Cyclic Functional Mapping:
Self-supervised correspondence between non-isometric deformable shapes

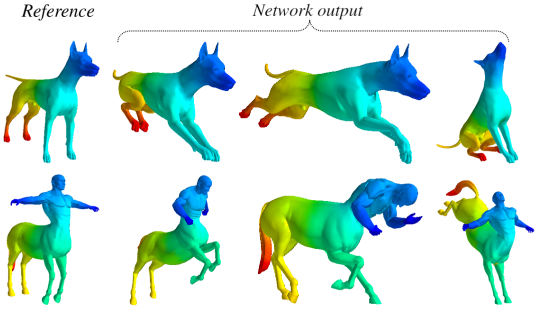
Dvir Ginzburg
dvirginzburg@mail.tau.ac.il

Dan Raviv
darav@tauex.tau.ac.il

Tel Aviv University

Abstract. We present the first spatial-spectral joint consistency network for self-supervised dense correspondence mapping between non-isometric shapes. The task of alignment in non-Euclidean domains is one of the most fundamental and crucial problems in computer vision. As 3D scanners can generate highly complex and dense models, the mission of finding dense mappings between those models is vital. The novelty of our solution is based on a cyclic mapping between metric spaces, where the distance between a pair of points should remain invariant after the full cycle. As the same learnable rules that generate the point-wise descriptors apply in both directions, the network learns invariant structures without any labels while coping with non-isometric deformations. We show here state-of-the-art-results by a large margin for a variety of tasks compared to known self-supervised and supervised methods.

Keywords: Dense shape correspondence, Self-supervision, One-shot learning, Spectral decomposition, 3D alignment



.

Fig. 1: TOSCA dataset results - similar colors represents correspondence mapping - we show excellent generalization after training for a single epoch on the TOSCA dataset with a pre-trained model on FAUST 6.5

1 Introduction

Alignment of non-rigid shapes is a fundamental problem in computer vision and plays an important role in multiple applications such as pose transfer, cross-shape texture mapping, 3D body scanning, and simultaneous localization and mapping (SLAM). The task of finding dense correspondence is especially challenging for non-rigid shapes, as the number of variables needed to define the mapping is vast, and local deformations might occur. To this end, we have seen a variety of papers focusing on defining unique key-points. These features capture the local uniqueness of the models using curvature [31], normals [48], or heat [45], for example, and further exploited for finding a dense mapping [12, 5].

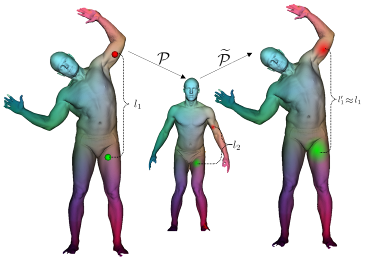


Fig. 2: Self-supervised dense correspondence using a cycle mapping architecture. By minimizing the geodesic distortion only on the source shape, we can learn complex deformations between structures.

A different approach used for alignment is based on pair-wise distortions, where angles [6, 24] or distances [15, 9, 35, 34], between pairs of points are minimized. Formulating this as a linear [52] or quadratic [9, 47, 2] optimization scheme showed a significant enhancement but with a painful time complexity even for small models. An interesting approach that addressed local deformations while still using Riemannian geometry introduced new metrics. Specifically, scale invariant [4, 37], equi-affine invariant [32] and affine-invariant [36] metric showed superior results for alignment tasks. To confront the challenges in alignment between stretchable shapes we recognize non-metric methods based on conformal mapping [24], experimenting with alternative metrics such as scale [4] or affine invariant metrics [33], or attempts to embed the shapes into a plane or a cone [11, 9], for example. A significant milestone named functional maps [30] has shown that such a mapping can be performed on the spectral domain, by aligning functions overlaid on top of the shapes.

Recently, a substantial improvement in dense alignment emerged using data-driven solutions, where axiomatic shape models and deformations were replaced by learnable counterparts. Among those methods a highly successful research direction was based on learning local features overlaid on the vertices of the shapes [25], where ResNet [19] like architecture is used to update SHOT descriptors [48].

The main challenge new data-driven geometric alignment algorithms need to face is the lack of data to train on or labeled data used for supervised learning. In many cases, the labeled data is expensive to generate or even infeasible to acquire, as seen, for example, in medical imaging.

A recent approach [18] showed that self-supervised learning could be applied for non-rigid alignment between isometric shapes by preserving the pair-wise distance measured on the source and on the target. While showing good results, an isometric limitation is a strong constraint that is misused in many scenarios. On a different note, self-supervised learning was recently addressed in images, where a cyclic mapping between pictures, known as cyclic-GAN, was introduced [53, 28, 51]. The authors showed that given unpaired collection of images from different domains, a cyclic-loss that measures the distortion achieves robust state-of-the-art results in unsupervised learning for domain transfer. To this end, some papers have even shown promising insights when linking cyclic-constraints and alignment of the spectra [42, 20, 40, 21]. Here we show that using the structure's metric one can align the spectrum of non-isometric shapes.

In this work, we claim that one can learn dense correspondence in a self-supervised manner in between non-isometric structures. We present a new learnable cyclic mechanism, where the same model is used both for forward and backward mapping learning to compensate for deformed parts. We measure the pair-wise distance distortion of the cyclic mapping on randomly chosen pair of points only from the source manifold. We show here state-of-the-art-results by a large margin for a variety of tasks compared to self-supervised and supervised methods, in isometric and non-isometric setups.

2 Contribution

We present an unsupervised learning scheme for dense 3D correspondence between shapes, based on learnable self similarities between metric spaces. The proposed approach has multiple advantages over other known methods. First, there is no need to define a model for the shapes or the deformations; Second, it has no need for labeled data with dense correspondence mappings. Third, the proposed method can handle isometric or non-isometric deformations and partial matching. The cyclic mapping approach allows our system to learn the geometric representation of manifolds by feeding it pairs of matching shapes, even without any labels, by measuring a geometric criterion (pair-wise distance) only on the source.

