Implicit Field Supervision For Robust Non-Rigid Shape Matching

Ramana Sundararaman, Gautam Pai, and Maks Ovsjanikov

LIX, École Polytechnique, IP Paris
{sundararaman, pai, maks}@lix.polytechnique.fr

Abstract. Establishing a correspondence between two non-rigidly deforming shapes is one of the most fundamental problems in visual computing. Existing methods often show weak resilience when presented with challenges innate to real-world data such as noise, outliers, self-occlusion etc. On the other hand, auto-decoders have demonstrated strong expressive power in learning geometrically meaningful latent embeddings. However, their use in *shape analysis* has been limited. In this paper, we introduce an approach based on an auto-decoder framework, that learns a continuous shape-wise deformation field over a fixed template. By supervising the deformation field for points on-surface and regularizing for points off-surface through a novel *Signed Distance Regularization* (SDR), we learn an alignment between the template and shape *volumes*. Trained on clean water-tight meshes, *without* any data-augmentation, we demonstrate compelling performance on compromised data and real-world scans. [1]

Keywords: Non-rigid 3D Shape correspondence, Neural Fields

1 Introduction

Understanding the relations between non-rigid 3D shapes through dense correspondences is a fundamental problem in computer vision and graphics. A common strategy is to leverage the underlying surfaces of shapes represented as triangle meshes. While recent advancements [65,21] demonstrate near-perfect correspondence accuracies, they strongly rely on idealistic settings of clean input data, which unfortunately is far from typical 3D acquisition setups. The question of generalizability of non-rigid shape correspondence to artifacts such as noise, outliers, self-occlusions, clutters, partiality, etc. which are innate to general 3D scans, is largely unanswered.

On the other hand, 3D shape representations through neural fields [80] or learned implicit functions have been shown to achieve remarkable accuracy, flexibility and generative power for a wide range of shape and scene modeling tasks [46,14,54,67]. Unlike standard shape representations, learning implicit functions through a neural network allows one to capture continuous surfaces, while seamlessly adapting to changes in topology. Indeed, implicit surface representations

[1] Our code is available at https://github.com/Sentient07/IFMatch

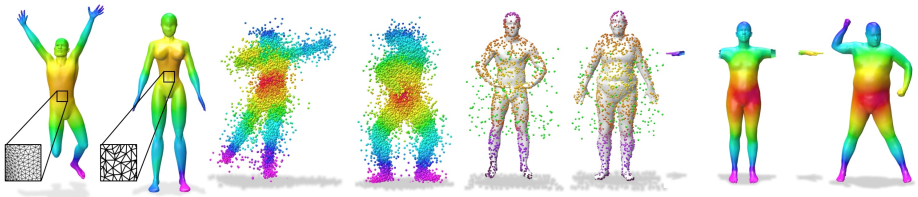

Fig. 1: Key advantages of our non-rigid shape correspondence pipeline: Our approach is extremely robust to common artifacts in 3D shapes like: (a) variations in sampling density, (b) significant noise, (c) cluttered outliers and (d) partiality.

not only allow to introduce an adaptive level of detail, but can also benefit from strong network regularization to control the desired resolution [71,66]. As a result, although initial efforts have focused on using implicit representations primarily for generative modeling and shape recovery, several recent works have shown their utility in other tasks including differentiable rendering for image synthesis [40,68], part-level shape decomposition [55], modeling dynamic geometry [51] and novel view synthesis [48,50] among many others.

This flexibility of implicit surface representations, however, comes at a cost, especially in applications that involve multiple shapes, such as shape correspondence or comparison. Since the surface is defined as the zero-level set of a function, individual points are no longer easily identifiable. As a result, recent methods based on implicit surface representations that have aimed at shape alignment, try to model a warping field over an underlying template [17,84,33], or between shape pairs [7]. All of these works, however, primarily focus on deformations across nearby, sufficiently similar 3D shapes.

