

DPFM: Deep Partial Functional Maps

Souhaib Attaiki Gautam Pai Maks Ovsjanikov
LIX, École Polytechnique, IP Paris

Abstract

We consider the problem of computing dense correspondences between non-rigid shapes with potentially significant partiality. Existing formulations tackle this problem through heavy manifold optimization in the spectral domain, given hand-crafted shape descriptors. In this paper, we propose the first learning method aimed directly at partial non-rigid shape correspondence. Our approach uses the functional map framework, can be trained in a supervised or unsupervised manner, and learns descriptors directly from the data, thus both improving robustness and accuracy in challenging cases. Furthermore, unlike existing techniques, our method is also applicable to partial-to-partial non-rigid matching, in which the common regions on both shapes are unknown a priori. We demonstrate that the resulting method is data-efficient, and achieves state-of-the-art results on several benchmark datasets. Our code and data can be found online: <https://github.com/pvnieo/DPFM>.

1. Introduction

Non-rigid shape correspondence is an essential problem in 3D Computer Vision, which enables a range of downstream tasks, such as statistical shape analysis [35] and deformation transfer [48] among many others.

A particularly challenging setting for this problem is computing correspondences between shapes in 3D that undergo both non-rigid deformations and exhibit strong partiality, such as missing parts or having holes, due, e.g., to acquisition errors. Several works have aimed to address this problem, in particular by adapting spectral techniques, within the functional map framework [41, 26, 1, 57, 58]. Spectral methods are inherently invariant under near-isometric deformations, making them attractive for non-rigid shape matching problems. Furthermore, objective functions promoting structural properties of functional maps have been designed through careful analysis of the alignment of Laplacian basis in the partial setting, e.g., [41, 26]. As a result, these approaches lead to moderately-sized optimization problems and can achieve accurate results even under significant pose changes.

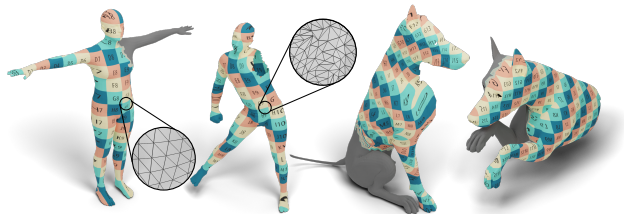


Figure 1: Our method is highly effective for non-rigid partial shape matching, even in the challenging cases of significant mesh and sampling variability (left), and partial-to-partial matching (right). Our method detects both the overlapping regions between the shapes (non-common regions are colored in grey) and produces dense pointwise correspondences, visualized via texture transfer.

At the same time, most existing methods strongly rely on hand-crafted input features, and especially the popular local SHOT descriptors [45]. Unfortunately, these descriptors are known to be very sensitive to local mesh structure and can fail when comparing two shapes that have different sampling or connectivity [14]. To remedy this issue, recent approaches have aimed at *learning* features for shape correspondence in both supervised and unsupervised settings, e.g., [28, 25, 18].

Despite this progress, however, existing learning-based methods are only geared towards computing correspondences between *complete* non-rigid shapes, and can easily fail in the presence of strong partiality. This is because special care must be taken to account for partiality during both training and testing, as most losses and feature extraction networks are typically not applicable.

In this paper, we present the first learning-based approach specifically geared towards finding correspondences between non-rigid shapes under strong partiality. For this, we design a novel architecture that can accommodate both partial-to-full and partial-to-partial correspondence. Our method can be trained in a supervised and unsupervised manner. We use the functional map framework but learn features from data, making our method robust both to change of poses and to changes in mesh sampling and connectivity (see Figure 1).

Our contributions can be summarized as follows:

- We propose a new deep learning method for non-rigid partial shape correspondence, which is robust under mesh changes, predicts both the correspondences *and the region of overlap*, and is the first learning-based approach for partial-to-partial non-rigid matching.
- We provide theoretical analysis of the limitations of a Siamese architecture for partial functional maps, and develop a module that enables communication between features.
- We study the robustness of existing methods and introduce two new datasets, both for partial-to-partial and partial-to-full matching with significant mesh variability. Our method achieves state-of-the-art on both standard and these new benchmarks.

2. Related Work

Shape matching is a vast and well-studied area of Computer Vision. Below we review works most closely related to ours, focusing especially on non-rigid partial shape correspondence, and refer the interested readers to recent surveys [51, 3, 44] for a more in-depth treatment of this field.

Partial rigid correspondence Partial matching has been particularly well-studied in the context of *rigid* shape alignment, motivated especially by shape reconstruction problems from partial scans [8, 2]. While early methods aimed to address this problem primarily using the classical ICP and its robust variants, e.g., [43, 9, 15, 5], more recent approaches focus on learning robust features that can be used for rigid alignment, [61, 13, 10, 16]. We note that several recent approaches in the latter category aim to directly address partiality, typically through outlier filtering, keypoint detection, or even explicit overlap prediction [54, 16, 34, 19]. While these methods achieve impressive performance even in the presence of extreme partiality [19], they are all limited to rigid alignment, which has a small number of degrees of freedom – indeed, only three point pair correspondences are sufficient to recover a rigid transformation.

Learning for full-to-full non-rigid correspondence

Learning strategies have also been applied to address non-rigid shape correspondence problems, especially pioneered by techniques in geometric deep learning [6, 7]. Most approaches in this field treat shape correspondence as a vertex labeling problem [31, 4, 28, 36, 55], thus mapping each shape to some predefined canonical template. In contrast, methods based on the functional map representation have recently allowed to predict and penalize *entire maps* between shape pairs [25, 18, 42, 14] achieving better accuracy and generalization power. While approaches in the former category have been applied in the context of partial-to-full

matching in, e.g., [4] and [24], they are unfortunately limited to matching shapes to a predefined template and furthermore have been observed to not be robust under sampling and connectivity changes [47]. On the other hand, existing learning functional maps-based techniques treat the correspondence problem globally but are geared towards complete shape matching. Some techniques can generalize to unseen partial matching [46] but do not outperform axiomatic methods, in part, as they are not explicitly trained to deal with partiality.

Spectral methods for partial non-rigid matching In the realm of axiomatic spectral techniques and optimization methods, the partial functional map framework (PFM) [41] is one of the important methods to tackle partial shape correspondence. The PFM scheme which is principally designed to tackle a part-to-full correspondence problem alternates between estimating a functional map and a part-region indicator function on the full shape, where both terms are strongly regularized. Follow-up works like [27, 12] extended the PFM approach to both cluttered non-rigid correspondence and multi-part settings where potentially multiple functional maps are optimized simultaneously.

Litany and colleagues [26] proposed a fully spectral approach for partial shape correspondence which avoids region-based optimization. Similar in spirit to joint diagonalization schemes like [23], this method employs a manifold optimization scheme to estimate a linear transformation to align the spectral basis between the shapes and is also posed to handle part-to-part shape correspondence.