Our main contribution is based on the understanding that a cyclic mapping between metric spaces which follows the same rules, forces the network to learn invariant parts. We built the cyclic mapping using the functional maps framework [30], optimizing for a soft correspondence between shapes on the spectral domains by updating a local descriptor per point. The proposed approach can be adapted to any dimension, and here we provide state-of-the-art results on surfaces. We show results that are comparable to supervised learning [25, 17, 23] methods in the rare case we possess dense correspondence labels, and outperforms self-supervised learning approaches [17, 18] when the shapes are isometric. Once the deformations are not isometric, our method stands out, and outperforms other methods by a large margin.

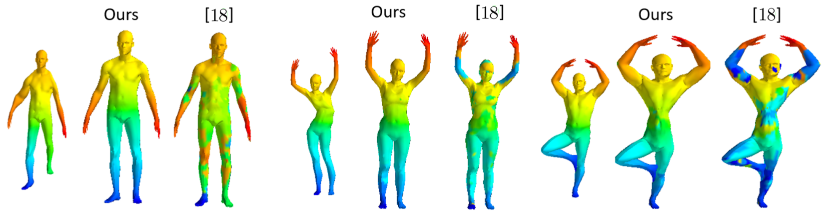


Fig. 3: Alignment between non-isometric shapes, where similar parts appear in similar colors. The shapes were locally scaled and stretched while changing their pose. Our approach learns the correct matching while [18] fails under local stretching.

3 Background

Our cyclic mapping is built on top of functional maps architecture. To explain the foundations of this approach, we must elaborate on distance matrices, functional maps and how to weave deep learning into functional maps. Finally, we discuss an isometric unsupervised approach for the alignment task and its limitations, which motivated this work.

3.1 Riemannian 2-Manifolds

We model 3D shapes as a Riemannian 2-manifold $(\mathcal{X}, g)$, where $\mathcal{X}$ is a real smooth manifold, equipped with an inner product g_p on the tangent space $T_p\mathcal{X}$ at each point p that varies smoothly from point to point in the sense that if $\mathcal{U}$ and $\mathcal{V}$ are differentiable vector fields on $\mathcal{X}$, then $p \to g_p(\mathcal{U}|_p, \mathcal{V}|_p)$ is a smooth function.

We equip the manifolds with a distance function $d\mathcal{X} : \mathcal{X} \times \mathcal{X} \to \Re$ induced by the standard volume form $d\mathcal{X}$. We state the distance matrix $\mathcal{D}_\mathcal{X}$, as a square symmetric matrix, represents the manifold's distance function $d\mathcal{X}$ such that

$$\mathcal{D}_{\mathcal{X}ij} = d\mathcal{X}(X_i, X_j). \tag{1}$$

3.2 Functional maps

Functional maps [30] stands for matching real-valued functions in between manifolds instead of performing a straight forward point matching. Using a spectral basis, one can extract a compact representation for a match on the spectral domain. The clear advantage here is that many natural constraints on the map become linear constraints on the functional map. Given two manifolds $\mathcal{X}$ and $\mathcal{Y}$, and functional spaces on top $F(\mathcal{X})$ and $F(\mathcal{Y})$, we can define a functional map using orthogonal bases ϕ and ψ

$$
\begin{aligned}
Tf &= T\sum_{i\geq 1}\langle f,\phi_i\rangle_{\mathcal{X}}\phi_i = \sum_{i\geq 1}\langle f,\phi_i\rangle_{\mathcal{X}}T\phi_i \\
&= \sum_{i,j\geq 1}\langle f,\phi_i\rangle_{\mathcal{X}}\underbrace{\langle T\phi_i,\psi_j\rangle_{\mathcal{Y}}}_{c_{ji}}\psi_j,
\end{aligned}
\tag{2}
$$

where $C \in \mathbb{R}^{k\times k}$ represents the mapping in between the domains given k matched functions, and every pair of corresponding functions on-top of the manifolds impose a linear constraint on the mapping. The coefficient matrix C is deeply dependent on the choice of the bases ϕ, ψ, and as shown in prior works [30, 38, 18] a good choice for such bases is the Laplacian eigenfunctions of the shapes.

3.3 Deep Functional Maps

Deep functional maps were first introduced in [25], where the mapping C between shapes was refined by learning new local features per point. The authors showed that using a ResNet [19] like architecture on-top of SHOT [48] descriptors, they can revise the local features in such a way that the global mapping is more accurate. The mapping is presented as a soft correspondence matrix P where P_{ji} is the probability $\mathcal{X}_i$ corresponds to $\mathcal{Y}_j$. The loss of the network is based on the geodesic distortion between the corresponding mapping and the ground truth, reading

$$
\mathcal{L}_{sup}(\mathcal{X},\mathcal{Y}) = \frac{1}{|\mathcal{X}|}\left\|\Big(P\circ(\mathcal{D}_{\mathcal{Y}}\Pi^*)\Big)\right\|_F^2, \tag{3}
$$

where $|\mathcal{X}|$ is the number of vertices of shape $\mathcal{X}$, and if $|\mathcal{X}| = n$, and $|\mathcal{Y}| = m$, then $\Pi^* \in \Re^{m\times n}$ is the ground-truth mapping between $\mathcal{X}$ and $\mathcal{Y}$, $\mathcal{D}_{\mathcal{Y}} \in \Re^{m\times m}$ is the geodesic distance matrix of $\mathcal{Y}$, $\circ$ denote a point-wise multiplication, and $||_F$ is the Frobenius norm. For each target vertex, the loss penalizes the distance between the actual corresponding vertex and the assumed one, multiplied by the amount of certainty the network has in that correspondence. Hence the loss is zero if

$$
P_{ji} = 1 \Leftrightarrow \Pi^*(\mathcal{X}_i) = \mathcal{Y}_j, \tag{4}
$$

as $\mathcal{D}(y,y) = 0\ \forall y \in \mathcal{Y}$.