In this paper, we introduce an efficient method for establishing correspondences across *arbitrary* non-rigid shapes, using neural field representations. To this end, we develop a new architecture based on the auto-decoder framework [54], that aims to recover a 3D deformation field between a fixed template and a target shape *volume*. The key ingredient of our architecture is defining the shape-wise deformation field from the latent embedding, augmented with two effective regularizations. First, we regularize the deformation field for arbitrary points in space through a novel *Signed Distance Regularization* (SDR). Second, we simultaneously condition the latent embedding to be compact and geometrically meaningful by learning a continuous Signed Distance Function (SDF) representation of the target shape. The resulting method is able to compute dense point-to-point correspondences between shapes while being extremely robust in the presence of varying sampling density, noise, cluttered outliers and missing parts as shown in Figure 1. To the best of our knowledge, ours is the first non-rigid correspondence method, based on neural field representation, that can be generalized to arbitrary shape categories such as articulated humans and animals.

Training on clean watertight meshes without any data-augmentation, we evaluate on a wide range of challenges across multiple benchmarks as well as real data captured by a 3D-scanner. Our approach shows compelling resilience to challenging artifacts and is more robust than existing point-based, mesh-based and spectral methods. In summary, our main contributions are: **(1)** We introduce an efficient approach based on the auto-decoder framework, capable of recovering a *volumetric* deformation field to align a source and a target shape *volumes*, even for significant non-rigid deformations. **(2)** We propose a novel way of regularizing the deformation of arbitrary points in space through the Signed Distance Regularization (SDR). **(3)** We perform rigorous evaluations by introducing challenges to existing benchmarks and on real-world data acquired by a 3D-scanner.

2 Related Work

2.1 Mesh-based Shape Correspondence

There is a large body of literature on shape matching, for shapes represented as triangle meshes. We refer interested readers to recent surveys [73,70,9,63] for a more comprehensive overview. Notable axiomatic approaches in this category are based on the functional maps paradigm [53,37,2,61,23,13]. Typically, these methods solve for near isometric shape correspondence by estimating linear transformations between spaces of real-valued functions, represented in a reduced functional basis. The conceptual framework of functional maps was further improved by learning-based formulations [39,30,62,19,22] that predict and penalize the map as a whole. Concurrently, recent advances in geometric deep learning have also tackled the correspondence problem by designing novel architectures for mesh and point cloud representation [49,12,43,58,78,38,83,21]. Such methods typically treat the correspondence learning problem as vertex labelling, which is learned efficiently using the respective architectures.

However, these methods that are predominantly based on mesh based representation of shapes are prone to sub-par performance when exposed to artifacts like sensitivity to variations in mesh discretization [65], sampling, missing or occluded parts, noise and other challenges that are common in typical 3D acquisition setups.

2.2 Template Based Shape Correspondence

Deforming a template to fit any given shape is a well-established technique in non-rigid shape registration [3,4]. The advent of learning-based skinning techniques [41,86,85] enabled deformation of a fixed template to an arbitrary shape and pose by calibrating a fixed set of SMPL model parameters. The introduction of parametric models has opened the avenue for generating copious amounts of training data [75,26,74] for data-driven methods. Such data-driven techniques have led to some seminal works in: 3D pose estimation [35,28,52], digitizing

humans [16,81] and even model-based 3D shape registration for articulated humans [10,57,11,56,8]. Most relevant to our work is LoopReg [8], which proposes to diffuse SMPL parameters in space to learn correspondence. In contrast, our approach does not require any parametric models as priors and can be generalized across arbitrary categories.

On the other hand, there are techniques that learn a model-free deformation to align a fixed template to a target shape [26,18,27,77]. Most notable among them is 3D-CODED [26], which learns to deform a fixed template mesh to a target shape. While this approach is succinct and well-founded, it requires significant amounts of training data to achieve optimal performance. Moreover, the deformation space is confined only to the surface of a mesh and can suffer from deformation artifacts. To ameliorate this, recent methods [17,84] have chosen to "implicitly define the template". However, their application in non-rigid shape matching is limited.

