descriptors, we learn them following [22] paradigm: we train the same architecture of \mathcal{E} , with the loss proposed by Equation 4 of [22].

DiffusionNet. We adopt a state-of-the-art feature extractor as the backbone. DiffusionNet [52] relies on a diffusion process over the surface to propagate the features information. A diffusion block is based on learning the parameters of a heat diffusion process, plus an anisotropic filter. We refer to the related paper for the technical details. We choose DiffusionNet since it is one of the most recent and promising architectures designed to exploit the structure of the 3D surfaces.

Notation. We denote with $\Phi^{\mathcal{E}}$ the embedding learned by minimizing $L_{\mathcal{E}}$ and with Φ^Δ the LBO basis (both coincide with a matrix), while we adopt *Our* when we use both learned embedding and descriptors.

4 Method

The main purpose of [22] was proposing a pipeline end-to-end (learning basis and descriptors) for sparse point clouds matching, where no good basis alternatives are known and require thousands of training data. We emphasize that our study is orthogonal to [22] and other approaches. Our work does not aim to introduce a novel framework but to analyze properties of existing ones by exploring different basis in a data-driven implementation similar to [22]. Remarkably, we propose a novel

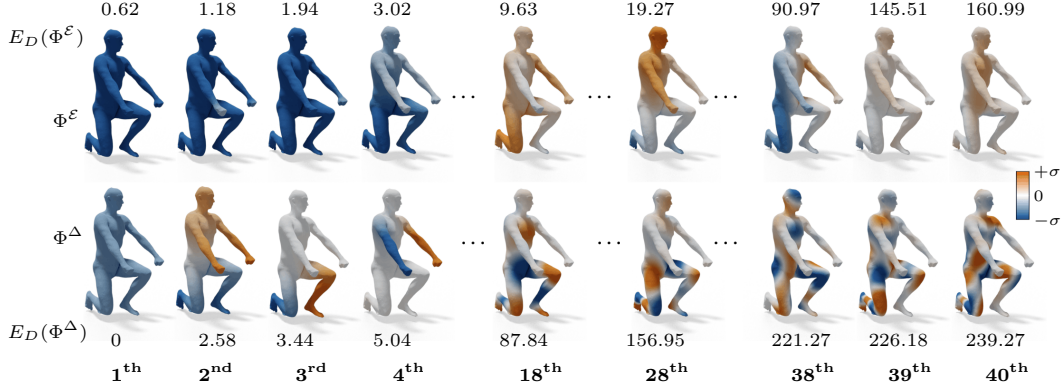


Figure 2: We visualize the dimensions of the embeddings as a function over the surface. For each one, we report its Dirichlet energy E_D as a measure of the smoothness.

perspective, a challenging setting, and a far more ambitious goal: overtaking LBO in its domain, with a limited training set, studying which properties make it competitive and which can be relaxed. We focus on the bases, considering descriptors only in application experiments. This approach lets us unveil theoretical properties and insights about good representations for embedding alignment.

The main properties of Φ^Δ are *orthogonality* and *smoothness*. Furthermore, these functions carry a natural order given by the corresponding eigenvalues. We describe the losses we adopt to force the learned basis to own the desired properties. All these losses do not require supervision at training time.

Smoothness. A good point-wise representation for the matching application should provide similar encoding for the nearest points. Smoothness could guarantee this property as preliminary proved in [22]. In the functional context, smoothness coincides with the Dirichlet energy (namely E_D): the smaller the energy, the smoother the function. We refer to the supplementary materials for its discretization. We foster this property through the following loss:

$$L_{E_D}(\Phi^\mathcal{E}_\mathcal{M}) = \sum \|diag(\Phi^\mathcal{E}_\mathcal{M} A \Phi^\Delta \Lambda \Phi^\Delta A \Phi^\mathcal{E}_\mathcal{M})\|_2^2 \quad (5)$$

L_{E_D} is a dense version of the Dirichlet energy, and thus its minimization promotes smoothness. We will show in our analysis the prominent role of smoothness in functional shape matching. A visualization of functions with different smoothness levels is depicted in Figure 2.

Orthogonality. Given a multi-dimensional representation of the points, looking at them a vertex-wise functions (i.e. columns of the embedding matrix), we impose orthogonality by minimizing the following loss:

$$L_\perp(\Phi^\mathcal{E}_\mathcal{M}) = \sum \|\Phi^\mathcal{E}_\mathcal{M} \Phi^\mathcal{E}_\mathcal{M} - Id\|_2^2, \quad (6)$$

where Id is the identity matrix. In the Appendix A, we prove that the orthogonality of the learned features promotes *weak bijectivity* in the matching. We refer to L_{w_\perp} when the loss is scaled by a 10^{-10} factor.

Sparsity. We do not limit our analysis to the properties owned by the LBO basis. Inspired by the great success of the *compressed sensing* in signal processing, we introduce a new loss to promote sparsity in the learned embedding. Given the importance of smoothness for the matching, we only combine the sparsity and the smoothness losses as follows:

$$L_{\ell_1+E_D} = L_{\ell_1} + w_{E_D} L_{E_D}, \quad (7)$$

$$\text{where } L_{\ell_1}(\Phi^\mathcal{E}_\mathcal{M}) = \sum \|\Phi^\mathcal{E}_\mathcal{M}\|_1. \quad (8)$$

Due to the different magnitude of the losses, in L_{ℓ_1+S} we scaled $Loss_S$ by $w_{E_D} = 10^6$. We highlight that, to assess the role of the matching loss in the proposed procedure, we do not optimize for the loss in Equation 4 when we adopt the $L_{\ell_1+E_D}$ one.

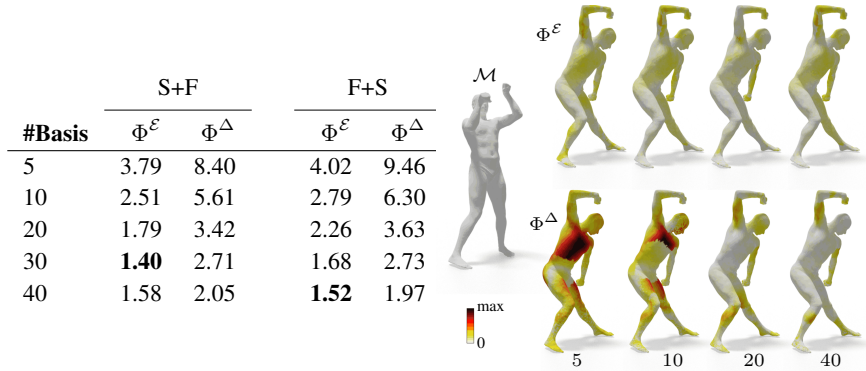


Figure 3: We investigate the best possible matching achievable with our embedding. On the left, we report the error varying the embedding dimension, comparing to the standard LBO basis Φ^Δ on the two settings S+F and F+S. Our basis always performs better. On the right, a qualitative example from F+S; the heatmap encodes the geodesic error of the correspondence.

5 Dataset and settings.

This section introduces the benchmarks and the parameters we select in our experiments.

Bijective. We namely refer to *bijective* setting when the shapes in the train and in the test share the same mesh and thus are in 1:1 correspondence. We consider two human datasets: SCAPE [3] (12500 vertices) and FAUST [5] (6890 vertices). As done in previous work [52], we train on one dataset and test on the other. We denote with **F+S** when we train on FAUST [3] and test on SCAPE [5]. **S+F** refers to the opposite. From now on, we adopt **+** to indicate training (on the left) and test set (on the right). Furthermore, we consider the SMAL dataset [66], composed of animal shapes with pose variations and from five species (big cats, canines, ovine, bovine, and hippos). The training set comprises 25 shapes, which is tight compared to the standard ones. The test set includes 300 pairs of 25 shapes unseen at training time. This setting, namely **SMAL**, includes strong non-isometric pairs.

Non-bijective. In this scenario, we take pairs of shapes that do not share the same connectivity at training or inference time. We indicate with \tilde{S} when they come from SCAPE, one at a full resolution and one remeshed with the 25% of the vertices. Similarly, we use \tilde{F} for FAUST.

Different densities between train and test. Finally, with $\widetilde{S+F}$, we denote the case in which, at training time, both shapes from SCAPE have a quarter of vertices, but at test time, we used the full resolution shapes from F . In this case, the discretization density at training differs from one of the test shapes. We choose this setting to emphasize the non-bijectivity since F is less isometric than S .

Settings. If not differently stated, we train our embedding networks to output 40-dimensional embeddings. If descriptors are used, they are 80 functions obtained as described in Section 3. Since [22] does not describe the relationship between basis and descriptors cardinalities, this choice is motivated by an analysis that we report in the Supplementary Material. In our quantitative evaluation, we compute the average of the geodesic error measured for each point as the geodesic distance between the estimated correspondence and the ground truth one. To facilitate the comparison, we visualize all the shapes with a similar orientation, but several rigid transformations in the 3D space occur among the shapes both at training and test time. During training, we do data augmentation by applying rotations on y -axis. We refer to the Supplementary Material for details on datasets, evaluation, training, and testing procedures.

6 Results

6.1 Analysis of the learned embeddings

We study the properties of our learned embedding, and we note that: i) It is always full rank, despite not being orthogonal (unlike the LBO basis, which is orthogonal by construction); ii) It is smooth. LBO eigenfunctions form the smoothest orthonormal basis. In Figure 2, we depict the embedding dimensions as functions over the surfaces and report their Dirichlet energy. Even if we do not explicitly impose smoothness, our embedding is significantly smoother than LBO one (smaller E_D); this is possible since we do not impose orthogonality.