3.4 Self-Supervised Deep Functional Maps

The main drawback of deep functional maps is the need for ground truth labels. Obtaining alignment maps for domains such as ours is a strong requirement, and is infeasible in many cases either due to the cost to generate those datasets, or even impractical to collect. In a recent paper [18], the authors showed that for isometric deformations, we can replace the ground truth requirement with a different geometric criterion based on pair-wise distances. In practice, they married together the Gromov-Hausdorff framework with the deep functional maps architecture.

The Gromov-Hausdorff distance which measures the distance between metric spaces, reads

$$d_{GH}(\mathcal{X}, \mathcal{Y}) = \frac{1}{2} \inf_{\pi}(dis(\pi)), \tag{5}$$

where the infimum is taken over all correspondence distortions of a given mapping $\pi \colon \mathcal{X} \to \mathcal{Y}$. This distortion can be translated to a pair-wise distance [15, 22] notation, which was used by [18] as a geometric criterion in the cost function of a deep functional map setup. Unfortunately, the pair-wise distance constraint is an extreme demand, forcing the models to be isometric, and can not be fulfilled in many practical scenarios.

4 Cyclic Self-Supervised Deep Functional Maps

The main contribution of this paper is the transition from the pair-wise distance comparison between source and target manifolds to a method that only examines the metric in the source manifold. Every pair of distances are mapped to the target and re-mapped back to the source. We use the same model for the forward and backward mapping to avoid a mode collapse, and we measure the distortion once a cyclic map has been completed, forcing the model to learn how to compensate for the deformations.

4.1 Correspondence Distortion

A mapping $\pi : X \to Y$ between two manifolds generates a pair-wise distortion

$$dis_{\pi}(\mathcal{X}, \mathcal{Y}) = \sum_{x_1, x_2 \in \mathcal{X}} \rho(d_{\mathcal{X}}(x_1, x_2), d_{\mathcal{Y}}(\pi(x_1), \pi(x_2)), \tag{6}$$

where ρ is usually an L_p norm metric, and $p = 2$ is a useful choice of the parameter.

As isometric mapping preserves pair-wise distances, minimizing the distances between those pairs provides a good metric-dependent correspondence. Specifically,

$$\pi^{iso}(\mathcal{X}, \mathcal{Y}) = \operatorname*{argmin}_{\pi : \mathcal{X} \to \mathcal{Y}} dis_{\pi}(\mathcal{X}, \mathcal{Y}). \tag{7}$$

Solving 7 takes the form of a quadratic assignment problem. The main drawback of this criterion, as the name suggests, is the isometric assumption. While it is a powerful tool for isometric mappings, natural phenomena do not follow that convention as stretching exists in the data. To overcome those limitations, we present here the cyclic distortion criterion.

4.2 Cyclic Distortion

We define a **cyclic distortion**[1] π^{cyc} as a composition of two mappings $\pi_{\rightarrow} : \mathcal{X} \rightarrow \mathcal{Y}$ and $\pi_{\leftarrow} : \mathcal{Y} \rightarrow \mathcal{X}$, which leads to a cyclic distortion

$$dis^{cyc}_{(\pi_{\rightarrow},\pi_{\leftarrow})}(\mathcal{X},\mathcal{Y}) = \sum_{x_1,x_2\in\mathcal{X}} \rho(d_{\mathcal{X}}(x_1,x_2), d_{\mathcal{X}}(\tilde{x}_1,\tilde{x}_2)), \tag{8}$$

where $\tilde{x}_1 = \pi_{\leftarrow}(\pi_{\rightarrow}(x_1))$ and $\tilde{x}_2 = \pi_{\leftarrow}(\pi_{\rightarrow}(x_2))$.

$\pi_{\rightarrow}$ and $\pi_{\leftarrow}$ are being optimized using the same sub-network, implemented as shared weights in the learning process. Every forward mapping $\pi_{\rightarrow}$ induce a backward mapping $\pi_{\leftarrow}$ and vice-versa. We call this coupled pair $\pi = (\pi_{\rightarrow}, \pi_{\leftarrow})$ a *conjugate mapping*, and denote the space of all conjugate mappings by $\mathcal{S}$. We define the cyclic mapping as

$$\pi^{cyc}(\mathcal{X},\mathcal{Y}) = \underset{\pi:(\mathcal{X}\rightarrow\mathcal{Y},\mathcal{Y}\rightarrow\mathcal{X})\in\mathcal{S}}{\operatorname{argmin}} dis^{cyc}_{\pi}(\mathcal{X},\mathcal{Y}). \tag{9}$$

4.3 Deep Cyclic Mapping

Following the functional map convention, given C, Φ, Ψ the *soft correspondence matrix* mapping between $\mathcal{X}$ to $\mathcal{Y}$ reads [25]

$$P = \left|\Phi C \Psi^T\right|_{\mathcal{F}_c}, \tag{10}$$

where each entry P_{ji} is the probability point j in $\mathcal{X}$ corresponds to point i in $\mathcal{Y}$. We further use $|\cdot|_{\mathcal{F}_c}$ notation for column normalization, to emphasize the statistical interpretation of P.

Let P represents the forward mapping $\pi_{\rightarrow}$ soft correspondence and $\tilde{P}$ the backward mapping $\pi_{\leftarrow}$. The cyclic distortion is defined by

$$\mathcal{L}_{cyclic}(\mathcal{X},\mathcal{Y}) = \frac{1}{|\mathcal{X}|^2} \left\| \left(D_{\mathcal{X}} - (\tilde{P}P) D_{\mathcal{X}} (\tilde{P}P)^T \right) \right\|^2_{\mathcal{F}}, \tag{11}$$

where $|\mathcal{X}|$ is the number of samples point pairs on $\mathcal{X}$.

