Deep Functional Maps:
Structured Prediction for Dense Shape Correspondence

Or Litany[1,2] Tal Remez[1]
Emanuele Rodolà[3,4] Alex Bronstein[2,5] Michael Bronstein[2,3]
[1]Tel Aviv University [2]Intel [3]USI Lugano [4]Sapienza University of Rome [5]Technion

Abstract

We introduce a new framework for learning dense correspondence between deformable 3D shapes. Existing learning based approaches model shape correspondence as a labelling problem, where each point of a query shape receives a label identifying a point on some reference domain; the correspondence is then constructed a posteriori by composing the label predictions of two input shapes. We propose a paradigm shift and design a structured prediction model in the space of functional maps, linear operators that provide a compact representation of the correspondence. We model the learning process via a deep residual network which takes dense descriptor fields defined on two shapes as input, and outputs a soft map between the two given objects. The resulting correspondence is shown to be accurate on several challenging benchmarks comprising multiple categories, synthetic models, real scans with acquisition artifacts, topological noise, and partiality.

1. Introduction

3D acquisition technology has made great progress in the last decade, and is being rapidly incorporated into commercial products ranging from Microsoft Kinect [43] for gaming, to LIDARs used in autonomous cars. An essential building block for application design in many of these domains is to recover 3D shape correspondences in a fast and reliable way. While handling real-world scanning artifacts is a challenge by itself, additional complications arise from non-rigid motions of the objects of interest (typically humans or animals). Most non-rigid shape correspondence methods employ local descriptors that are designed to achieve robustness to noise and deformations; however, relying on such "handcrafted" descriptors can often lead to inaccurate solutions in practical settings. Partial remedy to this was brought by the recent line of works on learning shape correspondence [28, 36, 29, 8, 10, 9, 31]. A key drawback of these methods lies in their emphasis on learning a descriptor that would help in identifying corresponding points, or on learning a labeling with respect to some reference domain. On the one hand, by focusing on the descriptor, the learning process remains agnostic to the way the final correspondence is computed, and costly post-processing steps are often necessary in order to obtain accurate solutions from the learned descriptors. On the other hand, methods based on a label space are restricted to a fixed number of points and rely on the adoption of an intermediate reference model.

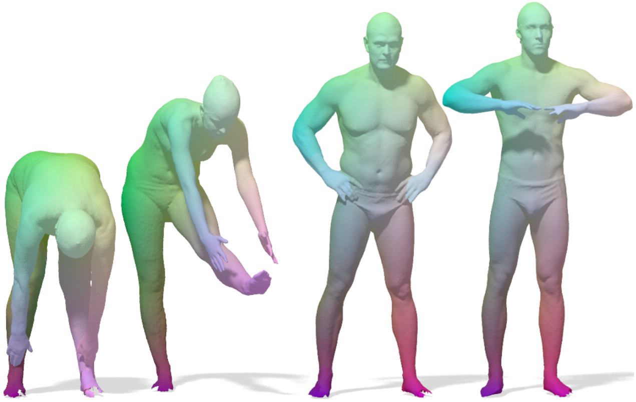

Figure 1. Correspondence results obtained by our network model on two pairs from the FAUST real scans challenge. Corresponding points are assigned the same color. The average error for the left and right pairs is 5.21cm and 2.34cm respectively. Accurate correspondence is obtained despite mesh "gluing" in areas of contact.

Contribution. In this work we propose a *task-driven* approach for descriptor learning, by including the computation of the correspondence directly as part of the learning procedure. Perfect candidates for this task are neural networks, due to their inherent flexibility to the addition of computational blocks. Our main contributions can be summarized as follows:

- We introduce a new *structured prediction* model for shape correspondence. Our framework allows end-to-end training: it takes base descriptors as input, and returns matches.

- We show that our approach consistently outperforms existing descriptor and correspondence learning methods on several recent benchmarks.

2. Related work

Shape correspondence is an active area of research in computer vision, graphics, and pattern recognition, with a variety of both classical and recent methods [11, 23, 13]. Since a detailed review of the literature would be out of scope for this paper, we refer the interested reader to the recent surveys on shape correspondence [39, 6]. More closely related to our approach is the family of methods based on the notion of *functional maps* [33], modeling correspondences as linear operators between spaces of functions on manifolds. In the truncated Laplacian eigenbases, such operators can be compactly encoded as small matrices – drastically reducing the amount of variables to optimize for, and leading to an increased flexibility in manipulating correspondences. This representation has been adopted and extended in several follow-up works [34, 22, 25, 3, 18, 35, 27, 26, 32], demonstrating state-of-the-art performance in multiple settings.

While being often adopted as a useful tool for the post-processing of some initial correspondence, functional maps have rarely been employed as a building block in correspondence learning pipelines.

Descriptor learning. Descriptor and feature learning are key topics in computer vision, and in recent years they have been actively investigated by the shape analysis community. Litman *et al*. [28] introduced *optimal spectral descriptors*, a data-driven parametrization of the spectral descriptor model (*i.e*., based on the eigen-decomposition of the Laplacian), demonstrating noticeable improvement upon the classical axiomatic constructions [37, 5]. A similar approach was subsequently taken in [42] and more recently in [8, 10, 17]. Perhaps most closely related to ours is the approach of Corman *et al*. [15], where combination weights for an input set of descriptors are learned in a supervised manner. Similarly to our approach, they base their construction upon the functional map framework [33]. While their optimality criterion is defined in terms of deviation from a ground-truth functional map in the *spectral* domain, we aim at recovering an optimal map in the *spatial* domain. Our structured prediction model will be designed with this requirement in mind.