Basis dimension. First, we compare our embedding $\Phi^\mathcal{E}$ with the LBO one Φ^Δ . Moreover, we evaluate how the performance depends on the embedding dimension. We train 5 embedding networks with a different number of output dimensions (i.e., 5, 10, 20, 30, and 40) both on $S + F$ and $F + S$. On the test dataset, to focus only on the embedding properties, we compute for each embedding the optimal transformation between the 50 pairs as done in Equation 1, and we estimate the correspondence solving the nearest neighbor search (NN) as $\Pi_{\mathcal{MN}} = NN(\Phi_{\mathcal{M}} C_{\mathcal{NM}}^{opt}, \Phi_{\mathcal{N}})$. Even using the ground truth $\Pi_{\mathcal{MN}}^{gt}$ to compute $C_{\mathcal{NM}}^{opt}$, the embeddings may present some differences which are not linearly alignable. We measure the error generated by this loss of information. This choice makes our comparison independent of the estimation of the linear transformation.

The Table on the left of Figure 3 contains the results. The first observation is that increasing the embedding size improves both $\Phi^\mathcal{E}$ and Φ^Δ representations. Secondly, we can compress the LBO representation to significantly smaller dimensions (i.e., with 5 learned dimensions, we have the same quality of 20 LBO basis; with 10 learned, the same of 30 LBO). But finally, and most importantly, our basis consistently outperforms the LBO basis. On the right of Figure 3, we also visualize the error of the two competitors. With small bases, the Φ^Δ cannot distinguish between the back and front of the shape, and at higher dimensions, the error peak at the protrusions. These are classical errors for Φ^Δ , which suffers the front-back symmetry and the missing details on the tiny components such as arms and legs due to the low pass representation. Learned embedding distributes the error more homogeneously over the surface.

Rotation invariance In shape matching, the rotation invariance is a desirable property. Our data augmentation is an effective technique to build resilience to rotations without giving up extrinsic information. To clarify this, we considered a pair of test shapes in the F+S setting. We fixed one of the two, applying 50 different random 3D rotations to the second. The error without rotations is 2.43, while the mean error after the rotations is 2.52 with a standard deviation of 0.11. Moreover, in the inset Table, we input 50 scales of HKS (a rotation-invariant descriptor) to train a 40-dimensional embedding as done in [52]. Even if $\Phi_{HKS}^\mathcal{E}$ does not overcome $\Phi^\mathcal{E}$ trained on coordinates, it is better than LBO, showing that our method could work with different input features.

	F+S
$\Phi^\mathcal{E}$	1.58
$\Phi_{HKS}^\mathcal{E}$	1.81
Φ^Δ	2.05

Comparison of learned embeddings. We compare $\Phi^\mathcal{E}$ and Φ^Δ to learned embeddings that exploit: i) a different architecture (L-Invariant obtained from [22]) and ii) a different loss (with $L_{\ell_1+E_D}$ and without $L_\mathcal{E}$). Again, we adopt the ground truth linear transformation C^{opt} highlighting the performance of each alternative in the optimal setting. In Table 1, we report the results for the matching task. As expected, the embeddings learned with the supervised loss $L_\mathcal{E}$ perform better. However, we note that $L_{\ell_1+E_D}$ preserves some matching capacity in the F+S setting, highlighting interesting relations between matching and general properties, which merit further explorations as future work. Finally, L-Invariant reveals that [22] cannot overtake Φ^Δ on meshes.

	S+F	F+S	SMAL
$\Phi^\mathcal{E}$	1.58	1.52	1.8
Φ^Δ	2.05	1.97	4.3
L-Invariant	2.94	2.85	2.4
$L_{\ell_1+E_D}$	13.18	5.72	14.9

Table 1: Quantitative comparison on shape matching, learning a 40 dimensional embedding and computing the ground truth transformation C^{opt} . Each row is a different training.

Regularizations in the bijective scenario. Noticing the potential of the learned embedding, we decided to explore additional regularizations during the training, starting from the bijective setting. Table 2 shows the comparisons obtained combining $L_\mathcal{E}$ with one or more losses from Section 4.

	S+F	F+S	SMAL
$\Phi^\mathcal{E}$	1.58	1.52	1.8
$\Phi^\mathcal{E}_{+L_{E_D}}$	1.29	1.34	2.0
$\Phi^\mathcal{E}_{+L_\perp}$	15.35	23.10	19.0
$\Phi^\mathcal{E}_{+L_{w\perp}}$	1.19	1.41	1.4
$\Phi^\mathcal{E}_{+L_{E_D}+L_{w\perp}}$	1.19	1.35	5.9

Table 2: Quantitative comparison on different regularizations, learning a 40 dimensional embedding, and computing the ground truth transformation C_{opt} . Each row is a different training.

	$\tilde{S}+F$	$\tilde{S}+\tilde{F}$	$\tilde{F}+S$	$\tilde{F}+\tilde{S}$	$\widetilde{S+F}$
Φ^Δ	2.05	1.88	1.97	2.04	2.05
$\Phi^\mathcal{E}$	1.12	0.76	1.24	1.27	1.07
$\Phi^\mathcal{E}_{+L_{E_D}}$	1.10	0.77	1.23	1.27	1.07
$\Phi^\mathcal{E}_{+L_{w\perp}}$	1.60	1.61	1.50	1.55	1.14
$\Phi^\mathcal{E}_{+L_{E_D}+L_{w\perp}}$	1.77	1.76	1.58	1.61	1.19

Table 3: Quantitative comparison on shape matching, learning a 40 dimensional embedding, and computing the ground truth transformation C_{opt} . Each row is a different training.

Promoting the smoothness ($\Phi^\mathcal{E}_{\ell+L_{E_D}}$) produces a slight improvement or keeps stable results both with and without L_\perp . Our analysis anticipated this result since $L_\mathcal{E}$ produces smoothness as a good property for the matching. Including L_\perp instead is disruptive due to its magnitude, which strongly bounds the network to favour the bijectivity induced by L_\perp over good matching. We thus scale L_\perp to a magnitude similar to L_{E_D} . With $\Phi^\mathcal{E}_{+L_{w\perp}}$, the network obtained a significant gain. We highlight that SMAL is the only dataset where smoothness seems slightly dangerous. We attribute this drop to the discontinuous self-intersections and artifacts present on the muzzle of animals.

Regularizations in challenging scenarios. Given the previous results, we verify our intuitions considering challenging contexts without bijection. Unlike previous works (that considered remeshed datasets where shapes have a similar density), we consider configurations with a significant change in the number of vertices. In Table 3, we train and test on different scenarios as described in Section 5. Once again, the smoothness does not change the performance. Adding orthogonality confirms our hypothesis: promoting bijectivity when this is not present worsen the results. Finally, in the last column, even if the shapes inside each dataset are in bijection, orthogonality does not help in the presence of huge density variations.

6.2 Applicative test

Learning descriptors Many previous works that we discussed in Section 2, validate their matching performance on the S+F and F+S settings. We report in the Supplementary Material the results of our approach on them, learning also the descriptors. Here we focus on more exciting challenges: non-isometric shapes from SMAL, in which the considered animals may have a significant difference in their surface metric. In the Table on the left of Figure 4, we report the results.

Competitors We considered four competitors in our analysis. *Universal* shares our same architecture but it does not foster a linear transformation to align them and is optimized with a different loss:

$$L_U(\Phi_{\mathcal{M}}^{uni}, \Phi_{\mathcal{N}}^{uni}) = \sum \|\Phi_{\mathcal{M}}^{uni} - \Pi_{\mathcal{M}\mathcal{N}}^{gt} \Phi_{\mathcal{N}}^{uni}\|_2^2. \quad (9)$$

L_U promotes universal signatures for the points in a similar fashion as the descriptor-based methods described in Section 2. This comparison is essential to highlight the merits of considering linear alignable embeddings. The other main competitor is [52], referred to as GFM_{DN} . As done in our pipeline, this method adopts the same architecture to learn the features, but it exploits the Φ^Δ instead of $\Phi^\mathcal{E}$. This comparison highlights the contributions of the learned embedding with respect to Φ^Δ . We will also consider two other methods to align Φ^Δ : i) GFM_{KP} [12], that uses a KPConv backbone [56], and ii) *Deep Shells* [13] that relies on a registration approach. For completeness, we also report the results of *L-Invariant* with its PointNet network [22]. In the experiments, we measure the error as the geodesic distance between the predicted match and the ground truth, and we report the average over all the points. While our analysis mainly focuses on theoretical properties of learned embeddings, we would emphasize contexts in which they show significant applicability. Not only we can learn descriptors that surpass state-of-the-art methods, but we also obtain better performance than the *theoretically best possible results with LBO*. Further preliminary evidence on descriptor learning, matching refinement, and functional reconstruction can be found in Supplementary Material.

	SMAL	
	ΦC	ΦC^{opt}
$Our^{\mathcal{E}}$	5.67	1.8
$Our^{\mathcal{E}}_{+L_{ED}}$	11.0	2.0
$Our^{\mathcal{E}}_{+L_{w\perp}}$	3.9	1.4
$Our^{\mathcal{E}}_{+L_{ED}+L_{w\perp}}$	11.5	5.9
Universal	5.91	4.3
GFM_{DN}	7.94	4.3
GFM_{KP} [12]	8.00	4.3
Deep Shells [13]	15.23	4.3
L-Invariant [22]	21.10	4.3



Figure 4: On the left, a quantitative comparison for non-isometric shape matching on SMAL dataset; the first column is learning also the descriptor, while the second is the theoretically best possible matching with that embedding using the optimal transformation C^{opt} . On the right, a pair of animals from TOSCA [7] dataset, showing our generalization capability.

We notice that, without any further regularization, we can recover good descriptors even for $\Phi^{\mathcal{E}}$. Once again, we observe that smoothness is not particularly favorable in this context. Finally, $Our^{\mathcal{E}}_{+L_{w\perp}}$ gives rise to the most exciting result, confirming our theoretical and empirical insights and outperforming the best possible performance achievable by LBO (that is 4.3, from Table 1).

Computational specifications We use an i9-9820X Intel processor at 3.31GHz and an RTX 2080 Ti GPU card for the experiments. Embedding training requires ~ 180 min for the F+S setting and 210min for S+F, regardless of the number of basis functions.