2.3 Neural Field Shape Representations

Coordinate-based neural networks are emerging methods for efficient, differentiable and high-fidelity shape representations [54,6,15,66,51,25,31,79,82] whose fundamental objective is to represent zero level-sets using parameters of neural network. In its most general form [54,66,67], these methods share two principal common goals - to perform differentiable surface reconstruction and to learn a latent shape embedding. This has given rise to numerous applications especially in the field of generative 3D modelling [72], such as shape editing [31,69,76], shape optimization [47,24] and novel view synthesis [48,71,50,64] to name a few. Most relevant to our work are DIF-Net [17] and SIREN [66] which achieve shape-specific surface reconstruction through Hyper-Networks [29]. However, leveraging the power of this representations in the domain of dense correspondence learning has so far been limited to nearly rigid objects [17,84,25,33].

3 Method

Notation: Throughout this manuscript, we use $\mathcal{S}$ to denote the target shape whose latent embedding is denoted by $\alpha_{\mathcal{S}} \in \mathbb{R}^{512}$ and $\mathcal{T}$ as the fixed template. $\mathcal{X}$ and $\mathcal{Y}$ denote an arbitrary pair of shapes between which we aim to find a correspondence. We let $\tilde{x} \in \partial\mathcal{S}$ be a point on the surface of the target shape $\mathcal{S}$, $x \in \mathbb{R}^3$ denotes a point in space and σ_x be its signed distance, $\sigma_x \coloneqq \mathrm{d}(x, \partial\mathcal{S})$. We define $[\mathcal{S}] \coloneqq \{x \in \mathbb{R}^3 | \sigma_x < \zeta\}$ to be the shape volume, which is the set of points sampled in space, in the vicinity of the shape surface $\partial\mathcal{S}$, with ζ being a constant. Analogously, $[\mathcal{T}] \coloneqq \{t_i \in \mathbb{R}^3 | \sigma_{t_i} < \zeta\}$ denotes the template volume.

3.1 Overview

Given a pair of shapes $\mathcal{X}$ and $\mathcal{Y}$, represented either as triangle meshes or point clouds, our goal is to estimate a point-wise map $\Pi : \mathcal{X} \rightarrow \mathcal{Y}$. To this end, we

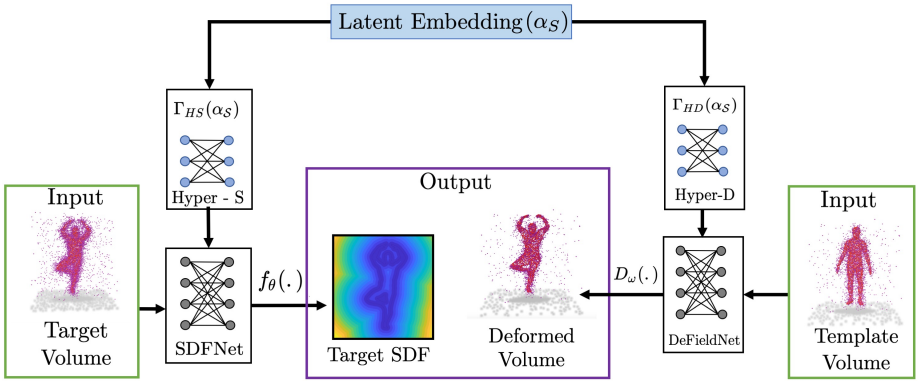

Fig. 2: Given a target shape volume (left) and a template volume (right) as input, DeFieldNet (Sec 3.2) aligns the template to target volume regularized by SDFNet (Sec 3.3). Shape-specific network weights are modeled by latent code (Sec 3.1). Points sampled within volumes (input) are shown only for visualization purposes to emphasize that our network operates over the 3D domain.