Correspondence learning. Probably the first example of learning correspondence for deformable 3D shapes is the "shallow" random forest approach of Rodolà *et al*. [36]. More recently, Wei *et al*. [41] employed a classical (extrinsic) CNN architecture trained on huge training sets for learning invariance to pose changes and clothing. Convolutional neural networks on non-Euclidean domains (surfaces) were first considered by Masci *et al*. [29] with the introduction of the geodesic CNN model, a deep learning architecture where the classical convolution operation is replaced by an intrinsic (albeit, non-shift invariant) counterpart. The framework was shown to produce promising results in descriptor learning and shape matching applications, and was recently improved by Boscaini *et al*. [9] and generalized further by Monti *et al*. [31]. These methods are instances of a broader recent trend of *geometric deep learning* attempting to generalize successful deep learning paradigms to data with non-Euclidean underlying structure such as manifolds or graphs [12].

3. Background

Manifolds. We model shapes as two-dimensional Riemannian manifolds $\mathcal{X}$ (possibly with boundary $\partial\mathcal{X}$) equipped with the standard measure $\mathrm{d}\mu$ induced by the volume form. Throughout the paper we will consider the space of functions $L^2(\mathcal{X}) = \{f : \mathcal{X} \to \mathbb{R} \mid \langle f, f\rangle_{\mathcal{X}} < \infty\}$, with the standard manifold inner product $\langle f, g\rangle_{\mathcal{X}} = \int_{\mathcal{X}} f \cdot g \,\mathrm{d}\mu$.

The positive semi-definite Laplace-Beltrami operator $\Delta_{\mathcal{X}}$ generalizes the notion of Laplacian from Euclidean spaces to surfaces. It admits an eigen-decomposition $\Delta_{\mathcal{X}}\phi_i = \lambda_i\phi_i$ (with proper boundary conditions if $\partial\mathcal{X} \neq \emptyset$), where the eigenvalues form a discrete spectrum $0 = \lambda_1 \leq \lambda_2 \leq \dots$ and the eigenfunctions $\phi_1, \phi_2, \dots$ form an orthonormal basis for $L^2(\mathcal{X})$, allowing us to expand any function $f \in L^2(\mathcal{X})$ as a Fourier series

$$f(x) = \sum_{i\geq 1} \langle \phi_i, f\rangle_{\mathcal{X}} \phi_i(x) \,. \tag{1}$$

Aflalo *et al*. [2] have recently shown that Laplacian eigenbases are optimal for representing smooth functions on manifolds.

Functional correspondence. In order to compactly encode correspondences between shapes, we make use of the functional map representation introduced by Ovsjanikov *et al*. [33]. The key idea is to identify correspondences by a linear operator $T : L^2(\mathcal{X}) \to L^2(\mathcal{Y})$, mapping functions on $\mathcal{X}$ to functions on $\mathcal{Y}$. This can be seen as a generalization of classical point-to-point matching, which is a special case where delta functions are mapped to delta functions.

The linear operator T admits a matrix representation $\mathbf{C} = (c_{ij})$ with coefficients $c_{ji} = \langle \psi_j, T\phi_i\rangle_{\mathcal{Y}}$, where $\{\phi_i\}_{i\geq 1}$ and $\{\psi_j\}_{j\geq 1}$ are orthogonal bases on $L^2(\mathcal{X})$ and $L^2(\mathcal{Y})$ respectively, leading to the expansion:

$$Tf = \sum_{ij\geq 1} \langle \phi_i, f\rangle_{\mathcal{X}} c_{ji} \psi_j \,. \tag{2}$$

A good choice for the bases $\{\phi_i\}$, $\{\psi_j\}$ is given by the Laplacian eigenfunctions on the two shapes [33, 2]. This choice is particularly convenient, since (by analogy with Fourier analysis) it allows to truncate the series (2) after the first k coefficients – yielding a band-limited approximation of the original map. The resulting matrix $\mathbf{C}$ is a $k \times k$ compact representation of a correspondence between the two shapes, where typically $k \ll n$ (here n is the number of points on each shape).

Functional correspondence problems seek a solution for the matrix $\mathbf{C}$, given a set of corresponding functions $f_i \in L^2(\mathcal{X})$ and $g_i \in L^2(\mathcal{Y})$, $i = 1, \dots, q$, on the two shapes. In the Fourier basis, these functions are encoded into matrices $\hat{\mathbf{F}} = (\langle \phi_i, f_j \rangle_{\mathcal{X}})$ and $\hat{\mathbf{G}} = (\langle \psi_i, g_j \rangle_{\mathcal{Y}})$, leading to the least-squares problem:

$$\min_{\mathbf{C}} \| \mathbf{C}\hat{\mathbf{F}} - \hat{\mathbf{G}} \|_F^2 \,. \tag{3}$$

In practice, dense q-dimensional descriptor fields (*e.g.*, HKS [37] or SHOT [38]) on $\mathcal{X}$ and $\mathcal{Y}$ are used as the corresponding functions.

Label space. Previous approaches at learning shape correspondence phrased the matching problem as a *labelling* problem [36, 29, 10, 9, 31]. These approaches attempt to label each vertex of a given query shape $\mathcal{X}$ with the index of a corresponding point on some reference shape $\mathcal{Z}$ (usually taken from the training set), giving rise to a dense point-wise map $T_{\mathcal{X}} : \mathcal{X} \to \mathcal{Z}$. The correspondence between two queries $\mathcal{X}$ and $\mathcal{Y}$ can then be obtained via the composition $T_{\mathcal{Y}}^{-1} \circ T_{\mathcal{X}}$ [36].

Given a training set $S = \{(x, \pi^*(x))\} \subset \mathcal{X} \times \mathcal{Y}$ of matches under the ground-truth map $\pi^* : \mathcal{X} \to \mathcal{Y}$, label-based approaches compute a descriptor $F_\Theta(x)$ whose optimal parameters are found by minimizing the *multinomial regression* loss:

$$\ell_{\mathrm{mr}}(\Theta) = - \sum_{(x,\pi^*(x)) \in S} \langle \delta_{\pi^*(x)}, \log F_\Theta(x) \rangle_{\mathcal{Y}} \,, \tag{4}$$

where $\delta_{\pi^*(x)}$ is a delta function on $\mathcal{Y}$ at point $\pi^*(x)$.