7 Conclusions

Insights. The observed saturation on the basis sets of increasing dimensions is caused by different reasons. Including more Φ^{Δ} basis functions introduces instability due to high frequencies that are hardly alignable linearly. Instead, $\Phi^{\mathcal{E}}$, being a learning method, is prone to overfitting if the representation is too rich. Smoothness seems an emerging natural property from matching. Intuitively, a linear alignment is simpler to achieve if there are no outliers or drastic changes; also, in our cases, the correspondence is a smooth function over the surface. However, if the data present discontinuities, the smoothness becomes dangerous. Orthogonality seems advantageous in contexts where bijectivity is a good prior. On the other hand, it requires proper weighting and could harm the results even if training and tests are not bijective. We believe this latter is a consequence of the global nature of orthogonality, which limits its adaptation if shapes change their number of vertices. We highlight that fully intrinsic approaches are rotation invariant but do not distinguish symmetries and may unrecoverably degrade the performance. Finally, our non-isometric and applicative domain results highlight the importance of having a proper basis to further push the matching results.

Limitations. While our work expands the understanding of learned embeddings, several questions remain. Investigating other classes of transformations different from the linear one is still entirely open. Other challenges like clutter and partial-to-partial matching (touched on in the Supplementary Materials) may require sophisticated strategies. Finally, cases in which the correspondence between two shapes is highly discontinuous would open to a completely different and intriguing analysis.

Future directions. In recent years, several methods have relied on refining the matching using ZoomOut [27]. This is a significantly important aspect for real applications, and it may seem a drawback of learned embedding of which dimension must be decided at training time. In the Supplementary Materials, we show that thanks to their smoothness, the learned basis can be augmented by ZoomOut, refining the matching by including the LBO eigenfunctions at increasing frequencies. These results are novel and exciting, but even more importantly, they pave the way for further exploration of possible refinement techniques designed explicitly for learned embeddings. Moreover, we focus on the shape matching task, but the learned embeddings can target other applications. In the

Supplementary Materials, we collect some coordinate function transfer results, and we foster other possible tasks as fascinating future work.

Our work shows a series of promising properties. We wish to motivate the community to foster research in this direction. Pointing out that smoothness seems a superior property to orthogonality in some contexts sheds new light on the study of shape embedding for matching, which for many years focused in a contrary direction [29, 41]. Generalization to challenging scenarios is appealing and highlights the importance of proper embedding instead of refining existing ones. Our joint analysis of learned and axiomatic embedding opens to look for other properties to impose and a broad set of different applications.

Acknowledgements This work is supported by the ERC Starting Grant No. 802554 (SPECGEO), the SAPIENZA BE-FOR-ERC 2020 Grant (NONLINFMAPS), KAUST OSR Award No. CRG-2017-3426, the ERC Starting Grant No. 758800 (EXPROTEA) and the ANR AI Chair AIGRETTE.

References

- [1] Scan the world project. <https://www.myminifactory.com/scantheworld>. [Online; accessed 16-November-2021].
- [2] Abubakar Abid, Ada Poon, and James Zou. Linear regression with shuffled labels. *arXiv preprint arXiv:1705.01342*, 2017.
- [3] 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.
- [4] Mathieu Aubry, Ulrich Schlickewei, and Daniel Cremers. The wave kernel signature: A quantum mechanical approach to shape analysis. In *Computer Vision Workshops (ICCV Workshops), 2011 IEEE International Conference on*, pages 1626–1633. IEEE, 2011.
- [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.
- [6] Alexander M Bronstein, Michael M Bronstein, and Ron Kimmel. Generalized multidimensional scaling: a framework for isometry-invariant partial surface matching. *Proceedings of the National Academy of Sciences*, 103(5):1168–1172, 2006.
- [7] Alexander M Bronstein, Michael M Bronstein, and Ron Kimmel. *Numerical geometry of non-rigid shapes*. Springer Science & Business Media, 2008.
- [8] Michael M Bronstein, Alexander M Bronstein, Ron Kimmel, and Irad Yavneh. Multigrid multidimensional scaling. *Numerical linear algebra with applications*, 13(2-3):149–171, 2006.
- [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.
- [10] Yoni Choukroun, A. Shtern, A. Bronstein, and R Kimmel. Hamiltonian operator for spectral shape analysis. *arXiv:1611.01990*, 2016.
- [11] Luca Cosmo, Emanuele Rodola, Michael M Bronstein, Andrea Torsello, Daniel Cremers, and Y Sahillioglu. Shrec’16: Partial matching of deformable shapes. *Proc. 3DOR*, 2(9):12, 2016.
- [12] Nicolas Donati, Abhishek Sharma, and Maks Ovsjanikov. Deep geometric functional maps: Robust feature learning for shape correspondence. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2020.
- [13] Marvin Eisenberger, Aysim Toker, Laura Leal-Taixé, and Daniel Cremers. Deep shells: Un-supervised shape correspondence with optimal transport. In H. Larochelle, M. Ranzato, R. Hadsell, M. F. Balcan, and H. Lin, editors, *Advances in Neural Information Processing Systems*, volume 33, pages 10491–10502. Curran Associates, Inc., 2020.
- [14] Danielle Ezuz and Mirela Ben-Chen. Deblurring and denoising of maps between shapes. *Computer Graphics Forum*, 36(5):165–174, 2017.
- [15] David A Field. Laplacian smoothing and delaunay triangulations. *Communications in applied numerical methods*, 4(6):709–712, 1988.
- [16] Thibault Groueix, Matthew Fisher, Vladimir G Kim, Bryan C Russell, and Mathieu Aubry. 3d-coded: 3d correspondences by deep deformation. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pages 230–246, 2018.

- [17] Xiaoshui Huang, Guofeng Mei, and Jian Zhang. Feature-metric registration: A fast semi-supervised approach for robust point cloud registration without correspondences. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2020.
- [18] Artiom Kovnatsky, Michael M Bronstein, Alexander M Bronstein, Klaus Glashoff, and Ron Kimmel. Coupled quasi-harmonic bases. In *Computer Graphics Forum*, volume 32, pages 439–448. Wiley Online Library, 2013.
- [19] Nathan Linial. Finite metric spaces–combinatorics, geometry and algorithms. *arXiv preprint math/0304466*, 2003.
- [20] 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.
- [21] Sven Loncaric. A survey of shape analysis techniques. *Pattern recognition*, 31(8):983–1001, 1998.
- [22] Riccardo Marin, Marie-Julie Rakotosaona, Simone Melzi, and Maks Ovsjanikov. Correspondence learning via linearly-invariant embedding. In H. Larochelle, M. Ranzato, R. Hadsell, M. F. Balcan, and H. Lin, editors, *Advances in Neural Information Processing Systems*, volume 33, pages 1608–1620. Curran Associates, Inc., 2020.
- [23] Riccardo Marin, Arianna Rampini, Umberto Castellani, Emanuele Rodolà, Maks Ovsjanikov, and Simone Melzi. Spectral shape recovery and analysis via data-driven connections. *International journal of computer vision*, 129(10):2745–2760, 2021.
- [24] Simone Melzi. Sparse representation of step functions on manifolds. *Computers & Graphics*, 82:117–128, 2019.
- [25] Simone Melzi, Riccardo Marin, Pietro Musoni, Filippo Bardoni, Marco Tarini, and Umberto Castellani. Intrinsic/extrinsic embedding for functional remeshing of 3d shapes. *Computers & Graphics*, 88:1 – 12, 2020.
- [26] S. Melzi, R. Marin, E. Rodolà, U. Castellani, J. Ren, A. Poulencard, P. Wonka, and M. Ovsjanikov. Matching Humans with Different Connectivity. In Silvia Biasotti, Guillaume Lavoué, and Remco Veltkamp, editors, *Eurographics Workshop on 3D Object Retrieval*. The Eurographics Association, 2019.
- [27] 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.
- [28] Simone Melzi, Emanuele Rodolà, Umberto Castellani, and Michael Bronstein. Shape analysis with anisotropic windowed fourier transform. In *International Conference on 3D Vision (3DV)*, 2016.
- [29] Simone Melzi, Emanuele Rodolà, Umberto Castellani, and Michael Bronstein. Localized manifold harmonics for spectral shape analysis. *Computer Graphics Forum*, 37(6):20–34, 2018.
- [30] Simone Melzi, Riccardo Spezialetti, Federico Tombari, Michael M. Bronstein, Luigi Di Stefano, and Emanuele Rodola. Gframes: Gradient-based local reference frame for 3d shape matching. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 4629–4638. IEEE, June 2019.
- [31] Mark Meyer, Mathieu Desbrun, Peter Schröder, and Alan H Barr. Discrete Differential-Geometry Operators for Triangulated 2-Manifolds. In *Visualization and mathematics III*, pages 35–57. Springer, New York, NY, 2003.
- [32] Walter Nelson, Marinka Zitnik, Bo Wang, Jure Leskovec, Anna Goldenberg, and Roded Sharan. To embed or not: network embedding as a paradigm in computational biology. *Frontiers in genetics*, 10:381, 2019.
- [33] Dorian Nogneng and Maks Ovsjanikov. Informative descriptor preservation via commutativity for shape matching. In *Computer Graphics Forum*, volume 36, pages 259–267. Wiley Online Library, 2017.
- [34] 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):30:1–30:11, 2012.
- [35] Maks Ovsjanikov, Etienne Corman, Michael Bronstein, Emanuele Rodolà, Mirela Ben-Chen, Leonidas Guibas, Frederic Chazal, and Alex Bronstein. Computing and processing correspondences with functional maps. In *SIGGRAPH 2017 Courses*. 2017.
- [36] Sebastian Padó, Michel Galley, Dan Jurafsky, and Christopher D Manning. Robust machine translation evaluation with entailment features. In *Proceedings of the Joint Conference of*