More recently, iterative spectral methods [30] have been shown to effectively handle the partial setting, where the slanted structure of the functional map is first estimated using the same spectral relation as proposed in PFM and then used in an upsampling strategy to refine an initial correspondence. Similarly, [58] and [57] use refinement via spectral upsampling by combining an array of objectives promoting structural properties of the computed functional maps, while [1] target partial shape matching using properties of the Nearest-Neighbor Field (NNF). Finally, [38] use a correspondence-free method for localizing the region on a full shape in the context of partial-to-full shape comparison.

We highlight that all of the existing axiomatic techniques strongly rely on input descriptors either for initialization or even for map optimization. Practical implementations typically employ SHOT descriptors within partial non-rigid matching pipelines. However, such handcrafted descriptors are known to be sensitive to the discretization of the surface, which severely affects the robustness of all resulting methods. Instead, we put special emphasis on the robustness of our approach and show that it can produce accurate results even under significant mesh changes.

Partial-to-partial Most existing methods assume that there exists a corresponding point on the target (full) shape

for every point on the source (partial) shape. To address the more general partial-to-partial matching, FSP [26] proposed a formulation for this setting, but without providing any practical evaluation.

Similar in spirit to [41], [12] demonstrates a partial-to-partial matching pipeline between a source (model) and a target shape (scene) - that potentially comprises an overlapped part of the model and other non-related cluttered objects. However, this method is highly dependent on an initialization procedure like [40] which ultimately depends on the quality of robust input descriptors, unlike SHOT which is usually chosen by default.

Finally, Litany and colleagues [27] proposed a solution to the non-rigid puzzle problem, where partial-to-full matching was extended to include possible additional clutter. However, their solution is constrained by assumptions on the existence of missing parts and ultimately still relies on a partial-to-full matching component.

In contrast to all of these existing techniques, our method completely avoids the need for hand-crafted input descriptors and allows to learn features directly from the 3D data. Furthermore, our functional map-based learning pipeline is geared towards partiality and can learn from limited training data, due to the strong spectral regularization. Thus, our method is virtually parameter-free and strongly outperforms existing methods, especially in the more realistic setting of shapes with significantly different mesh structures.

3. Motivation and Method Overview

As mentioned above, our main goal is to develop a learning-based method for partial non-rigid shape correspondence, while avoiding the use of hand-crafted features. For this, we follow the general strategy of recent learning-based spectral techniques [25, 18, 46, 14], and especially Deep Geometric Functional Maps introduced in [14]. This approach can be summarized as follows:

1. Given the source and target shapes \mathcal{X}, \mathcal{Y} , a trainable Siamese feature network \mathcal{F}_Θ produces m feature functions $\mathcal{F}_\Theta(\mathcal{X}) = \{f_1, f_2, \dots, f_m\}$ and $\mathcal{F}_\Theta(\mathcal{Y}) = \{g_1, g_2, \dots, g_m\}$.
2. These features are projected onto the Laplace-Beltrami basis to obtain their coefficients $\mathbf{a}_i = \Phi_{\mathcal{X}}^+ f_i$ and $\mathbf{b}_i = \Phi_{\mathcal{Y}}^+ g_i$, where $\Phi_{\mathcal{X}}, \Phi_{\mathcal{Y}}$ are the matrices that store, as columns, the pre-computed set of Laplace-Beltrami eigenfunctions of shapes \mathcal{X} and \mathcal{Y} respectively, while $+$ denotes the Moore-Penrose pseudo-inverse.
3. Finally, $\{\mathbf{a}_i, \mathbf{b}_i\}$ are stacked as columns of matrices \mathbf{A}, \mathbf{B} and fed to a regularized functional map layer, which uses a linear system to compute the functional

map [33] in a reduced basis:

$$C_{opt} = \arg \min_C \|C\mathbf{A} - \mathbf{B}\| + \lambda \|C\Delta_{\mathcal{X}} - \Delta_{\mathcal{Y}}C\|. \quad (1)$$

Here $\Delta_{\mathcal{X}}, \Delta_{\mathcal{Y}}$ are diagonal matrices of Laplace-Beltrami eigenvalues of the corresponding shapes and λ is a scalar hyper-parameter.

4. The computed functional map C_{opt} is compared to the ground truth C_{gt} one by imposing the Frobenius norm loss $\mathcal{L}(C_{opt}) = \|C_{gt} - C_{opt}\|_F^2$, and this loss is used to train the parameters Θ of the feature network \mathcal{F}_Θ in a supervised manner.

As mentioned in that work and several recent follow-ups [14, 46, 47], using the reduced Laplacian basis helps to regularize feature learning and allows to learn even from a limited set of training shapes, while avoiding the use of a predefined template as done in works that treat shape correspondence as a vertex labeling problem, e.g., [4, 31, 55].

Overview Our approach follows the same general approach as in [14], while making several key contributions (See Figure 2 for the overview).

First, we replace the Siamese network \mathcal{F}_Θ with a combination of a feature extractor and a *Cross Attention refinement* block. Our main motivation is that in the context of partial correspondence, the feature extractor on each shape should be made aware of the other shape. In the functional map language, if \mathbf{a} and \mathbf{b} are coefficients of the feature functions on the source and target respectively, and if $\mathbf{a} = 0$ because, e.g., a particular part is missing then $C\mathbf{a} = \mathbf{b}$ will only hold if $\mathbf{b} = 0$. However, a Siamese network is not aware of the descriptors on the source shape and would produce a non-zero feature \mathbf{b} that characterizes the existing part on the target shape. We provide an in-depth theoretical analysis of this effect, and highlight the limitations of a Siamese architecture for partial functional maps in the supplementary.

Motivated by these considerations we introduce the Attention-based Feature refiner (Section 3.2) that takes as input features on both shapes and produces refined features of each shape that are “informed” by the features on the other shape. We remark that since functional maps operate *in the opposite direction* of point-to-point ones, the analysis above applies when computing a pointwise map from either a full or a partial to another partial shape. However, we have found that the “communication” between features is generally useful across a range of matching settings.

Our second major contribution is that in addition to the functional map, we also train an Overlap predictor (Section 3.3) that produces the region on each shape that is expected to exist on the other shape. This helps us to both produce a dense pointwise map and the region in which it is expected to be accurate thus allowing us to deal with full-to-partial and partial-to-partial non-rigid matching.