Such an approach essentially treats the correspondence problem as one of classification, where the aim is to approximate as closely as possible (in a statistical sense) the correct label for each point. The actual construction of the full correspondence is done *a posteriori* by a composition step with an intermediate reference domain, or by solving the least-squares problem (3) with the learned descriptors as data.

Discretization. In the discrete setting, shapes are represented as manifold triangular meshes with n vertices (in general, different for each shape). The Laplace-Beltrami operator Δ is discretized as a symmetric $n \times n$ matrix $\mathbf{L} = \mathbf{A}^{-1}\mathbf{W}$ using a classical linear FEM scheme [30], where the *stiffness matrix* $\mathbf{W}$ contains the cotangent weights, and the *mass matrix* $\mathbf{A}$ is a diagonal matrix of vertex area elements. The manifold inner product $\langle f, g \rangle$ is discretized as the area-weighted dot product $\mathbf{f}^\top \mathbf{A}\mathbf{g}$, where the vectors $\mathbf{f}, \mathbf{g} \in \mathbb{R}^n$ contain the function values of f and g at each vertex. Note that under such discretization we have $\mathbf{\Phi}^\top \mathbf{A} \mathbf{\Phi} = \mathbf{I}$, where $\mathbf{\Phi}$ contains the Laplacian eigenfunctions as its columns.

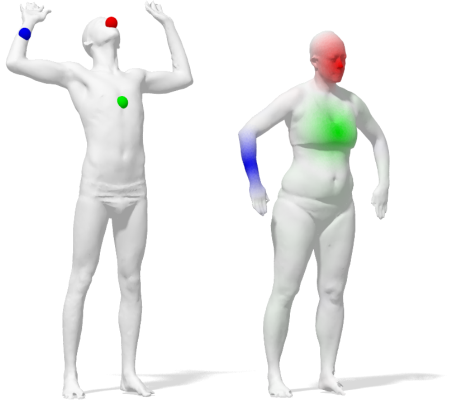

Figure 2. Given a source and a target shape as input, our network outputs a soft correspondence matrix whose columns can be interpreted as probability distributions over the target shape.

4. Deep Functional Maps

In this paper we propose an alternative model to the labelling approach described above. We aim at learning point-wise descriptors which, when used in a functional map pipeline such as (3), will induce an accurate correspondence. To this end, we construct a neural network which takes as input existing, manually designed descriptors and improves upon those while satisfying a *geometrically* meaningful criterion. Specifically, we consider the *soft error loss*

$$\ell_{\mathrm{F}} = \sum_{(x,y) \in (\mathcal{X}, \mathcal{Y})} P(x,y) d_{\mathcal{Y}}(y, \pi^*(x)) = \| \mathbf{P} \circ \mathbf{D}_{\mathcal{Y}} \|_{\mathrm{F}} \,, \tag{5}$$

where $\mathbf{D}_{\mathcal{Y}}$ is the $n \times n$ matrix of geodesic distances on $\mathcal{Y}$, $\circ$ is the element-wise product, and

$$\mathbf{P} = |\mathbf{\Psi} \mathbf{C} \mathbf{\Phi}^\top \mathbf{A}|^\wedge \tag{6}$$

is a *soft correspondence* matrix, which can be interpreted as the probability of point $x \in \mathcal{X}$ mapping to point $y \in \mathcal{Y}$ (see Figure 2); here, $\mathbf{\Phi}, \mathbf{\Psi}$ are matrices containing the first k eigenfunctions $\{\phi_i\}$, $\{\psi_j\}$ as their columns, $|\cdot|$ acts element-wise, and $\mathbf{X}^\wedge$ is a column-wise normalization of $\mathbf{X}$. In the formula above, the $k \times k$ matrix $\mathbf{C}$ represents a functional map obtained as the least-squares solution to (3) under *learned* descriptors $\mathbf{F}, \mathbf{G}$.

Matrix $\mathbf{P}$ represents a rank-k approximation of the spatial correspondence between the two shapes, thus allowing

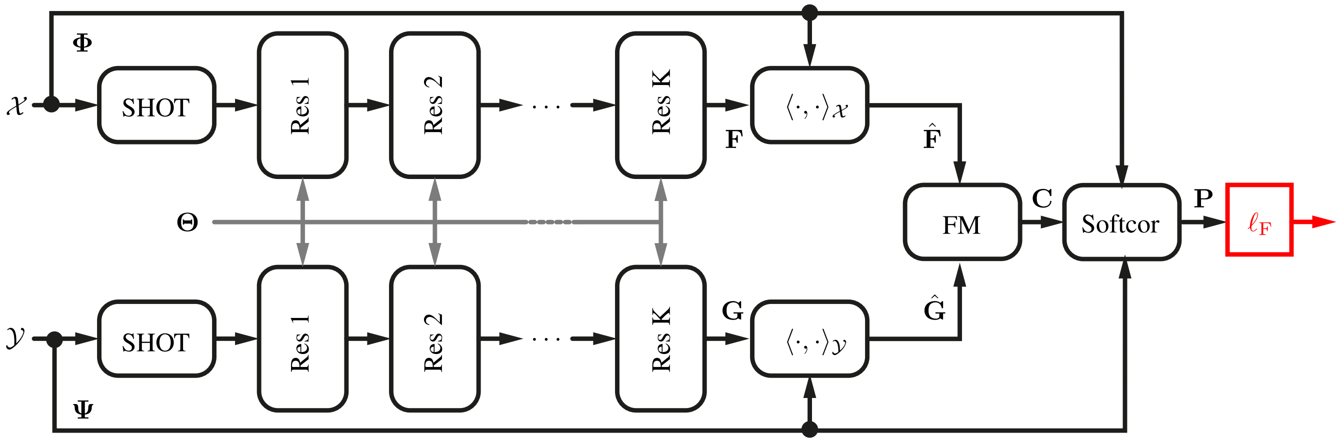

Figure 3. **FMNet architecture.** Input point-wise descriptors (SHOT [38] in this paper) from a pair of shapes are passed through an identical sequence of operations (with shared weights), resulting in refined descriptors $\mathbf{F}, \mathbf{G}$. These, in turn, are projected onto the Laplacian eigenbases $\mathbf{\Phi}, \mathbf{\Psi}$ to produce the spectral representations $\hat{\mathbf{F}}, \hat{\mathbf{G}}$. The functional map (FM) and soft correspondence (Softcor) layers, implementing Equations (3) and (6) respectively, are not parametric and are used to set up the geometrically structured loss ℓ_{F} (5).