- the 47th Annual Meeting of the ACL and the 4th International Joint Conference on Natural Language Processing of the AFNLP*, pages 297–305, 2009.
- [37] Gautam Pai, Jing Ren, Simone Melzi, Peter Wonka, and Maks Ovsjanikov. Fast sinkhorn filters: Using matrix scaling for non-rigid shape correspondence with functional maps. In *CVPR*, 2021.
 - [38] G Dias Pais, Srikumar Ramalingam, Venu Madhav Govindu, Jacinto C Nascimento, Rama Chellappa, and Pedro Miraldo. 3dregnet: A deep neural network for 3d point registration. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 7193–7203, 2020.
 - [39] Ulrich Pinkall and Konrad Polthier. Computing discrete minimal surfaces and their conjugates. *Experimental mathematics*, 2(1):15–36, 1993.
 - [40] 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.
 - [41] Arianna Rampini, Irene Tallini, Maks Ovsjanikov, Alex M Bronstein, and Emanuele Rodolà. Correspondence-free region localization for partial shape similarity via hamiltonian spectrum alignment. In *International Conference on 3D Vision (3DV)*, 2019.
 - [42] Vikram Ravindra, Huda Nassar, David F Gleich, and Ananth Grama. Rigid graph alignment. In *International conference on complex networks and their applications*, pages 621–632. Springer, 2019.
 - [43] Jing Ren, Simone Melzi, Maks Ovsjanikov, and Peter Wonka. Maptree: Recovering multiple solutions in the space of maps. *ACM Trans. Graph.*, 39(6), Nov. 2020.
 - [44] 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.
 - [45] Martin Reuter, Silvia Biasotti, Daniela Giorgi, Giuseppe Patanè, and Michela Spagnuolo. Discrete laplace–beltrami operators for shape analysis and segmentation. *Computers & Graphics*, 33(3):381–390, 2009.
 - [46] Emanuele Rodolà, Samuel Bulò, Thomas Windheuser, Matthias Vestner, and Daniel Cremers. Dense non-rigid shape correspondence using random forests. In *2014 IEEE Conference on Computer Vision and Pattern Recognition*, pages 4177–4184, 2014.
 - [47] Emanuele Rodolà, Luca Cosmo, Michael M Bronstein, Andrea Torsello, and Daniel Cremers. Partial functional correspondence. In *Computer Graphics Forum*, volume 36, pages 222–236. Wiley Online Library, 2017.
 - [48] Raif M Rustamov. Laplace-beltrami eigenfunctions for deformation invariant shape representation. In *Proceedings of the fifth Eurographics symposium on Geometry processing*, pages 225–233. Eurographics Association, 2007.
 - [49] Yusuf Sahillioğlu. Recent advances in shape correspondence. *The Visual Computer*, 36(8):1705–1721, 2020.
 - [50] Rodrigo Santa Cruz, Basura Fernando, Anoop Cherian, and Stephen Gould. Neural algebra of classifiers. In *2018 IEEE Winter Conference on Applications of Computer Vision (WACV)*, pages 729–737. IEEE, 2018.
 - [51] Vinit Sarode, Xueqian Li, Hunter Goforth, Yasuhiro Aoki, Rangaprasad Arun Srivatsan, Simon Lucey, and Howie Choset. Pcnnet: Point cloud registration network using pointnet encoding. *arXiv preprint arXiv:1908.07906*, 2019.
 - [52] Nicholas Sharp, Souhaib Attaiki, Keenan Crane, and Maks Ovsjanikov. Diffusion is all you need for learning on surfaces. *arXiv preprint arXiv:2012.00888*, 2020.
 - [53] Xiaofei Shi and Yanghua Xiao. Modeling multi-mapping relations for precise cross-lingual entity alignment. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, pages 813–822, Hong Kong, China, Nov. 2019. Association for Computational Linguistics.
 - [54] Dmitriy Smirnov and Justin Solomon. Hodgenet: learning spectral geometry on triangle meshes. *ACM Transactions on Graphics (TOG)*, 40(4):1–11, 2021.
 - [55] Jian Sun, Maks Ovsjanikov, and Leonidas Guibas. A concise and provably informative multi-scale signature based on heat diffusion. *Computer Graphics Forum*, 28(5):1383–1392, 2009.
 - [56] Hugues Thomas, Charles R. Qi, Jean-Emmanuel Deschaud, Beatriz Marcotegui, Francois Goulette, and Leonidas J. Guibas. Kpconv: Flexible and deformable convolution for point clouds. In *The IEEE International Conference on Computer Vision (ICCV)*, October 2019.

- [57] Federico Tombari, Samuele Salti, and Luigi Di Stefano. Unique signatures of histograms for local surface description. In *Proc. ECCV*, pages 356–369. Springer, 2010.
- [58] Warren S Torgerson. Multidimensional scaling: I. theory and method. *Psychometrika*, 17(4):401–419, 1952.
- [59] Oliver Van Kaick, Hao Zhang, Ghassan Hamarneh, and Daniel Cohen-Or. A survey on shape correspondence. In *Computer graphics forum*, volume 30, pages 1681–1707. Wiley Online Library, 2011.
- [60] Chenxu Wang, Zhiyuan Zhao, Yang Wang, Dong Qin, Xiapu Luo, and Tao Qin. Deepmatching: A structural seed identification framework for social network alignment. In *2018 IEEE 38th International Conference on Distributed Computing Systems (ICDCS)*, pages 600–610, 2018.
- [61] Yue Wang and Justin M Solomon. Deep closest point: Learning representations for point cloud registration. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 3523–3532, 2019.
- [62] Zhichun Wang, Qingsong Lv, Xiaohan Lan, and Yu Zhang. Cross-lingual knowledge graph alignment via graph convolutional networks. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, pages 349–357, Brussels, Belgium, Oct.-Nov. 2018. Association for Computational Linguistics.
- [63] Zhirong Wu, Shuran Song, Aditya Khosla, Fisher Yu, Linguang Zhang, Xiaoou Tang, and Jianxiong Xiao. 3d shapenets: A deep representation for volumetric shapes. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1912–1920, 2015.
- [64] Ruihan Yang, Dingsu Wang, Ziyu Wang, Tianyao Chen, Junyan Jiang, and Gus Xia. Deep music analogy via latent representation disentanglement. *arXiv preprint arXiv:1906.03626*, 2019.
- [65] Wei Zheng, Le Yang, Robert J Genco, Jean Wactawski-Wende, Michael Buck, and Yijun Sun. Sense: Siamese neural network for sequence embedding and alignment-free comparison. *Bioinformatics*, 35(11):1820–1828, 2019.
- [66] Silvia Zuffi, Angjoo Kanazawa, David Jacobs, and Michael J. Black. 3D menagerie: Modeling the 3D shape and pose of animals. In *IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, July 2017.

A Appendix

Given a network \mathcal{N} that takes a shape S and produces a d -dimensional embedding for every vertex on S . Stacking these values into a matrix (where rows correspond to vertices) we get a *basis matrix* Φ_S .

Now, suppose the network \mathcal{N} is trained so that for any input shape S , the basis matrix $\Phi_S = \mathcal{N}(S)$ should be approximately orthonormal: $\Phi_S^\top \Phi_S \approx Id$. We will call such a network \mathcal{N} an *orthonormal feature network*.

In a simpler scenario, suppose that we use the trained network to compute dense correspondences across shape pairs using nearest neighbor search in embedding space. Thus, given two shapes S_1, S_2 , we first compute the basis matrices $\Phi_1 = \mathcal{N}(S_1)$ and $\Phi_2 = \mathcal{N}(S_2)$. We then compute a map by solving the problem:

$$\Pi_{S_1 S_2} = \arg \min_{\Pi \in \mathcal{P}} \|\Phi_1 - \Pi \Phi_2\|^2. \quad (10)$$

Here \mathcal{P} is the space of all binary matrices that represent point-to-point maps: i.e., binary matrices with exactly one 1 per row, and we use the Frobenius norm. Note that the optimization problem in Eq.(10) is row separable, meaning that it can be solved via nearest neighbor search in the L_2 sense.

Lemma A.1. *If the basis matrices are exactly orthonormal and the residual when solving for the map in Eq. (10) is zero, then the optimal map $\Pi = \Pi_{S_1 S_2}$ must satisfy $\Phi_2^\top \Pi^\top \Pi \Phi_2 = Id$.*

Proof. By assumption we have $\Phi_1 = \Pi \Phi_2$. Moreover since $\Phi_1^\top \Phi_1 = Id$, left-multiplying the first equation by Φ_1^\top , we get: $\Phi_1^\top \Phi_1 = \Phi_1^\top \Pi \Phi_2$, so that $\Phi_1^\top \Pi \Phi_2 = Id$. Using again the fact that $\Phi_1 = \Pi \Phi_2$ we get $\Phi_2^\top \Pi^\top \Pi \Phi_2 = Id$. \square

Lemma A.2. *Under the assumptions of Lemma A.1, if, moreover, the dimensionality d of the basis matrices equals n , the number of points on S_1 , then the optimal map $\Pi = \Pi_{S_1 S_2}$ must be orthonormal, and thus represent a bijective map.*

Proof. If $d = n$ then Φ_2 is a square matrix, and since it is orthonormal, it must have an inverse $\Phi_2^{-1} = \Phi_2^\top$. Using this fact and applying Lemma A.1 we get $\Pi^\top \Pi = Id$. Since Π is, by assumption a binary matrix, it must therefore be doubly stochastic and represent a bijective map. \square

Lemma A.3. *More generally, for an arbitrary dimension d of the basis matrix, suppose that there is a subset of vertices $S \subseteq S_2$ such that the indicator function I_S lies within the span of the basis matrix Φ_2 . I.e., $I_S = \Phi_2 \mathbf{a}$ for some vector of coefficients \mathbf{a} . Then, again under the assumptions of Lemma A.1, we have $\|\Pi I_S\| = \|I_S\|$. Thus the map Π must preserve the cardinality of the set I_S , or, in other words, Π **must be injective** when restricted to S .*

Proof. We have $\|I_S\| = \mathbf{a}^\top \Phi_2^\top \Phi_2 \mathbf{a} = \mathbf{a}^\top \mathbf{a}$. On the other hand, $\|\Pi I_S\| = \mathbf{a}^\top \Phi_2^\top \Pi^\top \Pi \Phi_2 \mathbf{a} = \mathbf{a}^\top \mathbf{a}$ using Lemma A.1. \square

Lemma A.4. *In particular, if the feature matrix Φ_2 contains the constant function, regardless of the dimensionality, then under the assumptions of Lemma A.3, Π must be injective on the entire shape. Moreover, if size of Φ_1, Φ_2 is the same, then the map must be bijective.*

Proof. The first part follows directly from Lemma A.3. For the second part, observe that if $\|\Pi I_S\| = \|I_S\|$ then, if I_S is the indicator of the entire shape, the cardinality of ΠI_S is the same as the number of rows in Φ_1 , which means that the map is also surjective, and thus, must be bijective. \square

Supplementary Materials

Abstract

Here we list additional material as support to the main manuscript. We provide more details about the training processes and hyperparameters choice. We also offer an analysis on the relationship between the cardinality of basis and descriptors, quantitative results in isometric cases, and qualitative results both on humans and non-isometric animals. We describe more in detail the partiality setting, and finally, we include some more examples of ZoomOut process.