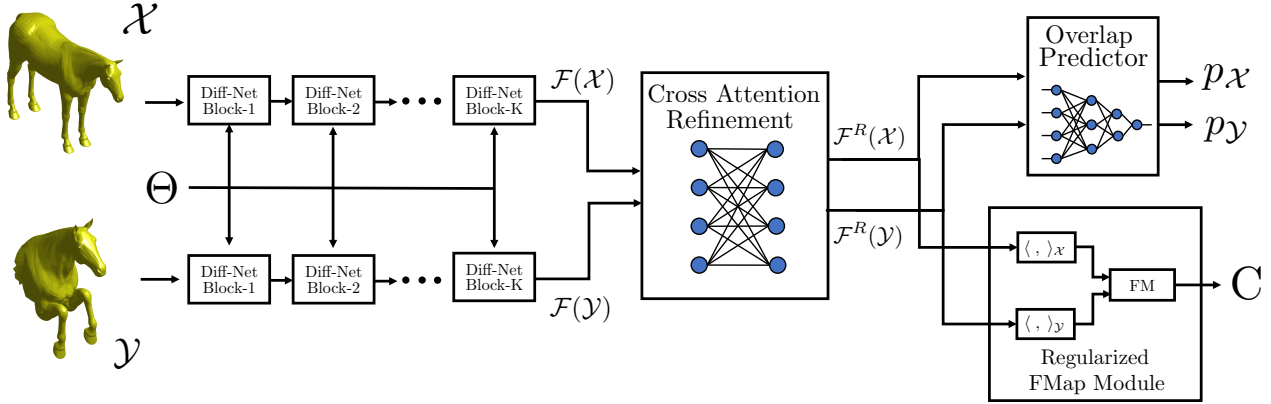


Figure 2: Overview of our approach. Raw input geometry of a pair of shapes is processed by a Siamese network with shared weights resulting in features $\mathcal{F}(\mathcal{X})$ and $\mathcal{F}(\mathcal{Y})$. These are then refined using a Cross Attention Refinement module allowing the features on the two shapes to communicate. The refined features are then used, first, to predict the overlap region using the Overlap Predictor Module, and second, to compute the functional map using the parameter-free Regularized FMap module.

Finally, we make several other architecture modifications that improve its accuracy and robustness: we use more powerful intrinsic and sparse feature extractors (Section 3.1) and a regularized functional map layer adapted to the partial setting (Section 3.4). Below we provide the details of each of these building blocks.

3.1. Feature extractor

As mentioned above, and as shown in Figure 2, the first block in our pipeline is a feature extractor module.

The goal of this module is to take as input the raw shape geometry and produce meaningful descriptors that will be used downstream for spectral matching. This module should be robust to rigid and non-rigid deformations, to the surface sampling, and more importantly, to the challenges introduced by partiality. Throughout our work, we use the recent DiffusionNet architecture [47] that is based on intrinsic operations for information diffusion on the surface, and has been shown to be robust under shape discretization changes. In cases of extreme partiality such as the Holes setting of the SHREC 2016 partiality benchmark [11], we also experimented with SparseConvNet [17], which can better handle very sparse data. This leads to a variant of our approach that we call DPFM_{sparse}. Unless specified otherwise, DPFM always denotes our approach with the DiffusionNet feature extractor.

Following [14], our feature extractor is applied in a Siamese way, i.e the same network with identical weights is applied to the source \mathcal{X} and target \mathcal{Y} shapes. This network produces an m -dimensional vector per point, or equivalently m feature functions on the surface, that we denote $\mathcal{F}(\mathcal{X})$ and $\mathcal{F}(\mathcal{Y})$ respectively. We observe that the surface-aware feature extractors like DiffusionNet and SparseConv typically outperform the point-based KPConv [49] used in [14].

3.2. Attention-based Feature Refiner

The features produced by the feature extractors encode the geometry and context of each shape independently and have no knowledge of the other shape, making predicting the overlap region impossible. We remedy this by letting the features *communicate* using a cross attention block, using a design inspired by the recent Predator network [19] architecture, that was introduced for partial *rigid* alignment.

Specifically, we construct a bipartite graph $(\mathcal{V}, \mathcal{E})$, where every point on shape \mathcal{X} is connected to all points on shape \mathcal{Y} , and we associate to each node the point-wise features learned from the feature extractor. The message passing formulation of [21] is used to propagate information through this graph.

Following the Transformer architecture [52], for each node i , a representation query q_i is learned, and it retrieves learned values v_j , based on their learned attributes, the keys k_j . The final message that is passed to node i has the form:

$$m_{\mathcal{E} \rightarrow i} = \sum_{j, (i,j) \in \mathcal{E}} \alpha_{ij} v_j$$

where α_{ij} are the attention weights, based on the key-query similarity $\alpha_{ij} = \text{Softmax}_j(q_i^T k_j / \sqrt{d})$.

The queries $q_i \in \mathbb{R}^d$, keys $k_i \in \mathbb{R}^d$ and values $v_i \in \mathbb{R}^d$ are learned as a linear projection of learned features of the previous block, using a learnable linear layer.

Finally, the value of the node i is updated using the formula:

$$x_i = x_i + \text{MLP}([x_i || m_{\mathcal{E} \rightarrow i}])$$

where $[.||]$ stands for concatenation, and MLP is a three layer perceptron [56] with ReLU activations [60] and instance normalization [50].

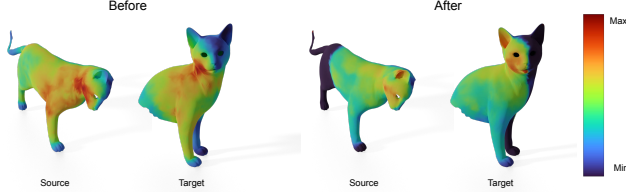


Figure 3: Visualizing the per-point feature magnitude, before and after the Cross-Attention Refinement module.

In practice and in line with existing literature [53], we improve the expressivity of the model using multi-head attention [52] with four heads. Also, in order to reduce the computation and memory footprint of this block, we sub-sample the two shapes using the farthest point sampling FPS [37] before constructing the bipartite graph. The learned features are then linearly interpolated to the missing points.

The feature refiner module takes as input features $\mathcal{F}(\mathcal{X})$, $\mathcal{F}(\mathcal{Y})$ and produces *refined* features that we denote $\mathcal{F}^R(\mathcal{X})$, $\mathcal{F}^R(\mathcal{Y})$.

As mentioned above, our Cross-Attention Refinement module enables the *communication* between features on the two shapes and thus allows the features on the overlap region to synchronize, while down-weighting the features outside the overlap. We illustrate this effect in Figure 3, where we show the intensity of the learned features, before and after the refinement module. Note that before refinement, the features are uniformly distributed, but after the refinement, the features on the missing region tend to zero, as suggested by our theoretical analysis.

Additional analysis and quantitative evaluation of the Cross-Attention Refinement module are provided in the supplementary materials.

3.3. Overlap Predictor

The combination of the two previous blocks produces features that capture geometric properties of each shape and are conditioned by the distribution of features on the other shape.

