



HAL
open science

Weakly Supervised Deep Functional Map for Shape Matching

Abhishek Sharma, Maks Ovsjanikov

► **To cite this version:**

Abhishek Sharma, Maks Ovsjanikov. Weakly Supervised Deep Functional Map for Shape Matching. 2020. hal-02872053v1

HAL Id: hal-02872053

<https://hal.science/hal-02872053v1>

Preprint submitted on 18 Jun 2020 (v1), last revised 28 Sep 2020 (v2)

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.

Weakly Supervised Deep Functional Map for Shape Matching

Abhishek Sharma
LIX, École Polytechnique
kein.iitian@gmail.com

Maks Ovsjanikov
LIX, École Polytechnique
maks@lix.polytechnique.fr

Abstract

A variety of deep functional maps have been proposed recently, from fully supervised to totally unsupervised, with a range of loss functions as well as different regularization terms. However, it is still not clear what are minimum and sufficient ingredients of a deep functional map pipeline and whether such ingredients unify or generalize all recent work on deep functional maps. With a slight abuse of notation, we show minimum and sufficient conditions for obtaining state of the art results with any loss function, supervised as well as unsupervised. Furthermore, we propose a novel framework designed for both full-to-full as well as partial to full shape matching that achieves state of the art results on benchmark datasets outperforming even the fully supervised methods by a large margin. Our code will be publicly available at <https://github.com/Not-IITian/Weakly-supervised-Functional-map>

1 Introduction

Shape correspondence is a fundamental problem in computer vision, computer graphics and related fields since it facilitates many applications such as texture or deformation transfer and statistical shape analysis Bogo et al. [2014] to name a few. While classical correspondence methods have been based on handcrafted features or deformation models Van Kaick et al. [2011], more recent approaches have focused on *learning* an optimal model directly from 3D data. This includes approaches based on template fitting and reconstruction Groueix et al. [2018, 2019], and methods that exploit different definitions of convolution and phrase correspondence as a dense labeling problem Wei et al. [2016], Masci et al. [2015], Boscaini et al. [2016] among others.

A prominent direction in learning-based shape matching was pioneered by the FMNet work Litany et al. [2017a] by exploiting the *functional map representation* Ovsjanikov et al. [2012] and learning features that recover optimal functional maps rather than e.g. individual point labels. The use of the functional map representation allows to efficiently impose global correspondence constraints, and has been recently extended in both unsupervised Halimi et al. [2019], Roufousse et al. [2019] and supervised settings Donati et al. [2020]. Despite significant progress in this area, there still exist three major issues. First, the most accurate recent approach Donati et al. [2020] is limited to supervised setting that requires ground truth correspondences that are difficult to obtain considering the cost of annotating a point to point map from the data. Second, despite a variety of deep functional maps-based methods, it is still not clear what are minimum and sufficient ingredients of a deep functional map pipeline. More importantly, do such minimum and sufficient ingredients unify or generalize all recent work on deep functional maps. While a battery of loss functions and regularization have been proposed for different deep functional maps, as we demonstrate below, the devil is not in the loss functions. Instead, using a low number of Laplacian eigen basis, very weak supervision in the form of rigid alignment and enforcing basic structural properties of resulting functional map are sufficient. Moreover, our findings generalize to all loss functions proposed recently. Third, recent

learning-based approaches are neither designed nor tested for the *partial shape matching problem* [Rodolà et al., 2017, Litany et al., 2017b] which is of great interest in robotics [Chavdar et al., 2012] and Virtual reality applications [Sharma et al., 2016]. To this end, we propose a weakly supervised framework that addresses all three major issues.

In this paper, weak supervision implies that datasets are only approximately rigidly aligned. Broadly, there are three main components to any deep functional map pipeline, namely feature extractor, choice of basis functions and design of empirical loss or regularization on functional Map. In this paper, we make contributions on all three fronts. First, we propose to learn feature descriptors directly from raw data with a very weak supervision and establish that for non-rigid shape correspondence, rigid alignment supervision turns out to be sufficient to obtain high quality results. Remarkably, it also outperforms the fully supervised state-of-the-art methods, which rely on ground truth point-to-point correspondences, on challenging benchmarks by a large margin. Secondly, we show that combination of our feature extractor projected to low number of Laplacian eigen basis (30) and unsupervised loss, consisting of regularization terms, is empirically a minimum and sufficient ingredient to obtain state of the art result for any recently proposed loss functions. Thirdly, to address partial shape matching, we propose a novel data driven method to learn an optimal alignment between source and target Laplacian eigen basis functions which paves the way for future work on deep functional maps in partial shape matching.