us to interpret the soft error (5) as a probability-weighted geodesic distance from the ground-truth. This measure, introduced in [25] as an evaluation criterion for soft maps, endows our solutions with guarantees of mapping nearby points on $\mathcal{X}$ to nearby points on $\mathcal{Y}$. On the contrary, the classification cost (4), adopted by existing label-based correspondence learning approaches, considers *equally* correspondences that deviate from the ground-truth, no matter how far. Further, notice that Equation (6) is asymmetric, implying that each pair of training shapes can be used twice for training (i.e., in both directions). Also note that, differently from previous approaches operating in the label space, in our setting the number of effective training examples (i.e. pairs of shapes) increases *quadratically* with the number of shapes in the collection. This is a significant advantage in situations with scarce training data.

We implement descriptor learning using a Siamese residual network architecture [21]. To this network, we concatenate additional *non*-parametric layers implementing the least-squares solve (3) followed by computation of the soft correspondence according to (6). In particular, the solution to (3) is obtained in closed form as $\mathbf{C} = \hat{\mathbf{G}}\hat{\mathbf{F}}^{\dagger}$, where † denotes the pseudo-inverse operation. The complete architecture (named "FMNet") is illustrated in Figure 3.

5. Implementation details

Data. For increased efficiency, we down-sample the input shapes to 15K vertices by edge contraction [19]; in case the input mesh has a smaller amount of vertices, it is kept at full resolution. As input feature for the network we use the 352-dimensional SHOT descriptor [38], computed on all vertices of the remeshed shapes. The choice of the descriptor is mainly driven by its fast computation and its local nature, making it a robust candidate in the presence of missing parts. Note that while this descriptor is *not*, strictly speaking, deformation invariant, it was shown to work well for the deformable setting in practice [35, 27]. Recent work on learning-based shape correspondence makes use of the same input feature [9, 31].

Network. Our network architecture consists of 7 fully-connected residual layers as described in [21] with exponential linear units (ELUs) [14] and no dimensionality reduction, implemented in TensorFlow[1] [1]. Depending on the dataset, we used 20K (FAUST synthetic), 100K (FAUST real scans), and 1K (SHREC'16) training mini-batches, each containing ∼1K randomly chosen ground-truth matches. For the FAUST real dataset, sampling was weighted according to the vertex area elements to prevent unduly aggregation of matches to high-resolution portions of the surface. For the three datasets, we used respectively $k = 120, 70, 100$ eigenfunctions for representing the $k \times k$ matrix $\mathbf{C}$ inside the network. Training was done using the ADAM optimizer [24] with a learning rate of $\alpha = 10^{-3}$, $\beta_1 = 0.9$, $\beta_2 = 0.999$ and $\epsilon = 10^{-8}$. The average prediction runtime for a pair of FAUST models is 0.25 seconds.

Upscaling. Given two down-sampled shapes $\tilde{\mathcal{X}}$ and $\tilde{\mathcal{Y}}$, the network predicts a $k \times k$ matrix $\tilde{\mathbf{C}}$ encoding the correspondence between the two. Since this matrix is expressed w.r.t. basis functions $\{\tilde{\phi}_i\}_i, \{\tilde{\psi}_j\}_j$ of the *low-resolution* shapes, it can not be directly used to recover a point-wise map between the full-resolution counterparts $\mathcal{X}$ and $\mathcal{Y}$. Therefore, we perform an upscaling step as follows.

Let $\pi_{\mathcal{X}} : \tilde{\mathcal{X}} \to \mathcal{X}$ be the injection mapping each point in $\tilde{\mathcal{X}}$ to the corresponding point in the full shape $\mathcal{X}$ (this

[1] Code and data are available at https://github.com/orlitany/DeepFunctionalMaps.

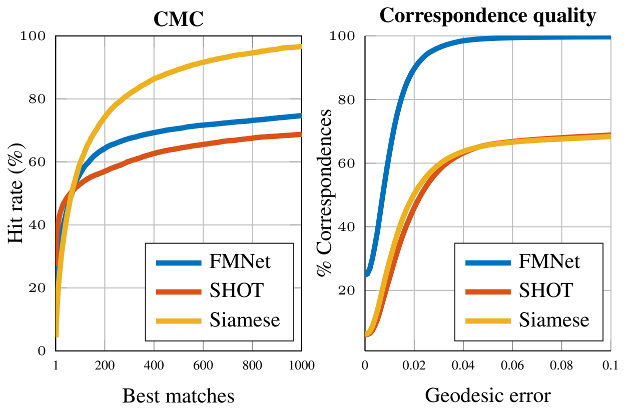

Figure 4. Comparison between our structured prediction model (FMNet), metric learning (Siamese), and baseline SHOT in terms of CMC (left) and geodesic error (right). While the Siamese model produces better descriptors in terms of proximity (left), these do not necessarily induce a good functional correspondence (right).