Note that if we assumed the shapes were isometric, then we would have expected $D_{\mathcal{Y}}$ to be similar or even identical to $P D_{\mathcal{X}} P^T$, which yields once plugged into (11) the isometric constraint

$$\mathcal{L}_{isometric}(\mathcal{X},\mathcal{Y}) = \frac{1}{|\mathcal{X}|^2} \left\| \left(D_{\mathcal{X}} - \tilde{P} D_Y \tilde{P}^T \right) \right\|^2_{\mathcal{F}}. \tag{12}$$

The cyclic distortion (11) is self-supervised, as no labels are required, and only use the pair-wise distances on the source manifold $\mathcal{X}$. The conjugate mappings are based on the functional maps architecture and use the geometry of both spaces, source and target. Since we constrain the mapping on the source's geometry, the mapping copes with stretching, and thus learning invariant representations.

---

[1] An illustration of the distortion process is shown in figure 2

5 Implementation

5.1 Hardware

The network was developed in TensorFlow [1], running on a GeForce GTX 2080 Ti GPU. The SHOT descriptor [48] was implemented in MATLAB, while the Laplace Beltrami Operator (LBO) [41] and geodesic distances were calculated in Python.

5.2 Pre-processing

We apply a sub-sampling process for shapes with more than 10,000 points using qSlim [16] algorithm. SHOT descriptor was computed on the sub-sampled shapes, generating a descriptor of length $s = 350$ per vertex. Finally, the LBO eigenfunctions corresponding to the most significant 70 eigenvalues (lowest by magnitude) were computed for each shape. The distance matrices were computed using the Fast Marching algorithm [43]. In order to initialize the conjugate mapping, we found that a hard constraint on P and $\tilde{P}$ coupling provides good results. Specifically we minimized in the first epoch the cost function $||P\tilde{P} - I||^2_{\mathcal{F}}$ before applying the soft cyclic criterion (11). Equation 11 consists of four probability matrices multiplication which leads to low values in the cost function and its derivatives. Hence, we found an advantage to start the optimization process with a simpler cost function in the first epoch. We examined several configurations of loss functions, including pre-training on non-cyclic loss, yet, we did not observe any improvement over direct optimization of equation (9).

5.3 Network Architecture

The architecture is motivated by [25, 18] and shown in Figure 4. The input to the first layer is the raw 3D triangular mesh representations of the two figures given by a list of vertices and faces. We apply a multi-layer SHOT [48] descriptor by evaluating the SHOT on $m \sim 5$ global scaled versions of the input. The figures vary from 0.2 to 2 times the size of the original figures, followed by a 1x1 convolution layers with $2m$ filters, to a 1×1 convolution layer with one filter, generating an output of $n \times s$ descriptor to the network. Besides, the relevant eigenfunctions and pair-wise distance matrix of the source shape are provided as parameters to the network.

The next stage is the ResNet [19] layers with the shared weights applied to both figures. Subsequently, the non-linear descriptors are multiplied by the $n \times k$ LBO eigenfunctions. We calculate the forward and backward mappings C and $\tilde{C}$ using the same network and evaluate the corresponding forward and backward mappings P and $\tilde{P}$, which are fed into the soft cyclic loss (11).

The fact we do not use geodesic distance metric on the target shape $D_{\mathcal{Y}}$ contributed to faster inference time, and lighter training batches, which accelerated the learning with comparison to [18, 25].

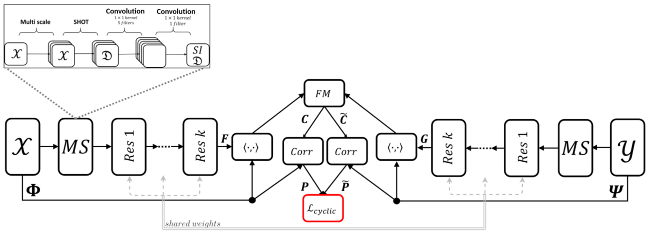


Fig. 4: Cyclic functional mapper in between two manifolds $\mathcal{X}$ and $\mathcal{Y}$ (left and right sides). The multi-scaled descriptors (top left, marked MS) based on shot [48] are passed to a ResNet like network, resulting in two corresponding coefficient matrices F and G. By projecting the refined descriptors onto the spectral space, two mappings, C and $\tilde{C}$, are computed. The two soft correspondence matrices P and $\tilde{P}$ are further used as part of the network cyclic loss $\mathcal{L}_{cyclic}$ as shown in equation 11.

6 Experiments

In this section, we present multiple experiments in different settings; synthetic and real layouts, transfer learning tasks, non-isometric transformations, partial matching and one-shot learning. We show benchmarks, as well as comparisons to state-of-the-art solutions, both for axiomatic and learned algorithms.

6.1 Mesh Error Evaluation

The measure of error for the correspondence mapping between two shapes will be according to the Princeton benchmark [22], that is, given a mapping $\pi_{\rightarrow}(\mathcal{X}, \mathcal{Y})$ and the ground truth $\pi^*_{\rightarrow}(\mathcal{X}, \mathcal{Y})$ the error of the correspondence matrix is the sum of geodesic distances between the mappings for each point in the source figure, divided by the area of the target figure.

$$\epsilon(\pi_{\rightarrow}) = \sum_{x \in \mathcal{X}} \frac{\mathcal{D}_{\mathcal{Y}}(\pi_{\rightarrow}(x), \pi^*_{\rightarrow}(x))}{\sqrt{area(\mathcal{Y})}}, \tag{13}$$

where the approximation of $area(\bullet)$ for a triangular mesh is the sum of its triangles area.

6.2 Synthetic FAUST

We compared our alignment on FAUST dataset [7] versus supervised [25] and unsupervised [18] methods. We followed the experiment as described in [25] and used the synthetic human shapes dataset, where the first 80 shapes (8 subjects with 10 different poses each) are devoted to training, and 20 shapes made of 2 different unseen subjects are used for testing.