2 Related Work

Functional Maps Computing point-to-point maps between two 3D discrete surfaces is a very well-studied problem. We refer to a recent survey Sahillioğlu [2019] for an in-depth discussion. Our method builds upon the functional map pipeline, introduced in Ovsjanikov et al. [2012] and then significantly extended in follow-up works Ovsjanikov et al. [2017]. Functional maps encode correspondences as small matrices, expressed in a reduced basis, which greatly simplifies the associated optimization problems. A range of recent works, including Kovnatsky et al. [2013], Huang et al. [2014], Burghard et al. [2017], Rodolà et al. [2017], Nogneng and Ovsjanikov [2017], Ren et al. [2018] among many others, have extended the generality and improved the robustness of the functional map estimation pipeline, by suggesting regularizers, robust penalties and powerful post-processing approaches. Nevertheless, existing non-learning based methods are strongly tied to the choice of descriptor (also known as “probe”) functions, which must be specified manually a priori. We also note that there also exist other techniques that learn correspondences without using the functional map representation, e.g., Wei et al. [2016], Boscaini et al. [2016], Monti et al. [2017]. However, such techniques typically either require significantly more training data (essentially because they treat shape correspondence as a dense labeling problem with a very large number of labels), or do not learn from 3D geometry which is the main goal of this paper.

Supervised Learning from raw 3D shape In contrast to axiomatic approaches that use hand-crafted features, a variety of methods have also been proposed to *learn* the optimal features or descriptors from 3D data. In the functional maps domain, this was first suggested in Corman et al. [2014] using classical optimisation techniques and then in the seminal Deep Functional Maps work Litany et al. [2017a] that proposed a deep learning architecture called FMNet to compute optimal features from data. This architecture was based on optimizing a non-linear transformation of SHOT descriptors Tombari et al. [2010] to obtain maps that are as close as possible to given ground truth correspondences. Follow-up works have extended this approach to the unsupervised setting Roufousse et al. [2019], Halimi et al. [2019] by modifying the training loss, but still used pre-defined descriptors for optimization. These methods generalize poorly across datasets as the input features such as SHOT descriptors Tombari et al. [2010] are sensitive to the triangle mesh structure, which varies drastically across different datasets.

Most recently, Groueix et al. [2018], Donati et al. [2020] have shown that feature functions can be learned directly from the *raw 3D data* without relying on pre-defined descriptors, resulting in a significantly more robust and accurate method. However, to obtain good results this work had to rely on ground truth correspondences and does not generalize its empirical success beyond its own setup. Although PointNet Qi et al. [2017a] and its variants (Qi et al. [2017b]) achieve impressive results from raw point clouds for classification tasks, they are not yet competitive for shape correspondence task.

Partial Shape Matching While some formulations of functional maps allow to deal with the lack of isometry and partiality, this framework is in principle not designed to deal with partial correspondence. Rodolà et al. [2017] provided an empirical evidence and theoretical analysis of a surprising property of interaction between Laplacian eigenfunctions as a result of removing parts from surfaces. This implies that there exists an unknown alignment between eigenfunctions of partial shapes and full shapes and knowing it results in a special slanted diagonal structure of the correspondence matrix. However, it requires a complicated alternating optimization over the spectral domain and the spatial domain. Instead, Litany et al. [2017b] proposed an efficient and fully spectral domain method for finding this transformation matrix between the two eigen space. However, it still relies on hand crafted features, optimization on Stiefel manifold and is instance specific. Besides, replacing handcrafted features by learnable feature descriptors is not straightforward due to manifold optimization involved in the process. We address both these issue by proposing a novel method that mitigates these issues by learning directly from raw data.

3 Background

As mentioned above, in this work we focus on the functional map representation, due to its efficiency, rich geometric structure and improved ability to generalize beyond a small training set, as was recently demonstrated in Donati et al. [2020].

Basic Deep Functional Map Pipeline Given a source and a target shape, S_1, S_2 , containing, respectively, n_1 and n_2 vertices, the basic pipeline for computing a map between them using the functional map framework Ovsjanikov et al. [2012] and its deep counterparts is as follows:

1. On each shape, compute a small set of k_1, k_2 of basis functions , e.g. by taking the first few eigenfunctions of the respective Laplace-Beltrami operators.
2. Compute a set of descriptor (also known as “probe”) *functions* on each shape that should be preserved by the unknown map. In our case, these functions are PointNet features that are further projected onto the first 30 Laplacian eigen functions and their coefficients are stored as columns of matrices \mathbf{A}, \mathbf{B} .
3. Compute the optimal *functional map* \mathbf{C} by solving the following optimization problem:

$$C_{\text{opt}} = \arg \min_{\mathbf{C}} E_{\text{desc}}(\mathbf{C}) + \alpha E_{\text{reg}}(\mathbf{C}), \quad (1)$$

where $E_{\text{desc}}(\mathbf{C}) = \|\mathbf{C}\mathbf{A} - \mathbf{B}\|^2$ aims at the descriptor preservation whereas the second term acts as a regularizer on the map by enforcing its overall structural properties. Eq 1 can be solved with any convex solver. However, when descriptor functions are neural network based, we need to differentiate the solution with respect to the spectral features \mathbf{A}, \mathbf{B} , which is challenging when \mathbf{C} is computed via an iterative solver. Deep functional maps avoid this by only optimizing \mathbf{C} the first part of the energy : $\|\mathbf{C}\mathbf{A} - \mathbf{B}\|^2$ by solving a simple linear system for which the derivatives can be computed in closed form.