map can be easily recovered by a simple nearest-neighbor search in $\mathbb{R}^3$), and similarly for shape $\mathcal{Y}$. Further, denote by $\tilde{T} : \tilde{\mathcal{X}} \to \tilde{\mathcal{Y}}$ the point-to-point map recovered from $\tilde{\mathbf{C}}$ using the baseline recovery approach of [33]. A map $T : \mathcal{X} \supset \mathrm{Im}(\pi_{\mathcal{X}}) \to \mathcal{Y}$ is obtained via the composition $T = \pi_{\mathcal{Y}} \circ \tilde{T} \circ \pi_{\mathcal{X}}^{-1}$. However, while $\tilde{T}$ is dense in $\tilde{\mathcal{X}}$, the map T is *sparse* in $\mathcal{X}$. In order to map *each* point in $\mathcal{X}$ to a point in $\mathcal{Y}$, we construct pairs of delta functions $\delta_{x_i} : \mathcal{X} \to \{0,1\}$ and $\delta_{T(x_i)} : \mathcal{Y} \to \{0,1\}$ supported at corresponding points $(x_i, T(x_i))$ for $i = 1,\dots,|\tilde{\mathcal{X}}|$; note that we have as many corresponding pairs as the number of vertices in the low-resolution shape $\tilde{\mathcal{X}}$. We use these corresponding functions to define the minimization problem:

$$\mathbf{C}^* = \arg\min_{\mathbf{C}} \|\mathbf{C}\hat{\mathbf{F}} - \hat{\mathbf{G}}\|_{2,1}\,, \tag{7}$$

where $\hat{\mathbf{F}} = (\langle \phi_i, \delta_{x_j} \rangle_{\mathcal{X}})$ and $\hat{\mathbf{G}} = (\langle \psi_i, \delta_{T(x_j)} \rangle_{\mathcal{Y}})$ contain the Fourier coefficients (in the full-resolution basis) of the corresponding delta functions, and the $\ell_{2,1}$-norm allows to discard potential mismatches in the data[2]. Problem (7) is non-smooth and convex, and can be solved globally using ADMM-like techniques. A dense point-to-point map between $\mathcal{X}$ and $\mathcal{Y}$ is finally recovered from the optimal functional map $\mathbf{C}^*$ by the nearest-neighbor approach of [33].

6. Results

We performed a wide range of experiments on real and synthetic data to demonstrate the efficacy of our method. Qualitative and quantitative comparisons were carried out with respect to the state of the art on multiple recent benchmarks, encapsulating different matching scenarios.

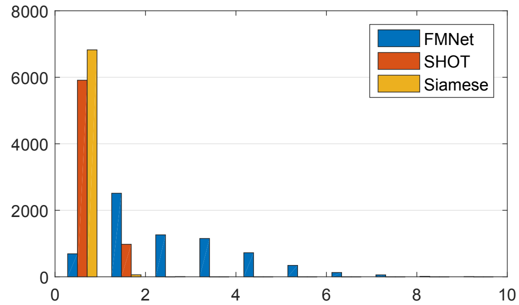

Figure 5. Distance distributions (in descriptor space) between correct matches. Since FMNet does not optimize a distribution criterion of this kind, it exhibits a heavy tail despite producing excellent correspondences.

Error measure. We measure correspondence quality according to the Princeton benchmark protocol [23]. Assume to be given a match $(x,y) \in \mathcal{X} \times \mathcal{Y}$, whereas the ground-truth correspondence is (x,y^*). Then, we measure the *geodesic error*:

$$\epsilon(x) = \frac{d_{\mathcal{Y}}(y,y^*)}{\mathrm{area}(\mathcal{Y})^{1/2}}\,, \tag{8}$$

having units of normalized geodesic length on $\mathcal{Y}$ (ideally, zero). We plot cumulative curves showing the percent of matches that have error smaller than a variable threshold.

Metric learning. As a proof of concept, we start by studying the behavior of our framework when the functional map layer is removed, and the soft error criterion (6) is replaced with the *siamese* loss [20]:

$$\begin{aligned}\ell_{\mathrm{s}}(\Theta) = &\sum_{x,x^+ \in S} \gamma \|F_\Theta(x) - F_\Theta(x^+)\|_2^2 \\ &+ \sum_{x,x^- \in D} (1-\gamma)(\mu - \|F_\Theta(x) - F_\Theta(x^-)\|_2)_+^2\,,\end{aligned} \tag{9}$$

where $\gamma \in (0,1)$ is a trade-off parameter, $\mu > 0$ is the margin, and $(x)_+ = \max(0,x)$. Here, the sets $S, D \subset \mathcal{X} \times \mathcal{Y}$ constitute the training data consisting of knowingly similar and dissimilar pairs of points respectively. By considering this loss function, we transform our *structured prediction* model into a *metric learning* model. The learned descriptors $F_\Theta(x)$ can be subsequently plugged into (3) to compute a correspondence; this metric learning approach was recently used in a functional map pipeline in [17]. For this test we use FAUST templates [7] as our data and SHOT [38] as an input feature.

From the CMC curves[3] of Figure 4 (left) and the distance distributions of Figure 5 we can clearly see that the model (9) succeeds at producing descriptors that attract each other

[2] The matrix norm $\|\mathbf{X}\|_{2,1}$ is defined as the sum of the ℓ_2 norms of the columns of $\mathbf{X}$.

[3] *Cumulative error characteristic* (CMC) curves evaluate the probability (y-axis) of finding the correct match within the first k best matches (x-axis), obtained as ℓ_2-nearest neighbors in descriptor space.

	inter AE	inter WE	intra AE	intra WE
Zuffi et al. [44]	3.13	6.68	1.57	5.58
Chen et al. [13]	8.30	26.80	4.86	26.57
FMNet	4.83	9.56	2.44	26.16

Table 1. Comparison with the state of the art in terms of average error (AE) and worst error (WE) on the FAUST challenge with real scans. The error measure is reported in cm.

at corresponding points, while mismatches are repulsed. However, as put in evidence by Figure 4 (right), these descriptors do not perform well when they are used for seeking a dense correspondence via (3). Contrarily, our structured prediction model yields descriptors that are optimized for such a correspondence task, leading to a noticeable gain in accuracy.