For a fair comparison between methods, we did not run the PMF cleanup filter [50] as this procedure is extremely slow and takes about 15 minutes for one shape built of $\sim 7k$ vertices on an i9 desktop.

We do not perform any triangular mesh preprocessing on the dataset, that is, we learn on the full resolution of 6890 vertices.

Each mini-batch is of size 4 (i.e 4 pairs of figures), using $k = 120$ eigenfunctions, and 10 bins in SHOT with a radius set to be 5% of the geodesic distance from each vertex. We report superior results for inter-subject and intra-subject tasks in Table 1, while converging faster (see Figure 6).

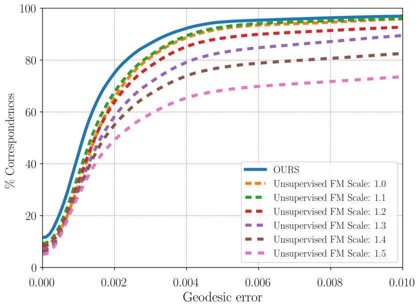
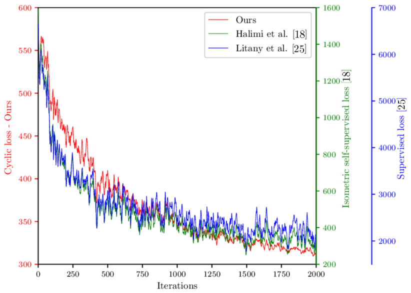


Fig. 5: Scale-Invariant Attributes - Network performance under global-scaling transformations

As shown in figure [5] global scaling is transparent to our cyclic consistent loss. We provide several evaluations on the Synthetic FAUST dataset where the target shape was global scaled by different factors during training. While we converge to the same result in each experiment, unsupervised methods as [18] that assume isometry fail to learn in such frameworks.



Method	Loss during their training	Loss during our training
Litany *et al.*[25]	~ 180	~ 150
Halimi *et al.*[18]	~ 20	~ 15

Fig. 6: We visualize our cyclic loss, the isometric constrained unsupervised loss [18], and the supervised loss [25] during the training of our cyclic loss on the synthetic FAUST dataset. We show that minimization of the cyclic loss on isometric structures provides lower loss values on previous architectures, without any geometric assumption.

6.3 Real Scans

We tested our method on real 3D scans of humans from the FAUST [7] dataset. While the synthetic samples had $\sim 6k$ vertices, each figure in this set has $\sim 150k$ vertices, creating the amount of plausible cyclic mappings extremely high. The dataset consists of multiple subjects in a variety of poses, where none of the poses (e.g., women with her hands up) in the test set were present in the training set. The samples were acquired using a full-body 3D stereo capture system, resulting in missing vertices, and open-end meshes.

The dataset is split into two test cases as before, the intra and inter subjects (60 and 40 pairs respectively), and ground-truth correspondences is not given. Hence, the geodesic error evaluation is provided as an online service. As suggested in [25], after evaluating the soft correspondence mappings, we input our map to the PMF algorithm [50] for a smoother bijective correspondence refined map. We report state of the art results on both inter and intra class mappings in comparison to all the unsupervised techniques. We provide visualization in Figure 7 and qualitative results in table 1.

Table 1: Average error on the FAUST dataset measured as distance between mapped points and the ground truth. We compared between our approach and other supervised and unsupervised methods.

Method	Scans[cm]		Synthetic[cm]	
	inter	intra	inter	intra
Ours	**4.068**	**2.12**	**2.327**	**2.112**
Litany *et al.*[25]	4.826	2.436	2.452	2.125
Halimi *et al.*[18]	4.883	2.51	3.632	2.213
Groueix[a] *et al.*[17]	4.833	2.789	—	—
Li *et al.*[23]	4.079	2.161	—	—
Chen *et al.*[13]	8.304	4.86	—	—

---

[a] Unsupervised training.

Fig. 7: Correspondence on FAUST real scans dataset, where similar colors represent the same correspondence. This dataset contains shapes made of $\sim 100k$ vertices with missing information in various poses. We use a post-matching PMF filter [50], and show qualitative results in Table 1. We outperform both supervised and unsupervised methods.

6.4 Non-Isometric Deformations

An even bigger advantage of the proposed method is its ability to cope with local stretching. Due to the cyclic mapping approach, we learn local matching features directly between the models and are not relying on a base shape in the latent space or assume isometric consistency. We experimented with models generated in Autodesk Maya that were locally stretched and bent. We remark that none of the stretched samples was present during training, which was done on the "standard" FAUST real scans dataset. We show visual results in Figure 3. The proposed approach successfully handles large non-isometric deformations.

6.5 TOSCA

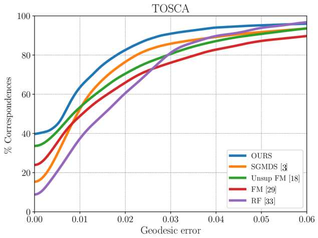


Fig. 8: Geodesic error on TOSCA dataset. We report superior results against other supervised and unsupervised learnable methods. Note that the compared methods did not run a post-processing optimization-based filter, or received partial matching as input.

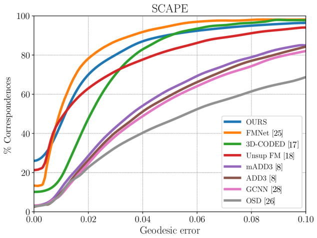


Fig. 9: Geodesic error on SCAPE dataset. Our network was trained on FAUST dataset and used to predict the mapping on SCAPE. We provide superior results on all unspervised and almost all supervised methods showing good generalization properties.

The TOSCA dataset [10] consists of 80 objects from different domains as animals and humans in different poses. Although the animals are remarkably different in terms of LBO decomposition, as well as geometric characteristics, our model achieves excellent performance in terms of a geodesic error on the dataset after training for a single epoch on it, using the pre-trained model from the real scans FAUST dataset [1].