learn a *shape-specific* deformation field $D_\omega(.) : \mathbb{R}^3 \rightarrow \mathbb{R}^3$ which when applied to a fixed template volume $[\mathcal{T}]$, yields the target shape volume $[\mathcal{S}]$. Then, by using $D_\omega(.)$ to independently align $[\mathcal{T}]$ to $[\mathcal{X}]$ and $[\mathcal{Y}]$, we obtain correspondence between $\mathcal{X}$ and $\mathcal{Y}$ through nearest neighbor search. We stress that differently from previous data-driven works [26,18] that align a template mesh to a target mesh, our approach aligns two *volumes*. This is because, we observe that learning a volumetric alignment between arbitrary points in space naturally leads to a more robust map estimation as the deformation field is not constrained to an underlying surface defined by a mesh or a point cloud. While aligning on-surface points is straightforward in the supervised setting, aligning off-surface points is ill-posed. To this end, we propose a novel *Signed Distance Regularization (SDR)* for constraining the change in the SDF brought about by the deformation field. Learning a continuous deformation field also allows us to impose useful smoothness and volume preservation constraints, for enhancing the regularity of the map.

To make the deformation shape-specific, we learn a latent embedding $\alpha_\mathcal{S}$, which governs the parameters $\omega_\mathcal{S}$ of $D_{\omega_\mathcal{S}}(\cdot)$. We drop the subscript of ω for the sake of brevity. This latent embedding is learned following the auto-decoder framework [54]. However, constructing an embedding based on the deformation field alone leads to topological inconsistencies as we discuss in the ablation studies (refer to Suppl). Therefore, we introduce a geometric prior to $\alpha_\mathcal{S}$ by learning a continuous Signed Distance Function (SDF) representation of the shape, resulting in two concurrent auto-decoder networks as shown in Figure 2. On one side (left), we learn the continuous Signed Distance Function (SDF) of the target shape, which we refer to as *SDFNet*. Simultaneously (right of Figure 2), we

learn a deformation field D_ω over $[\mathcal{T}]$ through *DeFieldNet*. The parameters of our SDFNet $\theta \coloneqq \Gamma_{HS}(\alpha_\mathcal{S})$ and DeFieldNet $\omega \coloneqq \Gamma_{HD}(\alpha_\mathcal{S})$ are defined as two functions of the latent embedding, through Hyper-S and Hyper-D respectively. We perform an end-to-end training, to jointly learn the latent embedding $\alpha_\mathcal{S}$, through the gradients of SDFNet and DeFieldNet, similar to [29,66]. In summary, we learn a latent embedding by concurrently learning a deformation field over the template volume and the target shape's SDF. We stress that our main objective is to learn a plausible deformation field (via DeFieldNet) and the role of learning an implicit surface (via SDFNet) is to act as a geometric regularizer.

3.2 DeFieldNet

The main objective of DeFieldNet is to learn a smooth continuous shape-specific deformation field over the fixed template volume. We apply on surface supervision and off-surface regularization in order to deform the template volume $[\mathcal{T}]$ to the target shape volume $[\mathcal{S}]$.

On Surface Supervision: For two corresponding points $\tilde{x_i} \in \partial\mathcal{S}$ and $\tilde{t_i} \in \partial\mathcal{T}$, where $\Pi(\tilde{x_i}) = \tilde{t_i}$, our goal is to find a deformation $D_\omega : \tilde{t_i} \in \mathbb{R}^3 \rightarrow \vec{v} \in \mathbb{R}^3,\ \ s.t.\ \tilde{t_i} + \vec{v} \approx \tilde{x_i}$. Thus, solving for the desirable deformation field amounts to optimising the following loss:

$$\begin{aligned} \mathcal{L}_{\text{surf}} = & \sum_{\tilde{x}_i \in \partial S} ||\tilde{x}_i - \hat{\tilde{x}}_i||_2 \\ \text{where}, \hat{\tilde{x}}_i = & D_\omega(\tilde{t}_i) + \tilde{t}_i \end{aligned} \tag{1}$$