We then use these features to predict the overlap probability $p_{\mathcal{X}} \in \mathbb{R}^{n_{\mathcal{X}}}$ and $p_{\mathcal{Y}} \in \mathbb{R}^{n_{\mathcal{Y}}}$ where $n_{\mathcal{X}}$ and $n_{\mathcal{Y}}$ is the number of points on shape \mathcal{X} and \mathcal{Y} respectively. For this we use a simple two layer MLP with ReLU activations, and a Sigmoid layer [32] at the end, applied independently to each point. We interpret, e.g., $p_{\mathcal{X}}(i)$ as the probability of vertex i of shape \mathcal{X} to exist on shape \mathcal{Y} .

3.4. Regularized FMap Layer

While the previous block predicts the *existence* of points across shapes, our main goal is to establish dense point-to-point correspondences between co-existing points. To this end we use a regularized functional map module that aims to predict the functional map between shape \mathcal{X} and \mathcal{Y} , using the refined features $\mathcal{F}^R_{\mathcal{X}}$ and $\mathcal{F}^R_{\mathcal{Y}}$. We therefore

project the feature functions onto their respective spectral basis consisting of the first k Laplace-Beltrami eigenfunctions: $\mathbf{A} = \Phi_{\mathcal{X}}^+ \mathcal{F}^R(\mathcal{X})$ and $\mathbf{B} = \Phi_{\mathcal{Y}}^+ \mathcal{F}^R(\mathcal{Y})$.

We then use these refined coefficient matrices to solve for the optimal functional map. Rather than optimizing the energy defined in Equation 1 above, as done in [14], we modify the regularizer for improved robustness.

Specifically, we first observe that the commutativity term $E_{comm}(C) = \|C\Delta_{\mathcal{X}} - \Delta_{\mathcal{Y}}C\|_F^2$ can be written as a *mask* applied on the functional map $E_{comm}(C) = \sum_{i,j} M_{ij} C_{ij}^2$, where $M_{ij} = (\Lambda_{\mathcal{Y}}(i) - \Lambda_{\mathcal{X}}(j))^2$ (see Eq. (3) in [39]). Then, we note that the closed form minimization technique used in [14] to make the computation of the optimal functional map is differentiable and can be applied to *any* mask matrix M regardless of whether it is associated with the Laplacian commutativity term or not.

This motivates our objective to choose a mask that is well-suited for the partial setting. In our experiments, we test the original mask formulation, dubbed the Laplacian mask, the slanted mask proposed in [41], and the resolvent mask proposed in [39]. As argued in [39], the resolvent mask has a more solid theoretical foundation. In our work, we also observed that the resolvent mask consistently leads to better results in practice, and we adopt it throughout all our experiments.

In summary, we compute the functional map by minimizing the following energy:

$$C_{opt} = \arg \min_C \|C\mathbf{A} - \mathbf{B}\|_F^2 + \lambda \sum_{ij} C_{ij}^2 M_{ij}, \quad (2)$$

where

$$M_{ij} = \left(\frac{\Lambda_{\mathcal{Y}}(i)^\gamma}{\Lambda_{\mathcal{Y}}(i)^{2\gamma+1}} - \frac{\Lambda_{\mathcal{X}}(j)^\gamma}{\Lambda_{\mathcal{X}}(j)^{2\gamma+1}} \right)^2 + \left(\frac{1}{\Lambda_{\mathcal{Y}}(i)^{2\gamma+1}} - \frac{1}{\Lambda_{\mathcal{X}}(j)^{2\gamma+1}} \right)^2$$

and γ is the resolvent Laplacian parameter, which we will take equal to 0.5 in all our experiments.

3.5. Training losses

As all of our architecture blocks are differentiable, we train the network end to end using ground truth correspondences using backpropagation. Our overall loss is composed of three main components:

Spectral Loss given a ground truth point to point map, we convert it to a functional map, and penalize the deviation of our predicted map from it using the Frobenius norm $\mathcal{L}_{spec}(C_{opt}) = \|C_{gt} - C_{opt}\|_F^2$ as done in [14]. The ground truth map is computed as follows: $C_{gt} = \Phi_{\mathcal{Y}}^+ \Pi_{\mathcal{X}\mathcal{Y}} \Phi_{\mathcal{X}}$ where $\Pi_{\mathcal{X}\mathcal{Y}}$ is the ground truth point to point matrix. I.e., $\Pi_{\mathcal{X}\mathcal{Y}}(i, j) = 1$ if and only if vertex i on \mathcal{X} is matched to the vertex j on \mathcal{Y} and 0 otherwise. Note that, unlike the full shape matching, in the case of partial correspondence, $\Pi_{\mathcal{X}\mathcal{Y}}$ might have entire rows that are zero, associated with vertices for which no match exists on the target.

PointInfoNCE Loss In our experiments, we observed that the spectral loss alone is not sufficient to train the feature extractor, especially in the challenging case of holes, where the ground truth functional map can fail to provide important high-frequency information. To remedy this, we use a contrastive loss, where the distance between extracted features of matched points should be minimized, meanwhile, the distance between unmatched points features should be maximized. In our experiments, we used PointInfoNCE Loss [59] which can be expressed as:

$$\mathcal{L}_{nce}(F_{\mathcal{X}}^R, F_{\mathcal{Y}}^R) = -\sum_{(i,j) \in \mathcal{P}} \log \frac{\exp(F_{\mathcal{X}}^R(i) \cdot F_{\mathcal{Y}}^R(j) / \tau)}{\sum_{(.,k) \in \mathcal{P}} \exp(F_{\mathcal{X}}^R(i) \cdot F_{\mathcal{Y}}^R(k) / \tau)}$$

Here \mathcal{P} is the set of matched points, computed using ground truth point to point map, τ is a scaling parameter, and $F_{\mathcal{X}}^R(i)$ is the refined feature of point i in shape \mathcal{X} . Note that we use $F_{\mathcal{X}}^R$ to denote the refined features $\mathcal{F}^R(\mathcal{X})$ for compactness.

Overlap Loss To train the overlap predictor module, we penalize the computed probabilities using a binary classification loss. The ground truth labels $y_{\mathcal{X}}$ and $y_{\mathcal{Y}}$ are extracted from the ground truth point to point map:

$$\mathcal{L}_{over}(p_{\mathcal{X}}, p_{\mathcal{Y}}) = \frac{1}{2} (\mathcal{L}_{classif}(p_{\mathcal{X}}, y_{\mathcal{X}}) + \mathcal{L}_{classif}(p_{\mathcal{Y}}, y_{\mathcal{Y}}))$$

where $\mathcal{L}_{classif}$ is the binary cross entropy loss.

$$\mathcal{L}_{classif}(x, y) = \frac{1}{|x|} \sum_{i=1}^{|x|} y_i \log(x_i) + (1 - y_i) \log(1 - x_i)$$

Total loss our final loss is just a weighted sum to the previous losses: $\mathcal{L} = \lambda_1 \mathcal{L}_{spec} + \lambda_2 \mathcal{L}_{nce} + \lambda_3 \mathcal{L}_{over}$.