In Figure 8, we show a comparison between our and other supervised and unsupervised approaches and visualize a few samples in Figure 1. Compared methods results were taken from [18]. Our network was trained for a single epoch on the dataset, with a pre-trained model of the real scans FAUST data and yet, shows great performance. We report state of the art results, compared to axiomatic, supervised, and unsupervised methods. Also note that while other methods mention training on each class separately, we achieve state-of-the-art results while training jointly.

6.6 SCAPE

To further emphasize the generalization capabilities of our network, we present our results on the SCAPE dataset [46], which is an artificial human shape dataset, digitally generated, with completely different properties from the FAUST dataset in any aspect (geometric entities, scale, ratio, for example). Nevertheless, our network that was trained on the real scan FAUST dataset performs remarkably well. See Figure 9. Compared methods results were taken from [18].

6.7 One-shot Single Pair Learning

Following the experiment shown in [18], we demonstrate that we can map in between two shapes seen for the first time without training on a large dataset.

Compared to optimization approaches we witness improved running time due to optimized hardware and software dedicated to deep learning in recent years. In Figure 10 we show such a mapping in between highly deformed shapes, and we found it intriguing that a learning method based on just two samples can converge to a feasible solution even without strong geometric assumptions. Note that in that case methods based on isometric criterion fail to converge due to the large non-isometric deformation. Running multiple epochs on two shapes not only converged to a pleasing mapping, but we found it to work well without any priors on the shapes or a need to engineer the initialization process. That is true for inter-class as well as highly deformed shapes (See figure 10) In this experiment we used our multi-SHOT pre-trained weights before we ran our cyclic mapper.

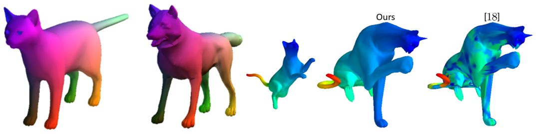


Fig. 10: Left - Highly deformed single pair, inter-class experiment.
Right - Single pair (one shot) learning on deformable non-isometric shapes. Supervised methods as [25] are irrelevant, where isometric self-learning approach fails [18].

6.8 Partial Shapes Correspondence

The partial shapes correspondence task is inherently more complicated than the full figure alignment problems. While in most experiments shown above, the number of vertices in both shapes differed by less than 5%, in the partial shapes task, we consider mappings between objects that differ by a large margin of up to 75% in their vertex count. To this end, numerous bijective solutions, such as [44, 49, 50] show degraded performance on the partial challenge, resulting in targeted algorithms [38, 26] for the mission. With that in mind, we show our results on the SHREC 2016 [14] partial shapes dataset [12][11]. We remark that our formulation does not require the map to be bijective. As $P \in \mathbb{R}^{n_1 \times n_2}$,$\tilde{P} \in \mathbb{R}^{n_2 \times n_1}$ where n_1, n_2 are the number of vertices in the shapes no bijection is enforced, rather than the composition is close to the identity. Thus, the distortion minimization function finds the best mapping it can even for partial to full correspondence.

Fig. 11: Partial to Full correspondence - Although the mapping is ill-posed by nature in the missing segments, the cyclic mapper acheives near-perfect results within the valid areas of the source shape.

We use the same architecture as described earlier, given hyperparameters and trained weights from the TOSCA 6.5 experiment, showing our network's generalization capabilities. As before, we have trained the network on this dataset only for a single epoch.

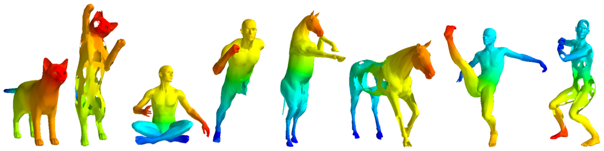


Fig. 12: Partial shapes correspondence on SHREC 16 [14] dataset after removing substantial parts (up to 75%). In every pair, we mapped the left shape into the right one, where similar mapped points share the color. Our method is robust to missing information.

7 Limitations

Our method uses functional maps architecture, which requires us to pre-compute sets of bases functions. To that end, this process can not be done in real-time in the current setup, and there might be an inconsistency in bases functions between shapes due to noise or large non-isometric deformations. While this method works well for isometric or stretchable domains, once the deformations are significantly large, we found that the current system does not converge to a reasonable geodesic error in terms of a pleasant visual solution, which makes it challenging to use in cross-domain alignments. We believe that the proposed approach can be used as part of semantic-correspondence to overcome those limitations. Furthermore, experimenting with mappings between partial shapes showed us (see Figure 11), that missing parts can suffers from ambiguity and can appear as a non-smooth mapping.

8 Summary

We presented here a cyclic architecture for dense correspondence between shapes. In the heart of our network is the bidirectional mapper, which jointly learns the mapping from the source to the target and back via a siamese-like architecture. We introduced a novel loss function which measures the distortion only on the source, while still using both geometries, allowing us to cope with non-isometric deformations. This approach is self-supervised, can cope with local stretching as well as non-rigid isometric deformations. While our concept specifically addresses the wrong assumption of isometry for inter-class subjects, we see superior results even for intra-class datasets. Our method outperforms other unsupervised and supervised approaches on tested examples, and we report state-of-the-art results in several scenarios, including real 3D scans and partial matching task.

Acknowledgment

D.R. is partially funded by the Zimin Institute for Engineering Solutions Advancing BetterLives, the Israeli consortiums for soft robotics and autonomous driving, and the Shlomo Shmeltzer Institute for Smart Transportation.