Signed Distance Regularization (SDR): In addition to supervising the deformation of points on the surface, we also regularize the deformation field applied to arbitrary points in the template volume $t \in [\mathcal{T}] \in \mathbb{R}^3$. For this, we propose a *Signed Distance Regularization* which *preserves* the Signed Distance Function under deformation for points sampled close to the surface. More specifically, given signed distances: $\sigma_{t_i}, \sigma_{\hat{x}_i}$ of points t_i, $\hat{x}_i$ respectively where $\hat{x}_i = D_\omega(t_i) + t_i$, we require $\sigma_{t_i} \approx \sigma_{\hat{x}_i}$, for all points sampled closed to the surface.

While σ_{t_i} is available as a result of pre-processing, computing $\sigma_{\hat{x}_i}$ requires a continuous signed distance estimator as the SDF is measured w.r.t deformed shape. Therefore we perform *discrete approximation* of the signed distance at any predicted point using Radial Basis Function (RBF) interpolation [32]. For any $\hat{x}_i \in \mathbb{R}^3$, we first construct the RBF kernel matrix Φ as a function of its neighbors in the target shape volume $\mathcal{N}(\hat{x}_i) \in [\mathcal{S}]$.

$$\Phi_{ij} \coloneqq \varphi(p_i, p_j) = \sqrt{\varepsilon_0 + ||p_i - p_j||^2} \tag{2}$$

Where, φ is the radial basis function and $p_{i,j} \in \mathcal{N}(\hat{x}_i)$. Assuming $\Delta = [\sigma_1 \dots \sigma_K]^T$ to be the vectorized representation of the SDF values of neighbors, the estimated SDF $\hat{\sigma}_{\hat{x}_i}$ of $\hat{x}_i$ w.r.t deformed template $\tilde{\mathcal{T}}$ is given as:

$$\hat{\sigma}_{\hat{x}_i} = \varphi\left(\hat{x}_i\right) \Phi^{-1} \Delta \tag{3}$$

We use shifted multiquadric functions as our RBF interpolant to avoid a singular interpolant matrix (refer to Suppl for more details). Therefore, our final SDF Regularization constraint can be written as:

$$\mathcal{L}_{SDR} = \sum_{t_i \in [\mathcal{T}]} ||\operatorname{clamp}(\sigma_{t_i}, \eta) - \operatorname{clamp}(\hat{\sigma}_{\hat{x}_i}, \eta)||_2 \tag{4}$$

Where $\operatorname{clamp}(x, \eta) := \min(\eta, \max(-\eta, x))$ is applied to make sure that the penalty is enforced only to points close to the surface. We highlight that this clamping is necessary, since the change in SDF under a considerable non-rigid deformation may differ significantly for points far from the surface.

Smooth Deformation: For the deformation field to be locally smooth, we ideally expect the flow vectors at neighboring points to be in "agreement" with each other. We enforce this constraint by encouraging the spatial derivatives to have minimal norm:

$$\mathcal{L}_{\text{Smooth}} = \sum_{t_i \in [\mathcal{T}]} || \nabla D_\omega(t_i)\|_2 \tag{5}$$

Volume Preserving Flow: Since a volume-preserving deformation field must be divergence-free, it must have a Jacobian with unit determinant [1].

$$\mathcal{L}_{\text{vol}} = \sum_{t_i \in [\mathcal{T}]} |\det(\nabla D_\omega(t_i)) - 1| \tag{6}$$

We use autograd to compute the Jacobian.