We use the same values of all of our hyperparameters throughout all of our experiments. We provide the complete implementation details and hyperparameter values in the supplementary.

Unsupervised variant While these losses depend on the presence of a ground truth correspondence, our pipeline can also be adapted to an *unsupervised* setting by using losses that promote desirable structural properties of the functional maps. We provide the details and evaluate one such possibility in the supplementary materials.

Ablation Study Our method comprises multiple building blocks that we consider essential for optimal performance. In the supplementary, we report an ablation study that demonstrates the efficacy of individual components. More significantly, we observe that the spectral loss is crucial for the convergence of the overlap predictor network and also report that the *resolvent* mask is best suited in our setting.

4. Experiments and Analysis

In this section, we provide extensive experiments highlighting the accuracy and robustness of our method on a range of non-rigid partial shape matching tasks. We consider three main challenges: partial-to-full, partial-to-partial, and finally partial-to-full under extreme variability in mesh discretization.

4.1. Datasets

We evaluate our method qualitatively and quantitatively using the following benchmarks. See the supplementary for visualization of shapes from each of the following datasets.

The SHREC’16 Partial Correspondence Benchmark [11] is a widely used benchmark to evaluate partial non-rigid shape matching pipelines. This dataset consists of 76 nearly isometric shapes, divided into 8 classes (humans and animals). Each class has a ‘null’ shape to which the partial shapes will be matched. As an additional challenge, all shapes are individually re-meshed to 10K vertices.

There are two subsets of **SHREC’16**, namely **CUTS** and **HOLES**. **CUTS** is constructed by cutting the shapes with a random plane at different orientations. The **HOLES** dataset is more challenging and contains shapes with surfaces that have been eroded by various random seeds, causing them to have actual holes and very irregular cuts.

SHREC’16 Part-to-Part dataset (CP2P): We introduce a new dataset to evaluate partial-to-partial shape matching. **CP2P** is based on the **CUTS** training set in [11] and contains the same 8 classes of humans and animals in different poses and partialities and establishes a total of 300 pairs for matching. The train and test sets are divided randomly with 80% pairs for training, and the remaining 20% for testing. We highlight that, similar to **SHREC’16** the ground truth map is only provided for shapes from the same class. The overlap between the source and target shapes varies from 10% to 90% of the total area.

Partial FARM dataset (PFARM): We introduce a new dataset with the objective of testing the robustness of partial shape matching methods to significantly varying mesh connectivity and sampling. Our dataset (**PFARM**) is designed using the SHREC’19 dataset [29] and is an extension of the recently introduced FARM partial dataset [22]. **PFARM** comprises of 27 test pairs resulting from 28 human shapes (one null shape). The shapes have significantly different meshing, vertex density, undergo varied rigid and non-rigid deformations and have significant partiality.

4.2. Baselines

We compare with the following prior partial shape matching methods: Partial Functional Maps (PFM) [41], Fully Spectral Partial Shape Matching (FSP) [26], Matching Deformable Objects in Clutter (DOC) [12] and also a recent

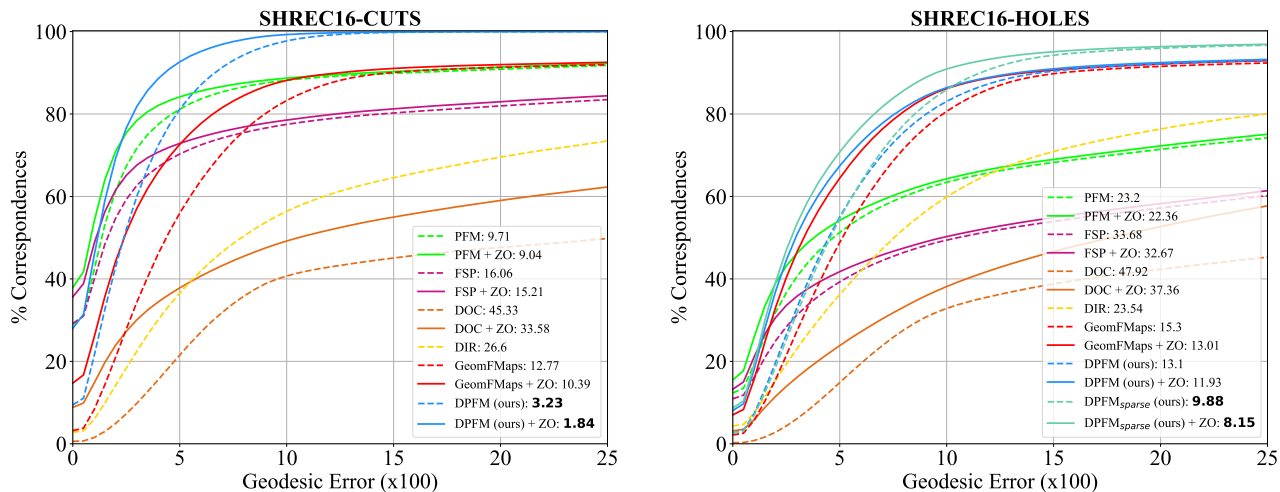


Figure 4: Correspondence quality of different methods on the test set of SHREC’16 Partial Benchmark, both on cuts (left) and holes (right). Our method outperforms all the competing methods and achieves state-of-the-art results. Mean errors of all methods are reported in the legend.

method: Dual Iterative Refinement (DIR) [58]. Since our work is the first to tackle a *learning* approach to non-rigid partial shape matching, we also include GeomFMaps [14] as a *learning* baseline, since that method is similar in spirit to our framework. In addition, we refine almost all methods with the partial variant of the Zoomout algorithm [30]. For all competing methods, we use the original code released by the authors and apply the best parameters reported in the respective papers.

Evaluation Protocol We follow the Princeton protocol [20] for evaluating all correspondences, as used in all recent works. Specifically, we compute the pointwise geodesic error between the predicted maps and the ground truth map and normalize by the square root of the total surface area of the target shape.

4.3. Part To Full Shape Matching

We first show the results of our method for partial-to-full matching on SHREC’16. We individually train our networks on the train set of SHREC’16 comprising of 120 pairs (CUTS) and 80 pairs (HOLES) respectively.

It should be noted that contrary to many of the previous works, which are evaluated on the train set of CUTS and HOLES, we follow the original SHREC’16 Benchmark protocol, and test our method, and all the baselines *only on the test set*, which contains 200 pairs each for CUTS and HOLES. The test set contains unseen shapes during training and provides a diverse and *significantly* more challenging set of partiality scenarios.

As reported in Figure 4, our method achieves state-of-the-art results across all settings in SHREC’16 outperforming both axiomatic and learning-based prior methods. Addition-

ally, we show the performance of our sparse variant Ours_{sparse} on the HOLES dataset, as the latter can handle particularly sparse data, and show that our architecture is not limited to the choice of the feature extractor.