Real scans. We carried out experiments on real 3D acquisitions using the recent FAUST benchmark [7]. The dataset consists of real scans (∼200K vertices per shape) of different people in a variety of poses, acquired with a full-body 3D stereo capture system. The benchmark is divided into two parts, namely the 'intra-class' (60 pairs of shapes, with each pair depicting different poses of the same subject) and the 'inter-class' challenge (40 pairs of different subjects in different poses). The benchmark does not provide ground-truth correspondence for the challenge pairs, whereas the accuracy evaluation is provided by an online service. Hence, for these experiments we only compare with methods that made their results publicly available via the official ranking: the recent convex optimization approach of Chen and Koltun [13], and the parametric method of Zuffi and Black [44].

As training data for our approach we use the official training set of 100 shapes provided with FAUST. Since ground truth correspondences are only given between low-resolution templates registered to the scans (and not between the scans themselves), during training we augmented our data by sampling, for each template vertex, one out of 10 nearest neighbors to the real scan; this step makes the network more robust to noise in the form of small vertex displacement.

The comparison results are reported in Table 1. Reading these results, we see that our approach considerably improves upon [13] (around 50%); note that while the latter method is not learning-based, it relies on a pose prior where the shapes are put into initial alignment in 3D in order to drive the optimization to a good solution. The approach of [44] obtains slightly better results than our method, but lacks in generality: it is based on a human-specific parametric model (called the 'stitched puppet'), trained on a collection of 6000 meshes in different poses from motion capture data. Our model is trained on almost two orders of magnitude less data, and can be applied to any shape category (*e.g.*, animals) as we demonstrate later in this Section.

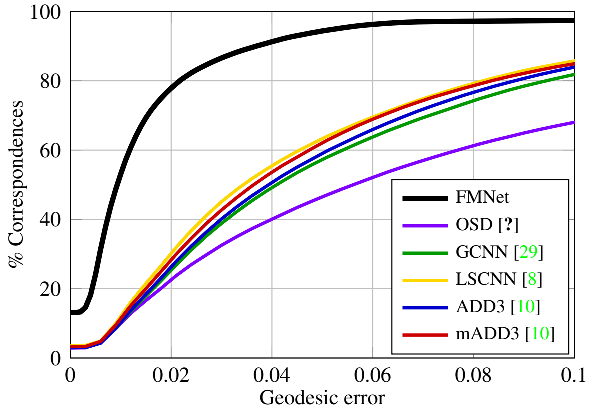

Figure 6. Comparison with learning-based shape matching approaches on the SCAPE dataset. Our method (FMNet) was trained on FAUST data, demonstrating excellent generalization, while all other methods were trained on SCAPE.

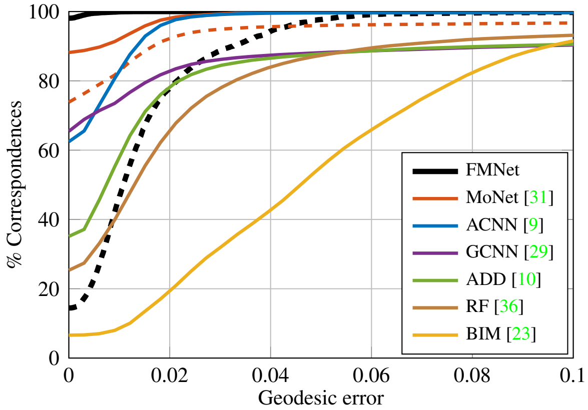

Figure 7. Comparison with learning-based approaches on the FAUST humans dataset. Dashed and solid curves denote performance before and after refinement respectively. FMNet has 98% correspondences with zero error (top left corner of the plot).

Transfer. We demonstrate the generalization capabilities of FMNet by performing a series of experiments on the SCAPE dataset of human shapes [4], where our network is trained on FAUST scans data as described previously. We compare with state-of-the-art learning-based approaches for deformable shape correspondence, namely optimal spectral descriptors (OSD) [28], geodesic CNNs (GCNN) [29], localized spectral CNNs (LSCNN) [8], and two variants of anisotropic diffusion descriptors (ADD3, mADD3) [10]. With the exception of FMNet, which was trained only on FAUST data, all the above methods were trained on 60 shapes from the SCAPE dataset. The remaining 10 shapes are used for testing. The results are reported in Figure 6; note that for a fair comparison, we show the *raw* predicted correspondence (*i.e.*, without post-processing) for all methods.

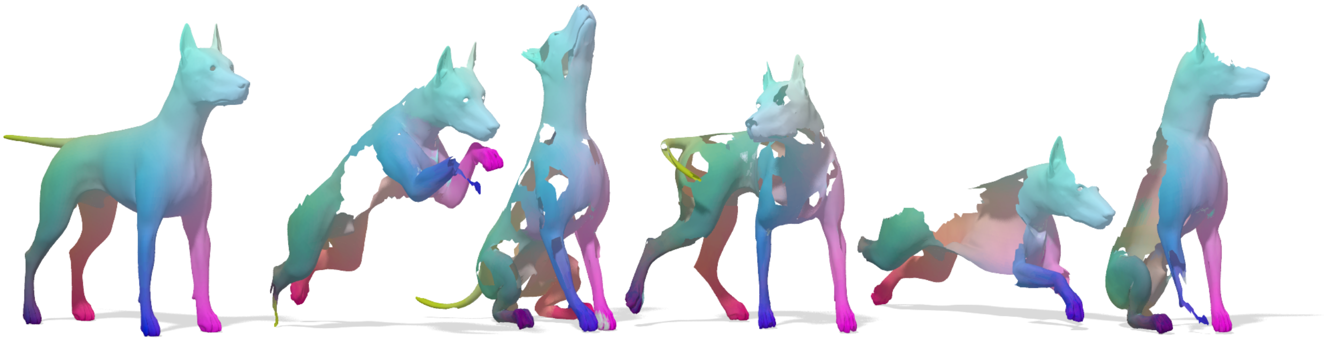

Figure 8. Results of FMNet on the SHREC'16 Partial Correspondence benchmark. Each partial shape is matched to the full shape on the left; the color texture is transferred via the predicted correspondence.