4. The functional map \mathbf{C} in spectral domain is then converted to a point-to-point spatial map using one of several techniques e.g nearest neighbor search in the spectral embedding.

4 Method

4.1 Overview

In this section, we first introduce our approach to learning descriptors from raw 3D shapes for full to full shape matching. Afterwards, we detail our novel partial shape matching algorithm that learns an optimal alignment of laplacian eigen basis functions given the spectrum of partial and full shape. Note that the feature descriptor extraction is common to both approaches. However, our unsupervised loss function is totally different for partial and full shape matching.

4.2 Supervision

In both full and partial matching cases, our method is “weakly supervised” in the sense that we expect the input non-rigid shapes to be approximately *rigidly* aligned. This means having a consistent

‘up’ direction (along, e.g., the y axis) and an approximate forward-facing direction (along, e.g., the z direction). Many existing datasets, such as FAUST Bogo et al. [2014], partial SHRECCosmo et al. [2016], already satisfy this assumption. When considering multiple datasets we only need to make sure that these axes are consistent, which can be done with very little manual intervention. We stress that we *do not use* ground truth point-to-point or functional correspondences, and that obtaining reliable detailed ground truth maps requires significant effort especially when considering cross-dataset learning. Please see the supplementary for example shapes and their approximate alignment in our training and test data.

4.3 Feature extractor

Our main goal is to learn functional characterizations of point clouds that will later be used to compute spectral descriptors and then functional maps. Thus, this network must be applied with the same weights to the source and target shapes in a siamese way using a shared learnable parameters. Our feature extractor is based on Pointnet ++ Qi et al. [2017b] that extracts local features capturing fine geometric structures from small neighborhoods. Such local features are further grouped into larger units and processed to produce higher level features. Our feature extraction network is based on the standard architecture consisting of 4 sampling layers, with first layer sampling 1024 points and 4 feature propagation layers such that final layer outputs 128 dimension feature descriptor for each input shape. For details on all the parameters of network, please see supplement.

4.4 Unsupervised loss for full shape matching

Given the extracted feature functions, we first project them onto the Laplacian basis and then compute the optimal functional map by minimizing $\min_C \|\mathbf{C}\mathbf{A} - \mathbf{B}\|^2$. As noted in Litany et al. [2017a], this leads to a simple linear system of equation, whose solution can be differentiated during training.

We therefore train feature extraction network by imposing an unsupervised loss on the optimized functional map. Our loss follows the approach of Roufousse et al. [2019] and is based on three key structural properties of a functional map between two full isometric shapes.

Bijection Transporting functions on a shape and transporting them back should yield the same functions. Following Eynard et al. [2016], Roufousse et al. [2019], we therefore enforce that composition between \mathbf{C}_{12} and \mathbf{C}_{21} turn out to be as closely as possible to \mathbf{I} , the identity matrix: $E_1 = \|\mathbf{C}_{12}\mathbf{C}_{21} - \mathbf{I}\|^2 + \|\mathbf{C}_{21}\mathbf{C}_{12} - \mathbf{I}\|^2$

Orthogonality As observed in the functional map literature Ovsjanikov et al. [2012], Rustamov et al. [2013], Roufousse et al. [2019] a point-to-point map is locally area preserving if and only if the corresponding functional map is *orthonormal*. Thus, for shape pairs, approximately satisfying this assumption, a natural penalty in our unsupervised pipeline is: $E_2 = \|\mathbf{C}_{12}^T\mathbf{C}_{12} - \mathbf{I}\|^2 + \|\mathbf{C}_{21}^T\mathbf{C}_{21} - \mathbf{I}\|^2$

Laplacian commutativity Having functional maps that commute with the Laplace-Beltrami operators is known to be a common regularizer in the functional map pipeline Rosenberg [1997], Ovsjanikov et al. [2012]. We recall that this constraint helps find better mappings since certain operators are preserved under isometries: $E_3 = \|\mathbf{C}_{12}\mathbf{\Lambda}_1 - \mathbf{\Lambda}_2\mathbf{C}_{12}\|^2 + \|\mathbf{C}_{21}\mathbf{\Lambda}_2 - \mathbf{\Lambda}_1\mathbf{C}_{21}\|^2$ where $\mathbf{\Lambda}_1$ and $\mathbf{\Lambda}_2$ are diagonal matrices of the Laplace-Beltrami eigenvalues on the two shapes.