References

1. Abadi, M., Agarwal, A., Barham, P., Brevdo, E., Chen, Z., Citro, C., Corrado, G.S., Davis, A., Dean, J., Devin, M., Ghemawat, S., Goodfellow, I., Harp, A., Irving, G., Isard, M., Jia, Y., Jozefowicz, R., Kaiser, L., Kudlur, M., Levenberg, J., Mané, D., Monga, R., Moore, S., Murray, D., Olah, C., Schuster, M., Shlens, J., Steiner, B., Sutskever, I., Talwar, K., Tucker, P., Vanhoucke, V., Vasudevan, V., Viégas, F., Vinyals, O., Warden, P., Wattenberg, M., Wicke, M., Yu, Y., Zheng, X.: TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems (2015), http://tensorflow.org/, software available from tensorflow.org
2. Aflalo, Y., Bronstein, A., Kimmel, R.: On convex relaxation of graph isomorphism. Proceedings of the National Academy of Sciences **112**(10), 2942–2947 (2015)
3. Aflalo, Y., Dubrovina, A., Kimmel, R.: Spectral Generalized Multi-dimensional Scaling. International Journal of Computer Vision **118**(3), 380–392 (2016)
4. Aflalo, Y., Kimmel, R., Raviv, D.: Scale invariant geometry for nonrigid shapes. SIAM Journal on Imaging Sciences **6**(3), 1579–1597 (2013)
5. Aubry, M., Schlickewei, U., Cremers, D.: The wave kernel signature: A quantum mechanical approach to shape analysis. In: 2011 IEEE international conference on computer vision workshops (ICCV workshops). pp. 1626–1633. IEEE (2011)
6. Ben-Chen, M., Gotsman, C., Bunin, G.: Conformal Flattening by Curvature Prescription and Metric Scaling. In: Computer Graphics Forum. vol. 27, pp. 449–458. Wiley Online Library (2008)
7. Bogo, F., Romero, J., Loper, M., Black, M.J.: FAUST: Dataset and evaluation for 3D mesh registration. In: Proceedings IEEE Conf. on Computer Vision and Pattern Recognition (CVPR). IEEE, Piscataway, NJ, USA (Jun 2014)
8. Boscaini, D., Masci, J., Rodolà, E., Bronstein, M.M., Cremers, D.: Anisotropic Diffusion Descriptors. Comput. Graph. Forum **35**, 431–441 (2016)
9. Bronstein, A.M., Bronstein, M.M., Kimmel, R.: Generalized multidimensional scaling: a framework for isometry-invariant partial surface matching. Proceedings of the National Academy of Sciences **103**(5), 1168–1172 (2006)
10. Bronstein, A.M., Bronstein, M.M., Kimmel, R.: Numerical geometry of non-rigid shapes. Springer Science & Business Media (2008)
11. Bronstein, M.M., Bronstein, A.M., Kimmel, R., Yavneh, I.: Multigrid multidimensional scaling. Numerical linear algebra with applications **13**(2-3), 149–171 (2006)
12. Bronstein, M.M., Kokkinos, I.: Scale-invariant heat kernel signatures for non-rigid shape recognition. In: 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. pp. 1704–1711. IEEE (2010)
13. Chen, Q., Koltun, V.: Robust Nonrigid Registration by Convex Optimization. In: Proceedings of the IEEE International Conference on Computer Vision. pp. 2039–2047 (2015)
14. Cosmo, L., Rodolà, E., Bronstein, M.M., Torsello, A., Cremers, D., Sahillioglu, Y.: SHREC'16: Partial matching of deformable shapes
15. Elad, A., Kimmel, R.: On bending invariant signatures for surfaces. IEEE Transactions on pattern analysis and machine intelligence **25**(10), 1285–1295 (2003)
16. Garland, M., Heckbert, P.S.: Surface simplification using quadric error metrics. In: Proceedings of the 24th annual conference on Computer graphics and interactive techniques. pp. 209–216. ACM Press/Addison-Wesley Publishing Co. (1997)
17. Groueix, T., Fisher, M., Kim, V.G., Russell, B.C., Aubry, M.: 3D-CODED : 3D Correspondences by Deep Deformation. CoRR **abs/1806.05228** (2018), http://arxiv.org/abs/1806.05228

18. Halimi, O., Litany, O., Rodola, E., Bronstein, A.M., Kimmel, R.: Unsupervised Learning of Dense Shape Correspondence. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 4370–4379 (2019)
19. He, K., Zhang, X., Ren, S., Sun, J.: Deep Residual Learning for Image Recognition. CoRR **abs/1512.03385** (2015), http://arxiv.org/abs/1512.03385
20. Huang, Q.X., Guibas, L.: Consistent shape maps via semidefinite programming. In: Computer Graphics Forum. vol. 32, pp. 177–186. Wiley Online Library (2013)
21. Huang, Q., Wang, F., Guibas, L.: Functional map networks for analyzing and exploring large shape collections. ACM Transactions on Graphics (TOG) **33**(4), 1–11 (2014)
22. Kim, V.G., Lipman, Y., Funkhouser, T.: Blended intrinsic maps. In: ACM Transactions on Graphics (TOG). vol. 30, p. 79. ACM (2011)
23. Li, C.L., Simon, T., Saragih, J., Póczos, B., Sheikh, Y.: LBS Autoencoder: Self-supervised Fitting of Articulated Meshes to Point Clouds. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 11967–11976 (2019)
24. Lipman, Y., Daubechies, I.: Conformal Wasserstein distances: Comparing surfaces in polynomial time. Advances in Mathematics **227**(3), 1047–1077 (2011)
25. Litany, O., Remez, T., Rodolà, E., Bronstein, A.M., Bronstein, M.M.: Deep Functional Maps: Structured prediction for dense shape correspondence. CoRR **abs/1704.08686** (2017), http://arxiv.org/abs/1704.08686
26. Litany, O., Rodolà, E., Bronstein, A.M., Bronstein, M.M.: Fully Spectral Partial Shape Matching. In: Computer Graphics Forum. vol. 36, pp. 247–258. Wiley Online Library (2017)
27. Litman, R., Bronstein, A.M.: Learning spectral descriptors for deformable shape correspondence. IEEE transactions on pattern analysis and machine intelligence **36**(1), 171–180 (2013)
28. Liu, M.Y., Breuel, T., Kautz, J.: Unsupervised Image-to-Image Translation Networks. In: Advances in neural information processing systems. pp. 700–708 (2017)
29. Masci, J., Boscaini, D., Bronstein, M., Vandergheynst, P.: Geodesic convolutional neural networks on Riemannian manifolds. In: Proceedings of the IEEE international conference on computer vision workshops. pp. 37–45 (2015)
30. Ovsjanikov, M., Ben-Chen, M., Solomon, J., Butscher, A., Guibas, L.: Functional maps: a flexible representation of maps between shapes. ACM Transactions on Graphics (TOG) **31**(4), 30 (2012)
31. Pottmann, H., Wallner, J., Huang, Q.X., Yang, Y.L.: Integral Invariants for Robust Geometry Processing. Computer Aided Geometric Design **26**(1), 37–60 (2009)
32. Raviv, D., Bronstein, A.M., Bronstein, M.M., Waisman, D., Sochen, N., Kimmel, R.: Equi-affine invariant geometry for shape analysis. Journal of mathematical imaging and vision **50**(1-2), 144–163 (2014)
33. Raviv, D., Bronstein, M.M., Bronstein, A.M., Kimmel, R., Sochen, N.: Affine-invariant diffusion geometry for the analysis of deformable 3d shapes. In: CVPR 2011. pp. 2361–2367. IEEE (2011)
34. Raviv, D., Dubrovina, A., Kimmel, R.: Hierarchical matching of non-rigid shapes. In: International Conference on Scale Space and Variational Methods in Computer Vision. pp. 604–615. Springer (2011)
35. Raviv, D., Dubrovina, A., Kimmel, R.: Hierarchical framework for shape correspondence. Numerical Mathematics: Theory, Methods and Applications **6**(1), 245–261 (2013)
36. Raviv, D., Kimmel, R.: Affine invariant geometry for non-rigid shapes. International Journal of Computer Vision **111**(1), 1–11 (2015)