Synthetic shapes. For these experiments we reproduce verbatim the experimental setup of [10, 9, 31]: the training set consists of the first 80 shapes of the FAUST dataset; the remaining 20 shapes are used for testing. Differently from the comparisons of Table 1, the shapes are now taken from the *synthetic* dataset provided with FAUST (∼7K vertices per shape), for which exact ground-truth correspondence is available. We compare with the most recent state of the art in learning-based shape correspondence: random forests (RF) [36], anisotropic diffusion descriptors (ADD) [10], geodesic CNNs (GCNN) [29], anisotropic CNNs (ACNN) [9], and the very recent MoNet model [31]. As a representative method for the family of axiomatic techniques, we additionally include blended intrinsic maps (BIM) [23] in the comparison.

The results are reported in Figure 7; here, all methods were post-processed with the correspondence refinement technique of [40]. For FMNet and MoNet (the top-performing competitor) we also report curves before refinement. We note that the raw prediction of MoNet has a higher accuracy at zero. This is to be expected due to the classifier nature of this method (and all the methods in this comparison): the logistic loss (4) aims at high pointwise accuracy, but has limited global awareness of the correspondence. Indeed, the task-driven nature of our approach induces lower accuracy at zero, but better global behavior – note how the curve "saturates" at around 0.06 while MoNet never does. As a result, FMNet significantly outperforms MoNet after refinement, producing almost ideal correspondence with zero error.

Partial non-human shapes. Our framework does not rely on any specific shape model, as it learns from the shape categories represented in the training data. In particular, it does not necessarily require the objects to be complete shapes: different forms of partiality can be tackled if adequately represented in the training set.

We demonstrate this by running our method on the recent SHREC'16 Partial Correspondence challenge [16]. The benchmark consists of hundreds of shapes of multiple categories with missing parts of various forms and sizes; a training set is also provided. We selected the 'dog' class from the 'holes' sub-challenge, being this among the hardest categories in the benchmark. The dataset is officially split into just 10 training shapes, and 26 test shapes. Qualitative examples of the obtained solutions are reported in Figure 8.

7. Discussion and conclusions

We introduced a new neural network based method for dense shape correspondence, structured according to the functional maps framework. Building upon the recent success of end-to-end learning approaches, our network directly estimates correspondences. This is in contrast to previous descriptor learning techniques, that do not account for post processing while training. We showed this methodology to be beneficial via an evaluation on a several challenging benchmarks, comprising synthetic models, real scans with acquisition artifacts, and partiality. Being model-free, we demonstrated our method can be adapted to different shape categories such as dogs. Furthermore, we showed our method is capable of generalizing between different datasets.

Limitations. Laplacian eigenfunctions are inherently sensitive to topological changes. Indeed, such examples proved to be more challenging for our method. A different choice of a basis may be useful in mitigating this issue.

As shown by recent works addressing partial correspondence using functional maps [35], special care should be taken when recovering matrix $\mathbf{C}$ from the spectral representation of the descriptors. While our method was able to recover most pairs with missing parts, it failed to recover correspondences under extreme partiality (see inset). This could be addressed by incorporating partiality priors into our structured prediction model.

Acknowledgments

The authors wish to thank Jonathan Masci, Federico Monti, Dan Raviv and Maks Ovsjanikov for useful discussions. OL, TR and AB are supported by the ERC grant no. 335491. ER and MB are supported by the ERC grants no. 307047 and 724228, a Google Research Faculty Award, a Radcliffe Fellowship, a Nvidia equipment grant, and a TUM-IAS Rudolf Diesel Industry Fellowship.

References

[1] M. Abadi, A. Agarwal, P. Barham, E. Brevdo, Z. Chen, C. Citro, G. S. Corrado, A. Davis, J. Dean, M. Devin, et al. Tensorflow: Large-scale machine learning on heterogeneous systems, 2015. *Software available from tensorflow. org*, 1, 2015. 4

[2] Y. Aflalo, H. Brezis, and R. Kimmel. On the optimality of shape and data representation in the spectral domain. *SIAM J. Imaging Sciences*, 8(2):1141–1160, 2015. 2, 3

[3] Y. Aflalo, A. Dubrovina, and R. Kimmel. Spectral generalized multi-dimensional scaling. *IJCV*, 118(3):380–392, 2016. 2

[4] D. Anguelov, P. Srinivasan, D. Koller, S. Thrun, J. Rodgers, and J. Davis. Scape: Shape completion and animation of people. *Trans. Graphics*, 24(3):408–416, 2005. 6

[5] M. Aubry, U. Schlickewei, and D. Cremers. The wave kernel signature: A quantum mechanical approach to shape analysis. In *Proc. ICCV*, 2011. 2

[6] S. Biasotti, A. Cerri, A. Bronstein, and M. Bronstein. Recent trends, applications, and perspectives in 3d shape similarity assessment. In *Computer Graphics Forum*, 2015. 2

[7] F. Bogo, J. Romero, M. Loper, and M. J. Black. FAUST: Dataset and evaluation for 3D mesh registration. In *Proc. CVPR*, June 2014. 5, 6

[8] D. Boscaini, J. Masci, S. Melzi, M. M. Bronstein, U. Castellani, and P. Vandergheynst. Learning class-specific descriptors for deformable shapes using localized spectral convolutional networks. *Computer Graphics Forum*, 34(5):13–23, 2015. 1, 2, 6

[9] D. Boscaini, J. Masci, E. Rodolà, and M. M. Bronstein. Learning shape correspondence with anisotropic convolutional neural networks. In *Proc. NIPS*, 2016. 1, 2, 3, 4, 6, 7

[10] D. Boscaini, J. Masci, E. Rodolà, M. M. Bronstein, and D. Cremers. Anisotropic diffusion descriptors. *Computer Graphics Forum*, 35(2):431–441, 2016. 1, 2, 3, 6, 7

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

[12] M. Bronstein, J. Bruna, Y. LeCun, A. Szlam, and P. Vandergheynst. Geometric deep learning: going beyond Euclidean data. *arXiv:1611.08097*, 2016. 2

[13] Q. Chen and V. Koltun. Robust nonrigid registration by convex optimization. In *Proc. ICCV*, 2015. 2, 6