Thus, our unsupervised loss function is a combination of all three structural properties and weighted as follows: $L = E_1 + E_2 + .001 * E_3$ where the weighing scalars are found empirically.

4.5 Basis Alignment for Partial Shape Matching

The basic pipeline described above for shape matching breaks down in the case of partial shape matching. This is primarily because structural properties of map such as bijection, area preservation (orthogonality) are not applicable anymore. Rodolà et al. [2017] showed that for each ‘‘partial’’ eigenfunction (i.e., each eigenfunction of a partial shape), there exists a corresponding ‘‘full’’ eigenfunction (i.e., some eigenfunction of the full shape). The problem then reduces to finding alignment in k dimensional eigen space that is achieved by optimizing for a new basis on one shape only and keeping the other fixed to the standard Laplacian eigenfunctions. Due to this coupling, the new basis functions

will behave consistently resulting in almost perfectly diagonal \mathbf{C} even in the absence of a perfect isometry. Mathematically, it is written as follows:

$$\min_{\mathbf{X}} \|\mathbf{A}_r - \mathbf{X}^\top \mathbf{B}\|^2 + \text{off}(\mathbf{X}^\top \mathbf{\Lambda} \mathbf{X}), \quad (2)$$

where \mathbf{A}_r contains the $r \times k$ submatrix of \mathbf{A}_r (the first r rows of matrix \mathbf{A}_r) and \mathbf{X} of size $k \times r$ is a transformation matrix between the two eigen spaces that stores the coefficients of desired linear combination. $\mathbf{\Lambda}$ is a diagonal matrix of the first k eigenvalues of partial shape. The second term in Eq. (2) is a regularizer on \mathbf{X} that ensures that resulting eigen basis functions on partial shape minimize the Dirchelet energy on its Laplacian Beltrami operator Δ . The value of r is estimated from the spectrum of partial and full shape as follows: $r = \max_{i=1}^{k_p} \{i \mid \lambda_i^p < \max_{j=1}^{k_f} \lambda_j^f\}$ after setting $k_p = k_f = 60$ where f denotes the full shape and p denotes the partial one. We upper bound the rank obtained by 40.

The method of Litany et al. [2017b] obtains the descriptor function matrix \mathbf{A} and \mathbf{B} using precomputed SHOT descriptors. Besides, it constrains \mathbf{X} to be an orthogonal matrix and thus optimize it using manifold optimization solver on Stiefel Manifold. However, we do not impose any orthogonality constraint on \mathbf{X} and optimize Eq. (2) differently since our descriptor functions are PointNet ++ based and need to be learned simultaneously. So, instead of optimizing over \mathbf{X} , we are optimizing the functional over \mathbf{X} , \mathbf{A} and \mathbf{B} .

$$\min_{\mathbf{X}, \mathbf{A}, \mathbf{B}} \|\mathbf{A}_r - \mathbf{X}^\top \mathbf{B}\|^2 + \text{off}(\mathbf{X}^\top \mathbf{\Lambda} \mathbf{X}), \quad (3)$$

We split the functional in Eq. 3 in two parts and first optimize for \mathbf{X} by solving $\|\mathbf{A}_r - \mathbf{X}^\top \mathbf{B}\|^2$ with a simple linear system for which the derivatives can be computed in closed form. Given this optimal \mathbf{X} , we then impose the loss on \mathbf{X} by computing the second part of Eq (3) and use this unsupervised loss to backpropagate gradients to learn the appropriate descriptor functions. Note that for partial matching this loss term is *the only* one we use, whereas in the full shape matching setting we use a more powerful loss described in Section 4.4.

Implementation We implemented our method in TensorFlowAbadi et al. [2015]. We train our network with a batch size of 8 shape pairs for 10000 steps. We use a learning rate of $1e-4$ with Adam optimizer. During training, we randomly sample 4000 points from each shape while training with Surreal dataset whose shapes contain 7000 points each. For other datasets such as Scape and Faust remesh, that contains roughly 5000 points each, we randomly sample 3000 points during for training. Since partial shape dataset contains very limited number of shapes, we describe its experimental setup later in Section 5.3. For a faircomparison with some baseline methods, we use a very recent and efficient refining algorithm, called ZoomOut Melzi et al. [2019] based on navigating between spatial and spectral domains while progressively increasing the number of spectral basis functions.

5 Results

This section is divided into three subsections where each provides a separate evaluation of our contributions. Section 5.1 shows the experimental comparison of our weakly supervised approach with fully supervised state-of-the art methods for near-isometric shape matching. Section 5.2 demonstrates that weak rigid alignment of datasets, low number of Laplacian eigen basis and enforcing structural properties of a map suffice to obtain excellent results across a variety of loss functions. Finally, Section 5.3 demonstrates the effectiveness of our novel partial shape matching framework. We evaluate all results by reporting the per-point-average geodesic distance between the ground truth map and the computed map. All results are multiplied by 100 for the sake of readability.