1 Training Details

We compose the Diffusion networks with 12 Diffusion Blocks in all cases. In the diffusion process, we use 128 eigenvalues and eigenvectors. We use an ADAM optimizer with an initial learning rate of $10e^{-4}$ and a Cosine Annealing Warm Restarts scheduler. Due to computational reasons, we use batch size 2 and accumulate the gradient for 4 batches to simulate a batch size of 8. We set the maximum number of epochs to 1200. We will release our data and method implementation.

Datasets . We consider two human datasets: the 100 shapes of FAUST [5], a dataset of ten different subjects in ten different poses, and the 71 shapes of SCAPE [3], a single subject in 71 different poses. Both of them are used with their original connectivity. At test time, we evaluate on 50 random pairs of the testing dataset.

SMAL dataset comprises different animal groups (big cats, canines, ovine, bovine, and hippos), presenting significant non-isometries. The training set is composed of 25 shapes of different species and poses, that is definitely small if compared to the ones generally adopted. The test set is composed of 300 pairs of 25 shapes unseen at training time.

1.1 Loss stability

In Figure 5, we report the behavior of training loss when we train our basis, our descriptor, and the universal embedding in the two settings of S+F and F+S. Both our networks show stability. We attribute the peaks to the Cosine Annealing scheduler restart, which increases the learning rate to escape from local minima. For this reason, we think our solution spaces are more regular than the one of the universal embedding.

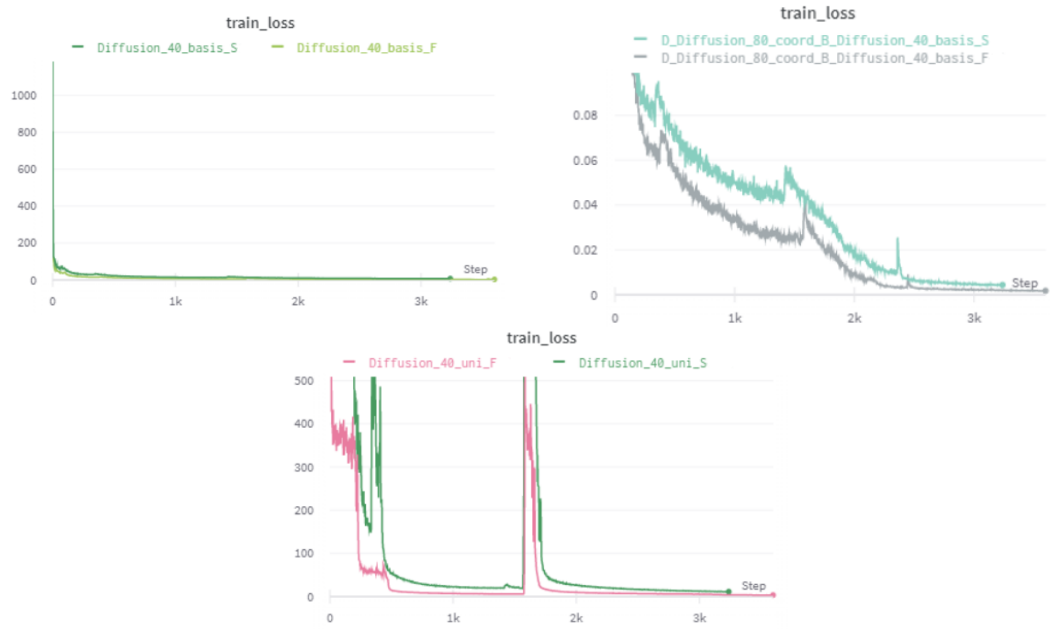


Figure 5: Training losses of our embedding (top left), features (top right) and Universal embedding formulation (bottom).

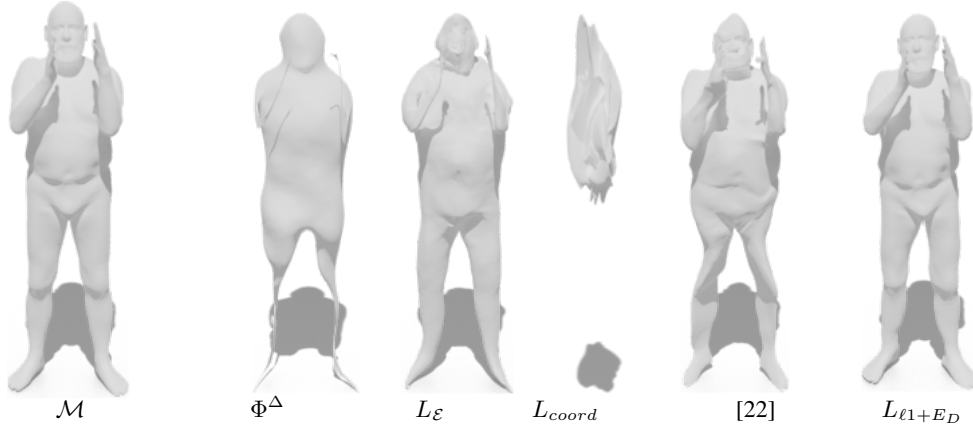


Figure 6: Coordinate reconstruction with different basis. Optimizing only the coordinates brings to a dramatical overfitting. LBO causes loss of information on the protrusions. Optimizing for the matching produce sharper results, while considering smoothness and sparsity may produce better detail recover.

2 Function representation and coordinates as proxies

Until now, the functional perspective has been carried only by the Functional Maps paradigm. How learned embeddings can represent functions was left unexplored by [22], but we think it is an important aspect, given the number of tools and analogies that it suggests. We focus on coordinates as a triplet of functions defined over the vertices. The interest in how the basis can represent such functions is multiple. Coordinates evolve smoothly across a shape, particularly in the human domain; hence, they are good examples of a continuous global function. Representing the coordinates of each point can be visualized, and it gives an interpretation of the loss of information. Finally, and importantly, reconstructing the coordinates means that the basis can represent the points on the surface. In this case, we considered three different trainings of $\Phi^\mathcal{E}$. In the first one, we used the $L_\mathcal{E}$ defined in Eq. 4 of the main manuscript. Then, we trained using $L_{\ell_1+E_D}$ presented in Equations 7 and 8 of the paper. To highlight the benefit of our mesh approach, we also compare [22] trained on $L_\mathcal{E}$ and $L_{\ell_1+E_D}$. Finally, we consider the case in which we train the features to represent the coordinates. We perform this training considering the following loss:

$$L_{coord}(\Phi^\mathcal{E}_\mathcal{M}) = \sum \|\Phi^\mathcal{E}_\mathcal{M} \Phi^{\mathcal{E}\dagger}_\mathcal{M} X_\mathcal{M} - X_\mathcal{M}\|_2^2, \quad (11)$$

which quantifies the error in the reconstruction of the coordinates functions $X_\mathcal{M}$, provided by the learned basis. In Table 4 we report the results of our experiment on F+S and S+F settings. Error is expressed as Equation 11. A qualitative example is reported in Figure 6.

Insights. Interestingly, learning for matching generates good bases also for function representation. The geometry of the shape is significantly more inflated than in the LBO case, in which the protrusions collapsed. $L_{\ell_1+E_D}$ also gave rise to an interesting result, in which we obtain a precise characterization, even in the finer details. This result shows that a functional basis that produces a good representation of the vertices does not imply an improved matching capability. The ad-hoc training with L_{coord} is the worst result, probably due to a lack of generalization: learning to reproduce specific functions for a small set of shapes may cause a quick overfit. Giving structure to the embeddings seems promising in obtaining more general tools for shape analysis. We finally highlight that matching loss for L-Invariant trained with $L_\mathcal{E}$ produces worst results on F+S settings but a better result in S+F where the poses seen at training time are more comprehensive than the ones observed at test time. The complete absence of intrinsic information in [22] makes it harder to generalize on new poses. Finally, L-Invariant cannot be trained with $L_{\ell_1+E_D}$; we believe this is a strong motivation for our approach.

	F+S	S+F
$\Phi_\mathcal{E}$	2.86	2.26
$\Phi_{\ell_1+E_D}$	1.07	0.89
Φ_{Coord}	47.58	45.90
Φ^Δ	2.96	3.06
L-Invariant [22]	3.62	1.91
L-Invariant $L_{\ell_1+E_D}$	25.35	44.40

Table 4: Results in coordinates representation.