In the supplementary materials, we also report two additional results in this setup. First, we plot the mean geodesic error as a function of the percentage of partiality. We observe that our method achieves the lowest mean error among all competing methods, and the error is stable even with an increasing amount of partiality, demonstrating the robustness of our method. Secondly, we also report the quality of overlap region prediction (which is a feature of only our method and PFM) by plotting the intersection over union (IOU) of the predicted region and the ground truth. We observe that our method achieves a high IOU, outperforming PFM, especially for CUTS, which demonstrate the quality of the predictions made by the overlap module.

4.4. Robustness to Mesh Discretization

Figure 5 compares the performance of all methods on the PFARM dataset. As the size of this dataset is small, and is not adapted to training, we only used the pretrained networks on CUTS for both our method and GeomFMaps.

We observe that competing methods that rely on hand-crafted features like SHOT, tend to *overfit* to the mesh connectivity and as a result see a significant performance drop on meshes with variable sampling and connectivity like PFARM. In contrast, our network which is *pre-trained on CUTS* from SHREC’16 generalises significantly better and has considerably lower errors than the competitors. Figure 6 shows a visualization of the predicted maps from all methods. It can be seen that only our method yields visually plausible correspondences on the shapes from PFARM.

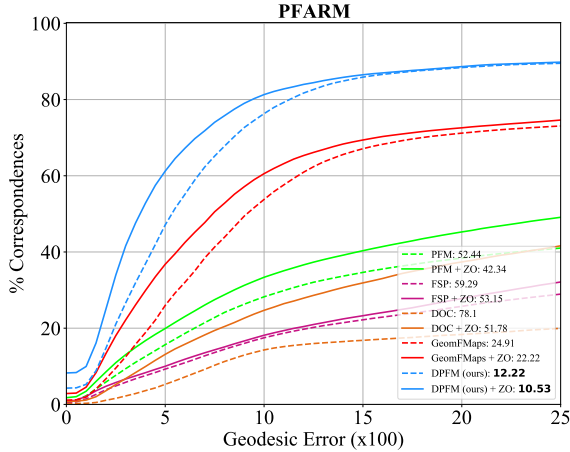


Figure 5: Results on the PFARM dataset. Our method is robust to significant changes in mesh discretization and outperforms all the baselines, whether learning-based or axiomatic.

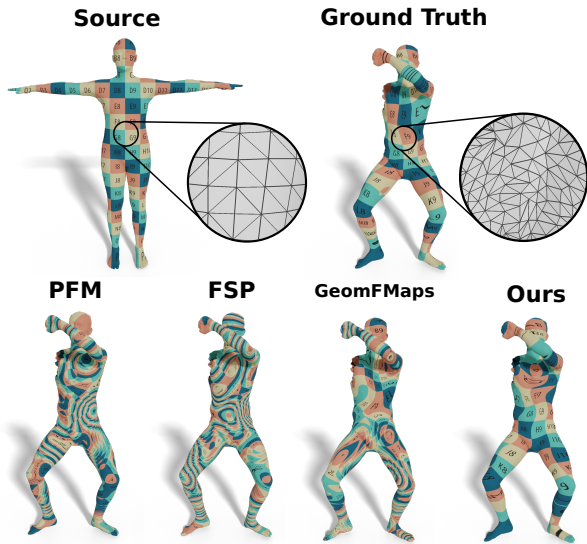


Figure 6: Qualitative results on PFARM using texture transfer. Only our method produces visually plausible results. Additional visualizations for SHREC’16 and CP2P are provided in the supplementary.

4.5. Part To Part Shape Matching

We evaluate our method on the challenging task of partial-to-partial shape matching, where only a portion of the source shape is mapped to a portion of the target shape. We compare our method against PFM [41], the formulation of FSP adapted to the partial-to-partial setting [26], DOC [12], and the learning-based baseline GeomFMaps [14] on the CP2P dataset. We apply PFM in both directions (source to target and target to source), thereby generating a predicted region of overlap on both source and target shapes. Since DOC [12] already optimizes for two overlap regions, we use their method as-is by initializing with SHOT descriptors. In order

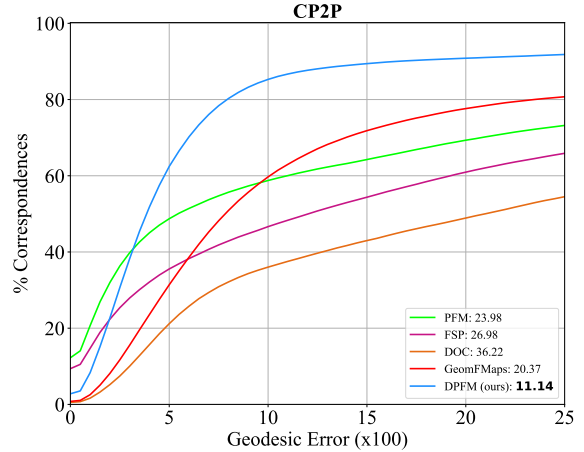


Figure 7: Quantitative results of the different methods on the CP2P dataset. Our method outperforms all baselines.

to evaluate the quality of matching, we show the error plot for correspondences *only in the overlapping region*. Figure 7 shows the superiority of our method over the baselines. We also plot the IOU of the predicted overlap region to assess the quality of the overlap region detection, our method gives superior results (see the supplementary).

5. Conclusion, Limitations, and Future work

In this paper, we present a learning framework specifically aimed for partial non-rigid shape matching. We exploit the versatility and accuracy of powerful geometric feature extractors along with the strong regularization provided by the functional map formulation. Key to our design is an attention-based feature refiner, which ensures cross-talk between features on the two shapes and is therefore crucial in the context of partial shape correspondence. Our method gives a significant improvement upon existing baselines, especially highlighted in challenging scenarios of considerable variability in mesh discretization.

Our approach still has some limitations and leads to possible exciting future work. Firstly, similar to many existing methods in practice, we assume that the shapes undergo a near-isometric deformation on the overlap region, a restrictive assumption that could be potentially lifted. In addition, extending our framework to more challenging representations such as noisy point clouds and graphs is also a fascinating area for future research.

Acknowledgements We gratefully acknowledge anonymous reviewers for their valuable suggestions. We also thank Rui Xiang for sharing code and help in executing the baselines as well as Simone Melzi and Emanuele Rodolà for fruitful discussions. Parts of this work were supported by the ERC Starting Grants No. 758800 (EXPROTEA) and the ANR AI Chair AIGRETTE.