5.1 Near-isometric Shape Matching

In this section we evaluate our method for complete (full to full) near isometric shape matching. We compare our method with state-of-the-art approaches while focusing especially on the the very recent functional map-based technique Donati et al. [2020], which was shown to outperform existing competitors.

Table 1: Comparative results ($\times 100$) of the different methods on FAUST and SCAPE remesh datasets.

Method \ Dataset	F	S	F on S	S on F
SURFMNet	15.	12.	32.	32.
SURFMNet+icp	7.4	6.1	19.	23.
Unsup FMNet	10.	16.	29.	22.
Unsup FMNet+pmf	5.7	10.	12.	9.3
FMNet	11.	17.	30.	33.
FMNet+pmf	5.9	6.3	11.	14.
3D-CODED	2.5	31.	31.	33.
GeomFmap	3.1	4.4	11.	6.0
GeomFmap +zo	1.9	3.0	9.2	4.3
Ours	3.6	6.2	11	6.2
Ours + zo	1.9	2.6	8.0	4.3

Table 2: Comparative results ($\times 100$) of the different methods when trained on the SURREAL dataset and tested on remeshed FAUST and SCAPE shapes.

Method \ Dataset	F	S
GeomFmap +Zo	2.5	9.2
3D-CODED	4.9	6.0
Ours	5.2	7.5
Ours+zo	2.5	4.5

Datasets For a fair comparison with Donati et al. [2020], we follow the same experimental setup and test our method on a wide spectrum of datasets: first, the re-meshed versions of FAUST dataset Bogio et al. [2014] and the SCAPE Anguelov et al. [2005], made publicly available by Ren et al. Ren et al. [2018]. Lastly, we also use the training dataset of 3D-CODED, consisting in 230K synthetic shapes generated using SURREAL Varol et al. [2017] with the parametric model SMPL introduced in Loper et al. [2015]. We use a subset of it for training purposes to compare the generalization ability of different methods to changes in connectivity and triangulation. This is achieved by training on this synthetic data and testing on re-meshed datasets such as FAUST and SCAPE.

Baselines We compare our method to several state of the art methods: the first category includes a variety of unsupervised deep functional maps proposed recently with SHOT descriptors. Second category includes supervised methods that directly learn from 3D data. This includes the supervised template based approach of 3D-CODED Groueix et al. [2018] as well as the recent work GeomFmap Donati et al. [2020]. All baseline results are taken from Donati et al. [2020]. In the case of SHOT based deep functional maps Litany et al. [2017a], Halimi et al. [2019], Roufousse et al. [2019], all results are invariant by any rigid transformation of the input shapes and therefore, no alignment is required. For a fair comparison with other methods, we show our results with and without ZoomOut Melzi et al. [2019] refinement, referred to as ZO. For conciseness, we refer to our method as Ours in the following text. We compare these different methods in Table 1.

Generalization Experiments Following the standard protocol, we split FAUST re-meshed and SCAPE re-meshed into training and test sets containing 80 and 20 shapes for FAUST, and 51 and 20 shapes for SCAPE. **F** and **S** in Table 1 shows the results for training and testing on same dataset, FAUST and SCAPE, respectively whereas **F on S** means trained on FAUST and tested on SCAPE. In Table 2, results are shown with the SURREAL dataset from which we sample 100 shapes for training and test the trained models on test sets of FAUST re-meshed, SCAPE re-meshed. We compare with 3D-CODED and GeomFMap since they outperform every other method and learn from raw 3D geometry. We report baseline numbers from Donati et al. [2020] which report performance of different methods by varying the size of training set from few hundred to thousands. We pick the best results obtained with any number of shapes. All results are multiplied by 100 for the sake of readability. We also report results without ZoomOut refinement.

Table 3: Comparative results of the different loss functions when trained with our framework on SURREAL dataset and tested on FAUST and SCAPE remesh.

Method \ Dataset	F	S
GeomFap+zo	2.5	9.2
Unsup FMNet loss + Ours	6.3	7.7
Unsup FMNet loss + Ours +zo	4.4	5.2
GeomFap loss + Ours	5.0	7.7
GeomFap loss + Ours +zo	2.7	4.6
Ours	5.2	7.5
Ours +zo	2.5	4.5

Results and Discussion As evident in Table 1, our method performs on par with the fully supervised approaches such as 3D-CODED Groueix et al. [2018] and GeomFMap Donati et al. [2020]. We observe substantial gains over supervised approach in Table 2. We obtain remarkable performance on the SCAPE dataset which is shown to be very challenging at test when trained with any other dataset. On FAUST, we are comparable with GeomFMap even though it is trained with ground truth correspondences.