[14] D.-A. Clevert, T. Unterthiner, and S. Hochreiter. Fast and accurate deep network learning by exponential linear units (elus). *arXiv:1511.07289*, 2015. 4

[15] É. Corman, M. Ovsjanikov, and A. Chambolle. Supervised descriptor learning for non-rigid shape matching. In *Proc. ECCV*, 2014. 2

[16] L. Cosmo, , et al. SHREC'16: Partial matching of deformable shapes. In *Proc. 3DOR*, 2016. 7

[17] L. Cosmo, E. Rodolà, J. Masci, A. Torsello, and M. Bronstein. Matching deformable objects in clutter. In *Proc. 3DV*, 2016. 2, 5

[18] D. Eynard, E. Rodola, K. Glashoff, and M. M. Bronstein. Coupled functional maps. In *Proc. 3DV*, 2016. 2

[19] M. Garland and P. S. Heckbert. Surface simplification using quadric error metrics. In *Proc. SIGGRAPH*, pages 209–216, 1997. 4

[20] R. Hadsell, S. Chopra, and Y. LeCun. Dimensionality reduction by learning an invariant mapping. In *Proc. CVPR*, 2006. 5

[21] K. He, X. Zhang, S. Ren, and J. Sun. Deep residual learning for image recognition. In *Proc. CVPR*, 2016. 4

[22] Q. Huang, F. Wang, and L. J. Guibas. Functional map networks for analyzing and exploring large shape collections. *Trans. Graphics*, 33(4):36, 2014. 2

[23] V. G. Kim, Y. Lipman, and T. A. Funkhouser. Blended intrinsic maps. *Trans. Graphics*, 30(4):79, 2011. 2, 5, 6, 7

[24] D. P. Kingma and J. Ba. ADAM: A method for stochastic optimization. In *ICLR*, 2015. 4

[25] A. Kovnatsky, M. M. Bronstein, X. Bresson, and P. Vandergheynst. Functional correspondence by matrix completion. In *Proc. CVPR*, 2015. 2, 4

[26] O. Litany, E. Rodolà, A. M. Bronstein, and M. M. Bronstein. Fully spectral partial shape matching. *Computer Graphics Forum*, 36(2), 2017. 2

[27] O. Litany, E. Rodolà, A. M. Bronstein, M. M. Bronstein, and D. Cremers. Non-rigid puzzles. *Computer Graphics Forum*, 35(5):135–143, 2016. 2, 4

[28] R. Litman and A. M. Bronstein. Learning spectral descriptors for deformable shape correspondence. *Trans. PAMI*, 36(1):170–180, 2014. 1, 2, 6

[29] J. Masci, D. Boscaini, M. M. Bronstein, and P. Vandergheynst. Geodesic convolutional neural networks on Riemannian manifolds. In *Proc. 3dRR*, 2015. 1, 2, 3, 6, 7

[30] M. Meyer, M. Desbrun, P. Schröder, and A. H. Barr. Discrete differential-geometry operators for triangulated 2-manifolds. *Visualization&Mathematics*, pages 35–57, 2003. 3

[31] F. Monti, D. Boscaini, J. Masci, E. Rodolà, J. Svoboda, and M. M. Bronstein. Geometric deep learning on graphs and manifolds using mixture model CNNs. In *Proc. CVPR*, 2017. 1, 2, 3, 4, 6, 7

[32] D. Nogneng and M. Ovsjanikov. Informative descriptor preservation via commutativity for shape matching,. *Computer Graphics Forum*, 36(2), 2017. 2

[33] M. Ovsjanikov, M. Ben-Chen, J. Solomon, A. Butscher, and L. Guibas. Functional maps: a flexible representation of maps between shapes. *Trans. Graphics*, 31(4):30:1–30:11, July 2012. 2, 3, 5

[34] J. Pokrass, A. M. Bronstein, M. M. Bronstein, P. Sprechmann, and G. Sapiro. Sparse modeling of intrinsic correspondences. *Computer Graphics Forum*, 32(2):459–468, 2013. 2

[35] E. Rodolà, L. Cosmo, M. M. Bronstein, A. Torsello, and D. Cremers. Partial functional correspondence. *Computer Graphics Forum*, 36(1):222–236, 2017. 2, 4, 7

[36] E. Rodolà, S. Rota Bulò, T. Windheuser, M. Vestner, and D. Cremers. Dense non-rigid shape correspondence using random forests. In *Proc. CVPR*, 2014. 1, 2, 3, 6, 7

[37] J. Sun, M. Ovsjanikov, and L. Guibas. A concise and provably informative multi-scale signature based on heat diffusion. In *Proc. SGP*, 2009. 2, 3

[38] F. Tombari, S. Salti, and L. Di Stefano. Unique signatures of histograms for local surface description. In *Proc. ECCV*, 2010. 3, 4, 5

[39] O. van Kaick, H. Zhang, G. Hamarneh, and D. Cohen-Or. A survey on shape correspondence. *Computer Graphics Forum*, 30(6):1681–1707, 2011. 2

[40] M. Vestner, R. Litman, E. Rodolà, A. Bronstein, and D. Cremers. Product manifold filter: Non-rigid shape correspondence via kernel density estimation in the product space. In *Proc. CVPR*, 2017. 7

[41] L. Wei, Q. Huang, D. Ceylan, E. Vouga, and H. Li. Dense human body correspondences using convolutional networks. In *Proc. CVPR*, 2016. 2

[42] T. Windheuser, M. Vestner, E. Rodolà, R. Triebel, and D. Cremers. Optimal intrinsic descriptors for non-rigid shape analysis. In *Proc. BMVC*, 2014. 2

[43] Z. Zhang. Microsoft Kinect sensor and its effect. *IEEE Multimedia*, 19(2):4–10, 2012. 1

[44] S. Zuffi and M. J. Black. The stitched puppet: A graphical model of 3D human shape and pose. In *Proc. CVPR*, 2015. 6