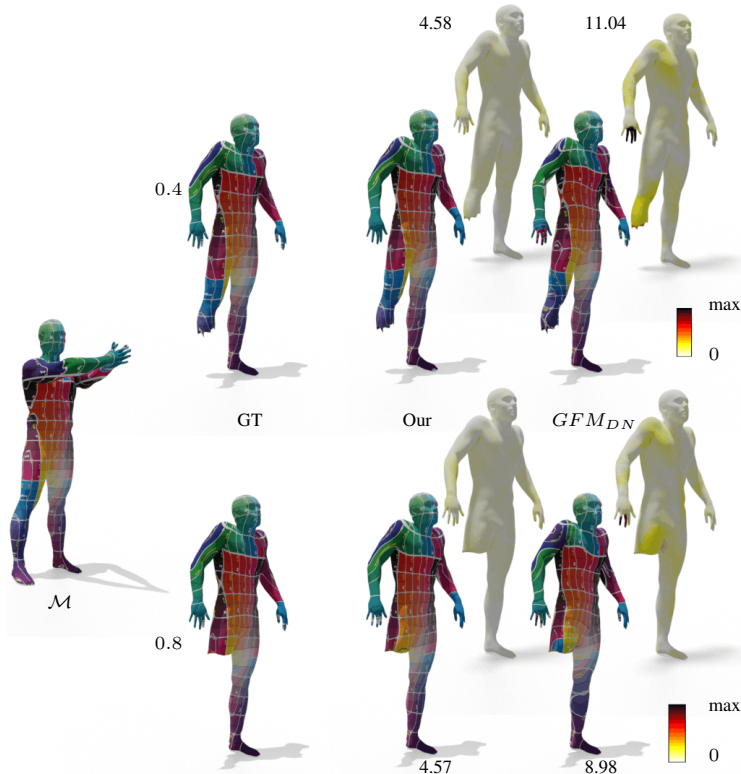


Figure 7: An example of partiality involved in our experiments. On the top, the 0.2 setting. On the bottom, the 0.4 one. Comparison between Our and GFM_{DN} . Near to each case we report the average geodesic error over the surface.

3 Partiality.

To test the resilience of learned features to missing geometry, we used $\Phi^{\mathcal{E}}$ trained on the *full* shapes of SCAPE (i.e., without any data augmentation with partial shapes). We constructed our dataset as follows:

1. We consider the 100 FAUST shapes.
2. For each shape, we take a landmark on the right foot.
3. We remove all surface that is within a certain geodesic distance to the landmark. We consider two different ranges: 0.4 and 0.8.

	0.0	0.4	0.8
$\Phi^{\mathcal{E}}C_{opt}$	1.58	6.8	10.5
$\Phi^{\Delta}C_{opt}$	2.05	7.3	10.6
Our $^{\mathcal{E}}$	3.80	8.4	12.0
GFM_{DN}	2.90	8.2	11.9

Table 5: Results with different partialities in S+F setting, training on full shapes.

Then, we kept the 50 pairs considered in the other experiments, substituting the source shape with the partial one. Hence, we look for a point on the complete shape for each point of the partial shape. Notice that the partial and the complete are different subjects in different poses. We want to remark that partiality has not to be seen at training time. We argue that this dataset is more interesting from an applicative perspective compared to SHREC16 [11]: the latter has been designed to analyze some specific theoretical properties of LBO basis. Instead, we are not aware of any other work that tackled the problem of missing limbs in human shape matching, also providing a more fair representation of different human beings. We show the results in Table 5; in the first two rows, we consider the optimal matching, while in the last two, we consider the one with the features obtained by \mathcal{F} . In Figures 7 and 8 we depict two qualitative examples. On the left, the full shape. The ground-truth, our and GFM_{DN} matchings in the 0.4 setting, with texture transfer and error on the top right. On the bottom right, the same for the 0.8 setting. The LBO bases are unstable, and the results of matching between 0.4 and 0.8 settings vary significantly even in regions far from the missing part (e.g., the right arm).

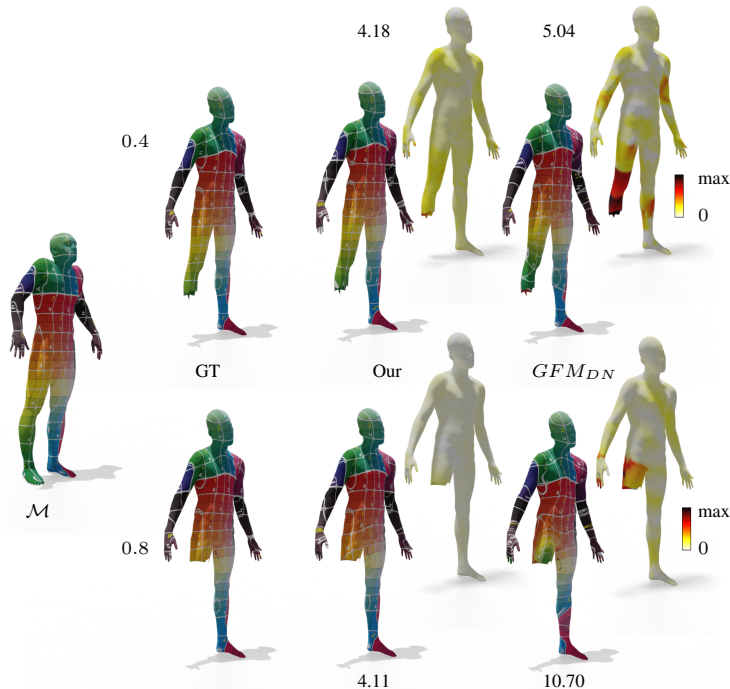


Figure 8: An example of partiality involved in our experiments. On the top, the 0.2 setting. On the bottom, the 0.4 one. Comparison between Our and $GFMDN$. Near to each case we report the average geodesic error over the surface.

	S+F				F+S			
	#Features				#Features			
	20	40	80	120	20	40	80	120
5 Basis	34.75	30.80	18.29	29.81	15.28	13.14	10.61	15.56
10 Basis	11.01	19.88	6.85	15.03	8.83	18.92	16.30	12.90
20 Basis	32.07	9.83	4.67	10.77	40.62	12.72	7.68	12.78
30 Basis	14.60	13.18	3.98	3.46	8.34	29.71	8.87	9.82
40 Basis	7.61	29.66	3.77	4.63	12.68	49.85	8.76	11.49

Table 6: Analysis with different number of basis embedding dimensions and features.

4 Matching

4.1 Basis and Descriptors relation

Given that [22] does not provide any insight on the relation between the basis and descriptors, we investigated this relation. In Table 6, we performed an extensive analysis with different combinations of embedding dimensions and a number of features. Each entry of Table 6 corresponds to a separate training. Concerning the dimensionality of the embedding, we notice that coherently with Figure 3 of the main manuscript, an increasing number of basis functions does not always produce better results. For example, in the F+S case, 40 basis functions tend to be hardly alignable. We believe this is mainly because the SCAPE dataset contains a wide variety of poses. Hence, overfitting FAUST by using a larger representation would make it difficult to generalize at test time.

Our analysis also gives a good rule of thumb for features: a number bounded between $\times 2$ and $\times 3$ seems the one producing the most stable results. We consider this number reasonable to overcome the noise and in line with the literature [2]. This analysis sheds light on the role of the two components of the matching pipeline; the embedding should provide a structure shared across the objects to be easily

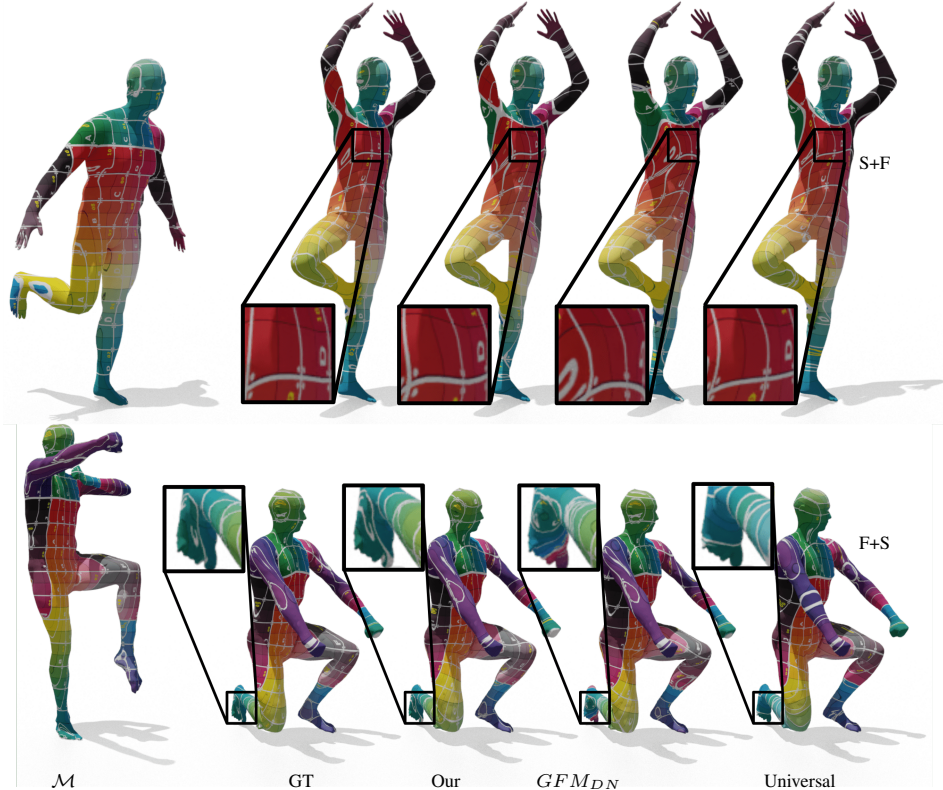


Figure 9: Texture transfer for matching quality comparison.

alignable. The features help to identify such structure. A complex structure is not desirable, while sometimes redundant information in the features can help to identify the correct transformation.

Given such analysis, we decide to keep 40 dimensions for the basis and 80 for the features for all our experiments. This choice provides coherence in the results and good performance in all the settings.

4.2 Near-Isometric matching

We tested our capability of matching near-isometric shapes using the learned embedding and features. In Table 7, we reported a comparison of our embeddings and main competitors in the S+F and F+S settings. For all the experiments, we kept a 40-dimensional embedding and 80. As can be seen, we perform better in F+S settings where generalization is relevant to obtain good results.

	S+F	F+S
Our ^E	3.8	8.8
GFM_{DN}	2.9	10.2
Universal	4.5	12.4

Table 7: Quantitative comparison on shape matching.