We would like to stress that baselines such as 3D-CODED and GeomFMap require hundreds of SURREAL shapes, 2000 for 3D-CODED, in order to obtain reasonable results on SCAPE whereas we obtain high quality results with 200 percent improvement over GeomFMap with as low as 100 and 50 shapes. Please see supplement on the effect of varying the training size. We stress that no other method is able to achieve such a generalization with this low number of shapes. We attribute our superior results over GeomFmap to a range of factors. First, in contrast to our unsupervised loss, GeomFmap uses a supervised loss without adequate regularization that leads to severe overfitting on challenging datasets with different poses such as SCAPE. This underscores the importance of enforcing structural properties of functional map in any loss function. Third, GeomFmap achieves a consistent forward facing direction with data augmentation and ground truth functional map supervision, whereas we align it manually. These experimental results confirm our findings that the devil in non-rigid shape matching lies in approximate *rigid* alignment and such weak supervision is equivalent to having supervised ground truth correspondence as we show next.

5.2 Deep Functional Maps with any Loss Function

The goal of this section is to unpack the minimum and sufficient ingredients of a deep functional map pipeline such that it leads to unification of all the recent work under these minimum and sufficient conditions. To this end, we test one representative deep functional map each from supervised setting and unsupervised setting with different loss functions. We optimize their loss functions with our PointNet++ feature extractor with low eigen basis (30) and our regularizers in both functional map pipeline and discard any other regularizer or feature extractor as proposed in these works. We train on SURREAL dataset from which we sample 100 shapes for training and test the trained models on test sets of FAUST re-meshed, SCAPE re-meshed.

Unsup FMNet loss + Ours Halimi et al. [2019] is an unsupervised approach that uses a soft correspondence based loss with geodesic matrix. Note that in their paper, Unsup FMNet relies on the SHOT descriptor that we replace with a PointNet++ feature extractor. We use their unsupervised loss in addition to our regularization terms.

GeomFmap loss + Ours We also evaluate Donati et al. [2020], a supervised approach where the ground truth functional map is computed in the spectral domain. We use this supervised loss function but discard their regularization proposed to alleviate overfitting. We also discard their feature extractor and do not perform any data augmentation. We sample 6000 vertices randomly for each shape as more vertices should lead to a better ground truth functional map estimation.

GeomFmap simply reports the performance of Donati et al. [2020] without any modifications.

Results and Discussion We summarize the findings in Table 3. Remarkably, we obtain state of the art results with both loss functions. In particular, GeomFmap supervised spectral loss when

Table 4: Comparative results on partial Shrec benchmark

Method \ Dataset	Holes
Litany et. al	16
Ours	13

optimized with our framework leads to 200 percent increase in accuracy on the challenging SCAPE dataset. This shows the generalization capability of our framework since GeomFmap already contains a feature extractor and a regularization layer especially designed to reduce overfitting which evidently are not sufficient. Similar performance boost is observed with Unsup FMNet on both datasets. It must be noted that memory footprint/training time of Halimi et al. [2019] is 50 times more as it requires either geodesic matrices to fit to RAM or load them on the fly for each pair. Overall, these results demonstrate the usage of low laplacian eigen basis, regularization by enforcing basic structural properties of map under weak supervision suffice to obtain state-of-the-art results with different loss functions. In the supplement, we show empirically that all three conditions are necessary and exclusion of anyone results in inferior performance.

5.3 Partial Shape Matching

Finally, we quantitatively evaluate our method in the partial matching scenario on the challenging SHREC’16 Partial Correspondence benchmarkCosmo et al. [2016]. The dataset is composed of 200 partial shapes (from a few hundred to 9K vertices each) belonging to 8 different classes (humans and animals), undergoing nearly-isometric deformations in addition to having missing parts of various forms and sizes. Each class comes with a “null” shape in a standard pose which is used as the full template to which partial shapes are to be matched. The dataset is split into two subsets, namely cuts (removal of a few large parts) and holes (removal of many small parts).

Experimental Setup The dataset contains several shapes whose number of points range from few hundreds to 2000. We use some of these shapes as a validation set and separate them from training or test set. Since the number of overall shapes are not enough in each subset for a learning based method to train and test separately, we first split each subset randomly into train and test and then combine the training set of both holes and cuts subset to create our training data. We show results on holes dataset for test as it is shown to be more challenging than cuts in Litany et al. [2017b]. We provide results on cuts in supplement. Our loss function for partial shape matching does not contain any hyperparameters. Thus, we use validation set to only validate the training iterations. We consider Litany et al. [2017b] as our main baseline as it is considered state of the art for partial shape matching. Remark that no existing functional maps learning-based approach has yet been proposed for *partial* non-rigid shape matching.