References

- [1] Nadav Yehonatan Arbel, Ayellet Tal, and Lihi Zelnik-Manor. Partial correspondence of 3d shapes using properties of the nearest-neighbor field. *Computers & Graphics*, 82:183–192, Aug. 2019. 1, 2
- [2] Ben Bellekens, Vincent Spruyt, Rafael Berkvens, and Maarten Weyn. A survey of rigid 3d pointcloud registration algorithms. In *AMBIENT 2014: the Fourth International Conference on Ambient Computing, Applications, Services and Technologies, August 24-28, 2014, Rome, Italy*, pages 8–13, 2014. 2
- [3] Silvia Biasotti, Andrea Cerri, Alex Bronstein, and Michael Bronstein. Recent trends, applications, and perspectives in 3d shape similarity assessment. *Computer Graphics Forum*, 35(6):87–119, 2016. 2
- [4] Davide Boscaini, Jonathan Masci, Emanuele Rodolà, and Michael Bronstein. Learning shape correspondence with anisotropic convolutional neural networks. In D. Lee, M. Sugiyama, U. Luxburg, I. Guyon, and R. Garnett, editors, *Advances in Neural Information Processing Systems*, volume 29. Curran Associates, Inc., 2016. 2, 3
- [5] Sofien Bouaziz, Andrea Tagliasacchi, and Mark Pauly. Sparse iterative closest point. In *Computer graphics forum*, volume 32, pages 113–123. Wiley Online Library, 2013. 2
- [6] Michael M Bronstein, Joan Bruna, Yann LeCun, Arthur Szlam, and Pierre Vandergheynst. Geometric deep learning: going beyond euclidean data. *IEEE Signal Processing Magazine*, 34(4):18–42, 2017. 2
- [7] Wenming Cao, Zhiyue Yan, Zhiquan He, and Zhihai He. A comprehensive survey on geometric deep learning. *IEEE Access*, 8:35929–35949, 2020. 2
- [8] Umberto Castellani and Adrien Bartoli. 3d shape registration. In *3D Imaging, Analysis and Applications*, pages 353–411. Springer, 2020. 2
- [9] Dmitry Chetverikov, Dmitry Svirko, Dmitry Stepanov, and Pavel Krsek. The trimmed iterative closest point algorithm. In *Object recognition supported by user interaction for service robots*, volume 3, pages 545–548. IEEE, 2002. 2
- [10] Christopher Choy, Jaesik Park, and Vladlen Koltun. Fully convolutional geometric features. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 8958–8966, 2019. 2
- [11] Luca Cosmo, Emanuele Rodola, Michael M Bronstein, Andrea Torsello, Daniel Cremers, and Y Sahillioglu. Shrec’16: Partial matching of deformable shapes. *Proc. 3DOR*, 2(9):12, 2016. 4, 6
- [12] Luca Cosmo, Emanuele Rodola, Jonathan Masci, Andrea Torsello, and Michael M. Bronstein. Matching deformable objects in clutter. In *2016 Fourth International Conference on 3D Vision (3DV)*. IEEE, Oct. 2016. 2, 3, 6, 8
- [13] Haowen Deng, Tolga Birdal, and Slobodan Ilic. PPFnet: Global context aware local features for robust 3d point matching. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 195–205, 2018. 2
- [14] Nicolas Donati, Abhishek Sharma, and Maks Ovsjanikov. Deep geometric functional maps: Robust feature learning for shape correspondence. In *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, June 2020. 1, 2, 3, 4, 5, 7, 8
- [15] Shaoyi Du, Jihua Zhu, Nanning Zheng, Yuehu Liu, and Ce Li. Robust iterative closest point algorithm for registration of point sets with outliers. *Optical Engineering*, 50(8):087001, 2011. 2
- [16] Zan Gojcic, Caifa Zhou, Jan D Wegner, and Andreas Wieser. The perfect match: 3d point cloud matching with smoothed densities. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 5545–5554, 2019. 2
- [17] Benjamin Graham, Martin Engelcke, and Laurens Van Der Maaten. 3d semantic segmentation with submanifold sparse convolutional networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 9224–9232, 2018. 4
- [18] Oshri Halimi, Or Litany, Emanuele Rodola, Alex M Bronstein, and Ron Kimmel. Unsupervised learning of dense shape correspondence. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 4370–4379, 2019. 1, 2, 3
- [19] Shengyu Huang, Zan Gojcic, Mikhail Usvyatsov, Andreas Wieser, and Konrad Schindler. Predator: Registration of 3d point clouds with low overlap. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 4267–4276, 2021. 2, 4
- [20] Vladimir G. Kim, Yaron Lipman, and Thomas Funkhouser. Blended intrinsic maps. *ACM Transactions on Graphics*, 30(4):1–12, July 2011. 7
- [21] Thomas N. Kipf and Max Welling. Semi-supervised classification with graph convolutional networks. In *International Conference on Learning Representations (ICLR)*, 2017. 4
- [22] Maxime Kirgo, Simone Melzi, Giuseppe Patanè, Emanuele Rodolà, and Maks Ovsjanikov. Wavelet-based heat kernel derivatives: Towards informative localized shape analysis. *Computer Graphics Forum*, 40(1):165–179, Nov. 2020. 6
- [23] Artiom Kovnatsky, Michael M Bronstein, Alexander M Bronstein, Klaus Glashoff, and Ron Kimmel. Coupled quasi-harmonic bases. In *Computer Graphics Forum*, volume 32, pages 439–448. Wiley Online Library, 2013. 2
- [24] Qinsong Li, Shengjun Liu, Ling Hu, and Xinru Liu. Shape correspondence using anisotropic chebyshev spectral cnns. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2020. 2
- [25] Or Litany, Tal Remez, Emanuele Rodola, Alex Bronstein, and Michael Bronstein. Deep functional maps: Structured prediction for dense shape correspondence. In *Proceedings of the IEEE international conference on computer vision*, pages 5659–5667, 2017. 1, 2, 3
- [26] O. Litany, E. Rodolà, A. M. Bronstein, and M. M. Bronstein. Fully spectral partial shape matching. *Computer Graphics Forum*, 36(2):247–258, May 2017. 1, 2, 3, 6, 8
- [27] O. Litany, E. Rodolà, A. M. Bronstein, M. M. Bronstein, and D. Cremers. Non-rigid puzzles. *Computer Graphics Forum*, 35(5):135–143, Aug. 2016. 2, 3
- [28] 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. 1, 2
- [29] S. Melzi, R. Marin, E. Rodolà, U. Castellani, J. Ren, A. Poulernard, P. Wonka, and M. Ovsjanikov. Matching Humans with Different Connectivity. In Silvia Biasotti, Guillaume Lavoué, and Remco Veltkamp, editors, *Eurographics Workshop on 3D Object Retrieval*. The Eurographics Association, 2019. 6
- [30] Simone Melzi, Jing Ren, Emanuele Rodolà, Abhishek Sharma, Peter Wonka, and Maks Ovsjanikov. ZoomOut: Spectral Upsampling for Efficient Shape Correspondence. *ACM Transactions on Graphics*, 38(6):1–14, Nov. 2019. 2, 7