37. Raviv, D., Raskar, R.: Scale invariant metrics of volumetric datasets. SIAM Journal on Imaging Sciences **8**(1), 403–425 (2015)
38. Rodolà, E., Cosmo, L., Bronstein, M.M., Torsello, A., Cremers, D.: Partial functional correspondence. In: Computer Graphics Forum. vol. 36, pp. 222–236. Wiley Online Library (2017)
39. Rodolà, E., Rota Bulo, S., Windheuser, T., Vestner, M., Cremers, D.: Dense Non-Rigid Shape Correspondence using Random Forests. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 4177–4184 (2014)
40. Roufosse, J.M., Sharma, A., Ovsjanikov, M.: Unsupervised deep learning for structured shape matching. In: Proceedings of the IEEE International Conference on Computer Vision. pp. 1617–1627 (2019)
41. Rustamov, R.M.: Laplace-beltrami eigenfunctions for deformation invariant shape representation. In: Proceedings of the Fifth Eurographics Symposium on Geometry Processing. pp. 225–233. SGP '07, Eurographics Association, Aire-la-Ville, Switzerland, Switzerland (2007), http://dl.acm.org/citation.cfm?id=1281991.1282022
42. Rustamov, R.M., Ovsjanikov, M., Azencot, O., Ben-Chen, M., Chazal, F., Guibas, L.: Map-based exploration of intrinsic shape differences and variability. ACM Transactions on Graphics (TOG) **32**(4), 1–12 (2013)
43. Sethian, J.A.: A fast marching level set method for monotonically advancing fronts. Proceedings of the National Academy of Sciences **93**(4), 1591–1595 (1996)
44. Starck, J., Hilton, A.: Spherical Matching for Temporal Correspondence of Non-Rigid Surfaces. In: Tenth IEEE International Conference on Computer Vision (ICCV'05) Volume 1. vol. 2, pp. 1387–1394. IEEE (2005)
45. Sun, J., Ovsjanikov, M., Guibas, L.: A Concise and Provably Informative Multi-Scale Signature Based on Heat Diffusion. In: Computer graphics forum. vol. 28, pp. 1383–1392. Wiley Online Library (2009)
46. Szeliski, R., Anguelov, D., Srinivasan, P., Koller, D., Thrun, S., Rodgers, J., Davis, J.: SCAPE: shape completion and animation of people, vol. 24 (2005)
47. Tevs, A., Berner, A., Wand, M., Ihrke, I., Seidel, H.P.: Intrinsic shape matching by planned landmark sampling. In: Computer Graphics Forum. vol. 30, pp. 543–552. Wiley Online Library (2011)
48. Tombari, F., Salti, S., Di Stefano, L.: Unique Signatures of Histograms for Local Surface Description. In: Computer Vision – ECCV 2010. pp. 356–369 (2010)
49. Vestner, M., Lähner, Z., Boyarski, A., Litany, O., Slossberg, R., Remez, T., Rodola, E., Bronstein, A., Bronstein, M., Kimmel, R., et al.: Efficient deformable shape correspondence via kernel matching. In: 2017 International Conference on 3D Vision (3DV). pp. 517–526. IEEE (2017)
50. Vestner, M., Litman, R., Rodolà, E., Bronstein, A., Cremers, D.: Product manifold filter: Non-rigid shape correspondence via kernel density estimation in the product space. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 3327–3336 (2017)
51. Yi, Z., Zhang, H., Tan, P., Gong, M.: Dualgan: Unsupervised dual learning for image-to-image translation. In: Proceedings of the IEEE international conference on computer vision. pp. 2849–2857 (2017)
52. Zaharescu, A., Boyer, E., Varanasi, K., Horaud, R.: Surface Feature Detection and Description with Applications to Mesh Matching. In: 2009 IEEE Conference on Computer Vision and Pattern Recognition. pp. 373–380. IEEE (2009)
53. Zhu, J.Y., Park, T., Isola, P., Efros, A.A.: Unpaired Image-to-Image Translation using Cycle-Consistent Adversarial Networks. In: Proceedings of the IEEE international conference on computer vision. pp. 2223–2232 (2017)