In Figure 9, we show texture transfer on S+F and F+S. While all methods produce an excellent matching, we observe a misalignment of high-frequency details.

4.3 Qualitative results

Here we report some additional qualitative results:

1. In Figure 10 we report another example of our matching on two statues. Notice that they do not share the same topology structure: despite this, we obtain a smooth matching.
2. In Figure 11 we report an example between two similar animals. Even in the isometric case (which is favorable to the LBO) we show better results.
3. In Figure 12 we have a significant non-isometric case; the error is localized on protrusions (ears, legs).



Figure 10: An example of extreme non-isometric matching. The two shapes are from [1].

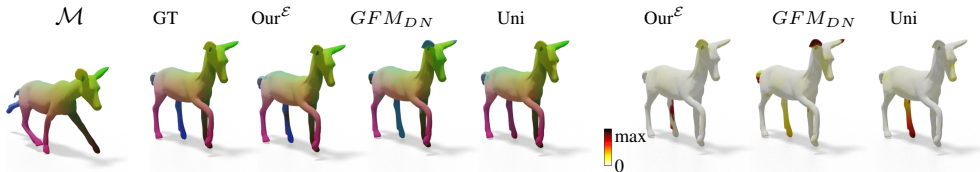


Figure 11: An isometric pair from the remeshed SMAL dataset. From the left: the source shape, the target shape with color transferred using $\Pi_{\mathcal{M}\mathcal{N}}^{gt}$. Then, the color transfer is performed by our, $GFMDN$, and the Universal Embedding Baseline. On the right, the geodesic error depicted over the surface.

4. In Figure 13 we show our results on three pairs of animals from the TOSCA dataset [7]. We matched highly non-isometric shapes to test our generalization capability. Notice that cats and horses are classes not seen at training time.
5. in Figures 14, 15, and 16 we show three more qualitative results. We would like to highlight that our methods seems to obtain better matching on protrusions (i.e., legs and arms).
6. in Figure 17 we show an example on two further different datasets. On the left, a pair from SHREC19 [26]; notice that the two shapes are different for their pose but also for their quality (i.e., the \mathcal{M} is a real scan, while the target one is a synthetic model). On the right, a woman is matched with a gorilla from TOSCA dataset [7]. We consider this case particularly extreme since arms and legs have entirely different proportions. Even if there are some evident artifacts on the stomach, we observe an overall coherence in the obtained matching.

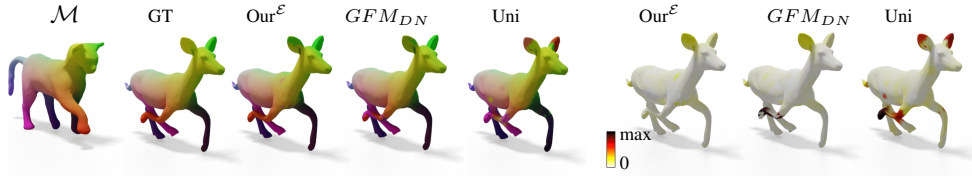


Figure 12: A non-isometric pair from the remeshed SMAL dataset. From the left: the source shape, the target shape with color transferred using $\Pi_{\mathcal{M}\mathcal{N}}^{gt}$. Then, the color transfer performed by our, GFM_{DN} , and the Universal Baseline. On the right, the geodesic error depicted over the surface.

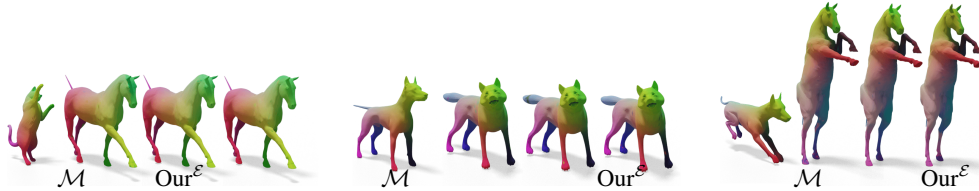


Figure 13: Three pair of animals from TOSCA [7] dataset, showing our generalization capability.

	S+F					F+S				
	Init	ZO10	ZO20	ZO40	ZO60	Init	ZO10	ZO20	ZO40	ZO60
$\mathbf{I}_{Our} \Phi^E$	3.77	3.18	3.16	3.18	3.21	8.76	7.33	7.20	7.30	7.30
$\mathbf{I}_{Our} \Phi^\Delta$		3.98	4.02	4.13	4.17		8.32	8.58	8.58	8.71
$\mathbf{I}_{GFM_{DN}} \Phi^E$	2.85	2.55	2.54	2.57	2.62	10.17	8.99	8.97	9.03	9.11
$\mathbf{I}_{GFM_{DN}} \Phi^\Delta$		2.96	3.04	3.12	3.18		9.63	9.72	10.03	10.08

Table 8: Matching results using ZoomOut as further refinement of an input matching. Replacing the first 40 Φ^Δ with Φ^E , we obtain a better matching even for the initialization provided by GFM_{DN} .

5 ZoomOut

5.1 Can we refine the learned embeddings?

One of the main advantages of a spectral embedding is its arbitrary dimensionality, which can be selected at test time. However, in [22] the basis dimension is fixed a priori. Inspired by this, we propose to incorporate ZoomOut [27] in our pipeline, considering our 40-dimensional embedding equivalent to the first 40 eigenfunctions of the LBO. Then, we apply ZoomOut by introducing LBO eigenfunctions (starting from the 41st). In the first two rows of Table 8, we initialize with the matching produced by our method the $C \in \mathbb{R}^{40 \times 40}$ (I_{Our}) exploiting our basis (1st row) or the first 40 LBO

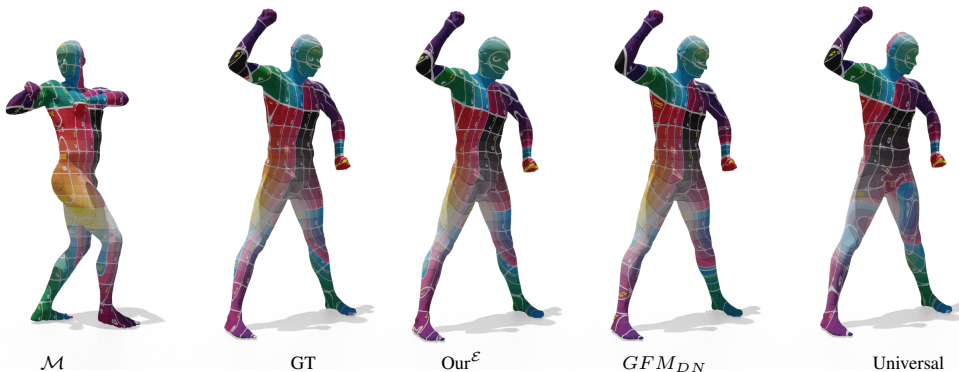


Figure 14: Texture transfer for matching quality comparison (F+S setting).

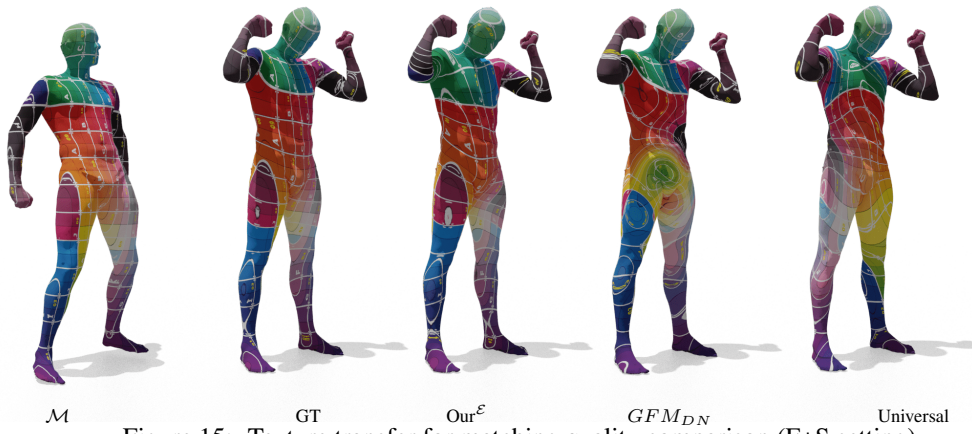


Figure 15: Texture transfer for matching quality comparison (F+S setting).

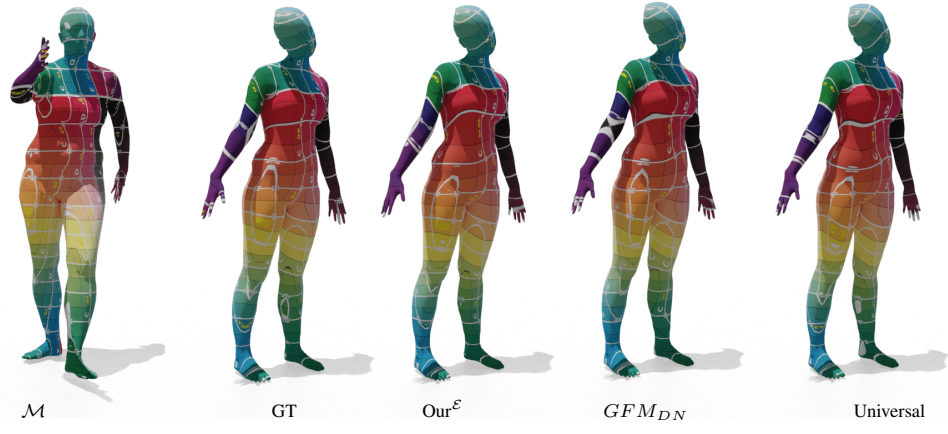


Figure 16: Texture transfer for matching quality comparison (S+F setting).