Results and Discussion We present our findings on partial shape matching in Table 4. Note that the method of Litany et al. [2017b] is not learning based but relies on expensive manifold optimization for every pair of shapes at test time. In contrast our method obtains a correspondence directly with pre-trained features and without the need for any test time optimization.

6 Conclusion and Future Work

We presented a novel weakly supervised method based on the functional map representation for both full and partial shape matching. Our main observation is that weak supervision in the form of approximately rigidly aligned input data is sufficient for learning powerful features to solve the non-rigid correspondence problem from raw data. Moreover, we establish that the key to cross dataset generalization lies in working with low number of eigen basis and enforcing very basic structural properties of a functional map. Our method for partial shape matching is also the first approach towards learning partial functional map and is of independent interest. We believe that this method will set the future direction of research, especially towards simpler techniques and weak supervision, in both near isometric as well as partial shape matching.

7 Acknowledgement

Parts of this work were supported by the ERC Starting Grant StG-2017-758800 (EXPROTEA), KAUST OSR Award No. CRG-2017-3426, and a gift from Nvidia. We thank Ruqi Huang and Simone Melzi for their help in performing quantitative comparisons on Partial SHREC dataset.

8 Broader Impact

Shape matching is a fundamental problem that arises in many areas of science and engineering from statistical shape analysis, creation of virtual avatars, medical imaging (for instance for detecting anomalies, and performing follow-up analysis) as well as other areas such as archeology and palaeontology (for comparing artefacts and anatomical features) among others. Efficient algorithms for solving the shape correspondence problem, therefore have immediate impact in those areas facilitating tedious manual intervention and requirements for expert knowledge, which are often not available. Our approach which achieves state-of-the-art results on challenging benchmarks can immediately be adapted and tested in such diverse scenarios. This is particularly true as our method does not place any assumptions on the shape category (e.g., not being restricted to only human shapes as some previous approaches), is efficient and fully automatic. We believe that the observations made in our work can also lead to new insights in other areas including graph matching and machine translation, where data is often represented as point clouds in some embedding space.

References

- Martín Abadi, Ashish Agarwal, and Paul Barham et al. TensorFlow: Large-scale machine learning on heterogeneous systems, 2015. URL <https://www.tensorflow.org/>. Software available from tensorflow.org.
- Dragomir Anguelov, Praveen Srinivasan, Daphne Koller, Sebastian Thrun, Jim Rodgers, and James Davis. SCAPE: Shape Completion and Animation of People. In *ACM Transactions on Graphics (TOG)*, volume 24, pages 408–416. ACM, 2005.
- 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.
- Davide Boscaini, Jonathan Masci, Emanuele Rodola, and Michael M. Bronstein. Learning shape correspondence with anisotropic convolutional neural networks. In *Proc. NIPS*, pages 3189–3197, 2016.
- Oliver Burghard, Alexander Dieckmann, and Reinhard Klein. Embedding shapes with Green’s functions for global shape matching. *Computers & Graphics*, 68:1–10, 2017.
- P Chavdar, H Sami, P Sven, Krieger Kai, and B. Darius. Rigid 3d geometry matching for grasping of known objects in cluttered scenes. *IJRR*, 31(4), 2012.
- Etienne Corman, Maks Ovsjanikov, and Antonin Chambolle. Supervised descriptor learning for non-rigid shape matching. In *Proc. ECCV Workshops (NORDIA)*, 2014.
- Luca Cosmo, Emanuele Rodola, Jonathan Masci, Andrea Torsello, and Michael M Bronstein. Matching deformable objects in clutter. In *3D Vision (3DV), 2016 Fourth International Conference on*, pages 1–10. IEEE, 2016.
- Nicolas Donati, Abhishek Sharma, and Maks Ovsjanikov. Deep geometric functional maps: Robust feature learning for shape correspondence. In *CVPR*, 2020.
- Davide Eynard, Emanuele Rodola, Klaus Glashoff, and Michael M Bronstein. Coupled functional maps. In *3D Vision (3DV)*, pages 399–407. IEEE, 2016.
- 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.