- [31] Federico Monti, Davide Boscaini, Jonathan Masci, Emanuele Rodola, Jan Svoboda, and Michael M Bronstein. Geometric deep learning on graphs and manifolds using mixture model cnns. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 5115–5124, 2017. 2, 3
- [32] Chigozie Nwankpa, Winifred Ijomah, Anthony Gachagan, and Stephen Marshall. Activation functions: Comparison of trends in practice and research for deep learning. *arXiv preprint arXiv:1811.03378*, 2018. 5
- [33] Maks Ovsjanikov, Mirela Ben-Chen, Justin Solomon, Adrian Butscher, and Leonidas Guibas. Functional maps: a flexible representation of maps between shapes. *ACM Transactions on Graphics (TOG)*, 31(4):1–11, 2012. 3
- [34] G Dias Pais, Srikumar Ramalingam, Venu Madhav Govindu, Jacinto C Nascimento, Rama Chellappa, and Pedro Miraldo. 3dregnet: A deep neural network for 3d point registration. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 7193–7203, 2020. 2
- [35] Leonid Pishchulin, Stefanie Wuhrer, Thomas Helten, Christian Theobalt, and Bernt Schiele. Building statistical shape spaces for 3d human modeling. *Pattern Recognition*, 67:276–286, 2017. 1
- [36] Adrien Poulernard and Maks Ovsjanikov. Multi-directional geodesic neural networks via equivariant convolution. *ACM Transactions on Graphics (TOG)*, 37(6):1–14, 2018. 2
- [37] 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*, page 5105–5114, Red Hook, NY, USA, 2017. Curran Associates Inc. 5
- [38] Arianna Rampini, Irene Tallini, Maks Ovsjanikov, Alex M Bronstein, and Emanuele Rodola. Correspondence-free region localization for partial shape similarity via hamiltonian spectrum alignment. In *2019 International Conference on 3D Vision (3DV)*, pages 37–46. IEEE, 2019. 2
- [39] Jing Ren, Mikhail Panine, Peter Wonka, and Maks Ovsjanikov. Structured regularization of functional map computations. *Computer Graphics Forum*, 38(5):39–53, Aug. 2019. 5
- [40] Emanuele Rodola, Alex M Bronstein, Andrea Albarelli, Filippo Bergamasco, and Andrea Torsello. A game-theoretic approach to deformable shape matching. In *2012 IEEE Conference on Computer Vision and Pattern Recognition*, pages 182–189. IEEE, 2012. 3
- [41] E. Rodolà, L. Cosmo, M. M. Bronstein, A. Torsello, and D. Cremers. Partial functional correspondence. *Computer Graphics Forum*, 36(1):222–236, Feb. 2016. 1, 2, 3, 5, 6, 8
- [42] Jean-Michel Roufousse, Abhishek Sharma, and Maks Ovsjanikov. Unsupervised deep learning for structured shape matching. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 1617–1627, 2019. 2
- [43] Szymon Rusinkiewicz and Marc Levoy. Efficient variants of the icp algorithm. In *Proceedings third international conference on 3-D digital imaging and modeling*, pages 145–152. IEEE, 2001. 2
- [44] Yusuf Sahillioğlu. Recent advances in shape correspondence. *The Visual Computer*, 36(8):1705–1721, 2020. 2
- [45] Samuele Salti, Federico Tombari, and Luigi Di Stefano. Shot: Unique signatures of histograms for surface and texture description. *Computer Vision and Image Understanding*, 125:251–264, 2014. 1
- [46] Abhishek Sharma and Maks Ovsjanikov. Weakly supervised deep functional maps for shape matching. *Advances in Neural Information Processing Systems*, 33, 2020. 2, 3
- [47] Nicholas Sharp, Souhaib Attaiki, Keenan Crane, and Maks Ovsjanikov. Diffusionnet: Discretization agnostic learning on surfaces. *arXiv preprint arXiv:2012.00888*, 2021. 2, 3, 4
- [48] Robert W Sumner and Jovan Popović. Deformation transfer for triangle meshes. *ACM Transactions on graphics (TOG)*, 23(3):399–405, 2004. 1
- [49] Hugues Thomas, Charles R Qi, Jean-Emmanuel Deschaud, Beatriz Marcotegui, François Goulette, and Leonidas J Guibas. Kpconv: Flexible and deformable convolution for point clouds. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 6411–6420, 2019. 4
- [50] Dmitry Ulyanov, Andrea Vedaldi, and Victor Lempitsky. Instance normalization: The missing ingredient for fast stylization. *arXiv preprint arXiv:1607.08022*, 2016. 4
- [51] Oliver Van Kaick, Hao Zhang, Ghassan Hamarneh, and Daniel Cohen-Or. A survey on shape correspondence. *Computer Graphics Forum*, 30(6):1681–1707, 2011. 2
- [52] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is all you need. In *Advances in neural information processing systems*, pages 5998–6008, 2017. 4, 5
- [53] Petar Veličković, Guillem Cucurull, Arantxa Casanova, Adriana Romero, Pietro Lio, and Yoshua Bengio. Graph attention networks. *arXiv preprint arXiv:1710.10903*, 2017. 5
- [54] Yue Wang and Justin M Solomon. PRnet: Self-supervised learning for partial-to-partial registration. *arXiv preprint arXiv:1910.12240*, 2019. 2
- [55] Ruben Wiersma, Elmar Eisemann, and Klaus Hildebrandt. Cnns on surfaces using rotation-equivariant features. *ACM Transactions on Graphics (TOG)*, 39(4):92–1, 2020. 2, 3
- [56] E. Wilson and D.W. Tufts. Multilayer perceptron design algorithm. In *Proceedings of IEEE Workshop on Neural Networks for Signal Processing*, pages 61–68, 1994. 4
- [57] Yan Wu, Jun Yang, and Jinlong Zhao. Partial 3d shape functional correspondence via fully spectral eigenvalue alignment and upsampling refinement. *Computers & Graphics*, 92:99–113, Nov. 2020. 1, 2
- [58] Rui Xiang, Rongjie Lai, and Hongkai Zhao. A dual iterative refinement method for non-rigid shape matching. In *Proceedings of the IEEE/CVF Conference on Computer Vision and*

Pattern Recognition (CVPR), pages 15930–15939, June 2021.
[1](#), [2](#), [7](#)

- [59] Saining Xie, Jiatao Gu, Demi Guo, Charles R. Qi, Leonidas Guibas, and Or Litany. Pointcontrast: Unsupervised pre-training for 3d point cloud understanding. *Lecture Notes in Computer Science*, page 574–591, 2020. [6](#)
- [60] Bing Xu, Naiyan Wang, Tianqi Chen, and Mu Li. Empirical evaluation of rectified activations in convolutional network. *arXiv preprint arXiv:1505.00853*, 2015. [4](#)
- [61] Andy Zeng, Shuran Song, Matthias Nießner, Matthew Fisher, Jianxiong Xiao, and Thomas Funkhouser. 3dmatch: Learning local geometric descriptors from rgb-d reconstructions. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1802–1811, 2017. [2](#)