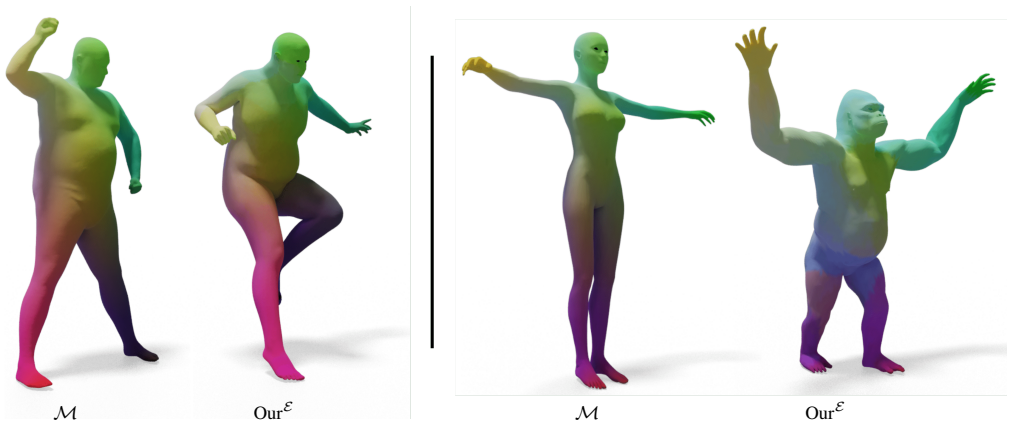


Figure 17: Color transfer for matching quality comparison on two out of distributions couples.

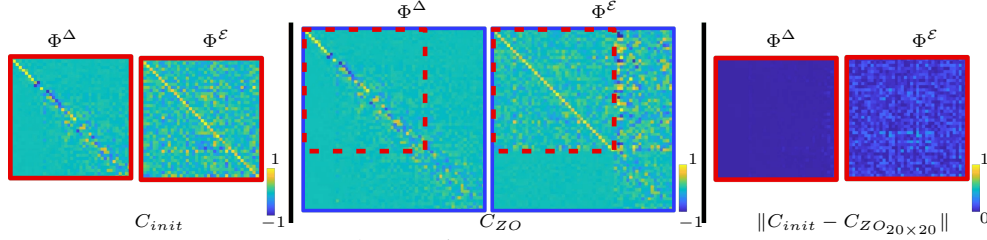


Figure 18: ZoomOut applied to Φ^ε and Φ^Δ . On the left, the two transformation matrices have been initialized by the same correspondence $\Pi_{\mathcal{MN}}$. Then, in the middle, we increased their size by 20 dimensions, using in both cases the LBO basis from 41 to 60. We show the absolute difference between the initialization and the new upper right part of the matrix on the right.

basis functions (2nd row). Then, we applied ZoomOut to increase the set with higher frequencies (adding 5 basis functions at each iteration in the ZoomOut pipeline). We observe that our basis can replace the low frequencies of the LBO. In the 3rd and 4th rows, we repeated the experiment initializing with the matching provided by $GFMDN$ (I_{GFMDN}), improving a matching optimized for another basis. ZoomOut shows instability for some dimensionality while, remarkably, our embedding seems more stable.

We depict in Figure 18 the different effects of ZoomOut on the linear transformation. Starting with $\Pi_{\mathcal{MN}}$, we obtain the 40×40 transformation using either our embedding ($C_{init}^{\Phi^\varepsilon}$) or the first 40 LBO basis functions ($C_{init}^{\Phi^\Delta}$). Then, we applied ZoomOut obtaining $C_{ZO}^{\Phi^\varepsilon}$ and $C_{ZO}^{\Phi^\Delta}$ of dimension 60×60 . We observe that $C_{ZO}^{\Phi^\varepsilon}$ presents a block structure: the top-left block highlights how Φ^ε have been recombined to match higher frequencies of Φ^Δ ; the bottom-left block is almost 0. On the right, we report the variations in the upper left 20×20 ; the Φ^Δ one is left unchanged, while our basis shows flexibility thanks to its non-orthogonality.

In Figures 19 and 20 we report two other cases of ZoomOut and their impact on the initialization matrix (the considered pairs are the ones of Figure 14 and 15 respectively). In the first one, we observe that the initial LBO matrix has a less diagonal behaviour than the latter. This is in general due to a non-isometric deformation that in this case, could be reasonably caused by a twist of the torso of Figure 14. The C_{init} seems to be almost preserved by the ZoomOut process. In the second case, the initialization is more diagonal, witnessing two almost isometrical shapes. In this case, the ZoomOut process can recombine a few parts of the LBO C_{init} , while this impacts mainly the last dimensions. However, in both cases, our matrix shows better flexibility. Also, in all experiments, we observe the block division of our C_{ZO} matrix discussed above.

Insights. This experiment shows that learned embeddings are a good initialization for refinement techniques and improve existing learning pipelines. The structure of the final C matrix also reveals the reason behind this. In the obtained matrix using only Φ^Δ , the top right and bottom left rectangles are almost empty, and the only significant interaction appears in the added frequencies. Considering the C produced with Φ^ε , this interaction is almost unchanged. However, the top right block shows the interaction between Φ^ε and higher frequencies. This opens to a new perspective in the field of shape matching: instead of seeking competition between representations, incorporating the two seems a promising direction. Note that our merging is naive, and more sophisticated combinations are possible; we leave this analysis for future work.

5.2 Step size ablation study

In Table 8, we reported the results by using ZoomOut and step size 5 (i.e., including 5 more basis at each ZoomOut iteration). Here, we report also results using a step of 1 (Table 9) and 2 (Table 10). We show that our basis is always a better choice to improve the matching. We noticed that by reducing the step size, the results at higher C dimensions become more unstable. We believe this is due to the more iterations required to reach the same C dimensions, introducing noise in the correspondence.

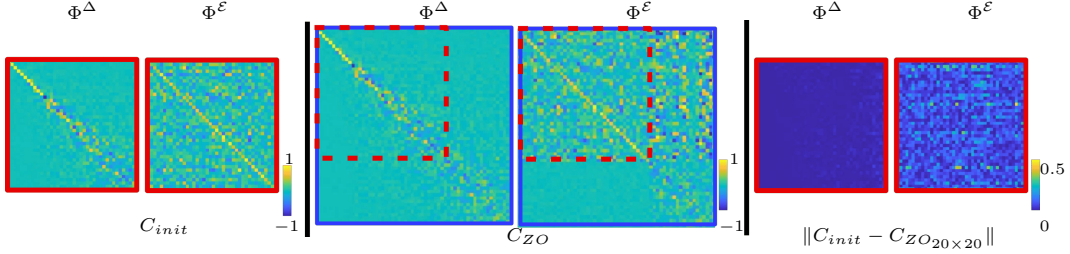


Figure 19: ZoomOut applied to Φ^E and Φ^Δ . On the left, the two transformation matrices have been initialized by the same correspondence $\Pi_{\mathcal{M}\mathcal{N}}$. Then, in the middle, we increased their size of 20 dimensions, using in both cases the LBO basis from 41 to 60. On the right, we show the absolute difference between the initialization and the new upper left part of the matrix. The considered example refers to the pair shown in Figure 14.

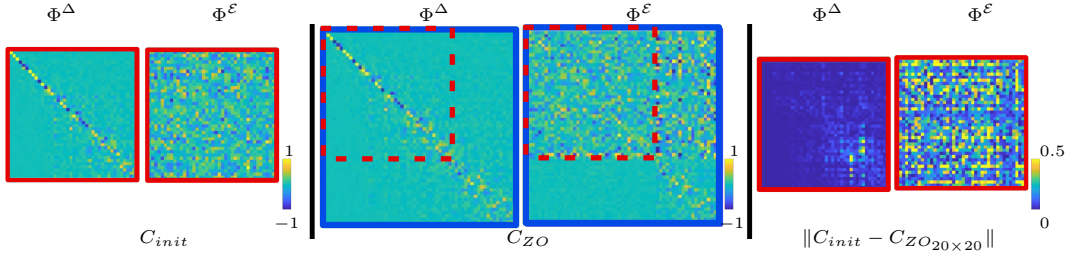


Figure 20: ZoomOut applied to Φ^E and Φ^Δ . On the left, the two transformation matrices have been initialized by the same correspondence $\Pi_{\mathcal{M}\mathcal{N}}$. Then, in the middle, we increased their size of 20 dimensions, using in both cases the LBO basis from 41 to 60. On the right, we show the absolute difference between the initialization and the new upper left part of the matrix. The considered example refers to the pair shown in Figure 15.

	S+F					F+S				
	Init	ZO10	ZO20	ZO40	ZO60	Init	ZO10	ZO20	ZO40	ZO60
$\mathbf{I}_{\text{Our}} \Phi^E$	3.77	3.45	3.45	3.45	3.67	8.76	7.78	7.50	7.37	7.42
$\mathbf{I}_{\text{Our}} \Phi^\Delta$	3.77	3.81	4.11	4.55	4.82	8.76	8.44	8.66	8.86	8.99
$\mathbf{I}_{\text{GFMDN}} \Phi^E$	2.85	2.88	2.93	3.00	3.22	10.17	9.44	9.35	9.43	9.38
$\mathbf{I}_{\text{GFMDN}} \Phi^\Delta$	2.85	2.92	3.28	3.70	3.95	10.17	9.72	9.89	10.27	10.41

Table 9: Matching results using ZoomOut with step size 1.

	S+F					F+S				
	Init	ZO10	ZO20	ZO40	ZO60	Init	ZO10	ZO20	ZO40	ZO60
$\mathbf{I}_{\text{Our}} \Phi^E$	3.77	3.34	3.36	3.52	3.54	8.76	7.54	7.35	7.27	7.28
$\mathbf{I}_{\text{Our}} \Phi^\Delta$	3.77	3.89	4.21	4.49	4.49	8.76	8.37	8.61	8.75	8.94
$\mathbf{I}_{\text{GFMDN}} \Phi^E$	2.85	2.76	2.83	3.03	3.03	10.17	9.28	9.18	9.22	9.34
$\mathbf{I}_{\text{GFMDN}} \Phi^\Delta$	2.85	2.97	3.25	3.51	3.56	10.17	9.65	9.99	10.19	10.48

Table 10: Matching results using ZoomOut with step size 2.