- Thibault Groueix, Matthew Fisher, Vladimir G Kim, Bryan C Russell, and Mathieu Aubry. Unsupervised cycle-consistent deformation for shape matching. In *Computer Graphics Forum*, volume 38, pages 123–133. Wiley Online Library, 2019.
- Oshri Halimi, Or Litany, Emanuele Rodolà, Alex Bronstein, and Ron Kimmel. Unsupervised learning of dense shape correspondence. In *CVPR*, 2019.
- Qixing Huang, Fan Wang, and Leonidas Guibas. Functional map networks for analyzing and exploring large shape collections. *ACM Transactions on Graphics (TOG)*, 33(4):36, 2014.
- 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, 2013.
- Or Litany, Tal Remez, Emanuele Rodolà, Alexander M. Bronstein, and Michael M. Bronstein. Deep functional maps: Structured prediction for dense shape correspondence. *2017 IEEE International Conference on Computer Vision (ICCV)*, pages 5660–5668, 2017a.
- Or Litany, Emanuele Rodolà, Alex M Bronstein, and Michael M Bronstein. Fully spectral partial shape matching. In *Computer Graphics Forum*, volume 36, pages 247–258. Wiley Online Library, 2017b.
- Matthew Loper, Naureen Mahmood, Javier Romero, Gerard Pons-Moll, and Michael J. Black. Smpl: A skinned multi-person linear model. *ACM Trans. Graph.*, 34(6):248:1–248:16, October 2015. ISSN 0730-0301. doi: 10.1145/2816795.2818013. URL <http://doi.acm.org/10.1145/2816795.2818013>.
- Jonathan Masci, Davide Boscaini, Michael Bronstein, and Pierre Vandergheynst. Geodesic convolutional neural networks on riemannian manifolds. In *Proceedings of the IEEE international conference on computer vision workshops*, pages 37–45, 2015.
- Simone Melzi, Jing Ren, Emanuele Rodola, Maks Ovsjanikov, and Peter Wonka. Zoomout: Spectral upsampling for efficient shape correspondence. *ACM Transactions on Graphics (Proc. SIGGRAPH Asia)*, 2019.
- Federico Monti, Davide Boscaini, Jonathan Masci, Emanuele Rodolà, Jan Svoboda, and Michael M. Bronstein. Geometric deep learning on graphs and manifolds using mixture model cnns. In *CVPR*, pages 5425–5434. IEEE Computer Society, 2017.
- Dorian Nogneng and Maks Ovsjanikov. Informative descriptor preservation via commutativity for shape matching. *Computer Graphics Forum*, 36(2):259–267, 2017.
- Maks Ovsjanikov, Mirela Ben-Chen, Justin Solomon, Adrian Butscher, and Leonidas Guibas. Functional Maps: A Flexible Representation of Maps Between Shapes. *ACM Transactions on Graphics (TOG)*, 31(4):30, 2012.
- 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 *ACM SIGGRAPH 2017 Courses, SIGGRAPH '17*, pages 5:1–5:62, 2017.
- Charles R Qi, Hao Su, Kaichun Mo, and Leonidas J Guibas. Pointnet: Deep learning on point sets for 3d classification and segmentation. In *Proc. CVPR*, pages 652–660, 2017a.
- Charles R. Qi, Li Yi, Hao Su, and Leonidas J. Guibas. Pointnet++: Deep hierarchical feature learning on point sets in a metric space. In *Proceedings of the 31st International Conference on Neural Information Processing Systems, NIPS'17*, pages 5105–5114, USA, 2017b. Curran Associates Inc. ISBN 978-1-5108-6096-4. URL <http://dl.acm.org/citation.cfm?id=3295222.3295263>.
- Jing Ren, Adrien Poulencard, Peter Wonka, and Maks Ovsjanikov. Continuous and orientation-preserving correspondences via functional maps. *ACM Transactions on Graphics (TOG)*, 37(6), 2018.
- 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.

- Steven Rosenberg. *The Laplacian on a Riemannian manifold: an introduction to analysis on manifolds*, volume 31. Cambridge University Press, 1997.
- Jean-Michel Roufosse, Abhishek Sharma, and Maks Ovsjanikov. Unsupervised deep learning for structured shape matching. In *ICCV*, pages 1617–1627, 2019.
- Raif Rustamov, Maks Ovsjanikov, Omri Azencot, Mirela Ben-Chen, Frederic Chazal, and Leonid Guibas. Map-based exploration of intrinsic shape differences and variability. *ACM Trans. Graphics*, 32(4):72:1–72:12, July 2013. ISSN 0730-0301.
- Yusuf Sahillioğlu. Recent advances in shape correspondence. *The Visual Computer*, pages 1–17, 2019.
- Abhishek Sharma, Oliver Grau, and Mario Fritz. Vconv-dae: Deep volumetric shape learning without object labels. In *ECCV*, 2016.
- Federico Tombari, Samuele Salti, and Luigi Di Stefano. Unique signatures of histograms for local surface description. In *International Conference on Computer Vision (ICCV)*, pages 356–369, 2010.
- 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, 2011.
- Gül Varol, Javier Romero, Xavier Martin, Naureen Mahmood, Michael J. Black, Ivan Laptev, and Cordelia Schmid. Learning from synthetic humans. In *CVPR*, 2017.
- Lingyu Wei, Qixing Huang, Duygu Ceylan, Etienne Vouga, and Hao Li. Dense human body correspondences using convolutional networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 1544–1553, 2016.