3.3 SDFNet

Given a set of N target shapes $\{\mathcal{S}_0 \dots \mathcal{S}_N\}$, our goal is to *regularize* their latent embedding $\{\alpha_{\mathcal{S}_0} \dots \alpha_{\mathcal{S}_N}\}$ through implicit surface reconstruction. We adopt the modified auto-decoder [66] framework with sinusoidal $\mathcal{C}^\infty$ activation function as our SDFNet. Given $f_\theta(\cdot) : x \in \mathbb{R}^3 \to \sigma_x \in \mathbb{R}$ to be the function that predicts the Signed Distance for a point $x \in [\mathcal{S}]$, SDFNet's learning objective is given by,

$$\begin{aligned}\mathcal{L}_{\mathcal{SDF}} = &\sum_{x \in [\mathcal{S}]} \left(\left| \|\nabla_x f_\theta(x)\|_2 - 1 \right| + |f_\theta(x) - \sigma_x| \right) + \sum_{\tilde{x} \in \partial S} \left(1 - \langle \nabla_x f_\theta(\tilde{x}), \hat{\mathbf{n}}(\tilde{x}) \rangle\right) \\ &+ \sum_{x \setminus \partial S} \psi(f(x))\end{aligned} \tag{7}$$

The first term penalizes the discrepancy in the predicted signed distance and enforces the Eikonal constraint for points in the shape volume. The second term encourages the gradient along the shape boundary to be oriented with surface normals. The last term applies an exponential penalty where $\psi := \exp(-C \cdot |\sigma_x|), C \gg 0$, for wrong prediction of $f_\theta(x) = 0$.

3.4 Training Objective:

In summary, the energy minimized at training time can be formulated as a combination of aforementioned individual constraints:

$$\mathcal{E}_{\text{Train}} = \Lambda_1 \mathcal{L}_{\mathcal{SDF}} + \Lambda_2 \mathcal{L}_{surf} + \Lambda_3 \mathcal{L}_{\mathcal{SDR}} + \Lambda_4 \mathcal{L}_{Smooth} + \Lambda_5 \mathcal{L}_{vol} \tag{8}$$

Here, Λ_i are scalars provided in Sec 3.6. The first term helps to regularize the latent space, while the other terms encourage a plausible deformation field.

3.5 Inference

At inference time, given $\mathcal{X}, \mathcal{Y}$ to be a pair of unseen shapes, our approach is three-staged. First, we find the optimal deformation function D_ω associated with $\mathcal{X}, \mathcal{Y}$ to deform $[\mathcal{T}]$. We solve for optimal parameters for our deformation field ω through Maximum-a-Posterior (MAP) estimation as:

$$\begin{aligned} \alpha_i &= \underset{\alpha_i}{\text{argmin}} \;\; \Lambda_1 \mathcal{L}_{\mathcal{SDF}} + \Lambda_3 \mathcal{L}_{\mathcal{SDR}} \\ \omega &\coloneqq \Gamma_{HD}(\alpha_i) \end{aligned} \tag{9}$$

Second, similar to [26] we enhance the deformation field applied by minimizing the bi-directional Chamfer's Distance

$$\alpha_{\mathbf{opt}} = \underset{\alpha_i}{\text{argmin}} \sum_{\tilde{\mathbf{s}} \in \partial \mathcal{S}} \min_{\tilde{\mathbf{t}}_i \in \partial \mathcal{T}} \left| D_\omega(\tilde{\mathbf{t}}_i) - \tilde{\mathbf{s}} \right|^2 + \sum_{\tilde{t}_i \in \partial \mathcal{T}} \min_{\tilde{\mathbf{s}} \in \partial \mathcal{S}} \left| D_\omega(\tilde{\mathbf{t}}_i) - \tilde{\mathbf{s}} \right|^2 \tag{10}$$

Finally, we establish the correspondence between $\mathcal{X}, \mathcal{Y}$ through their respective deformed templates using a nearest neighbor search.

3.6 Implementation details

Our two Hyper-Networks, SDFNet and DeFieldNet all use 4-layered MLPs with 20% dropout. SDFNet uses sinusoidal activation [66] while DeFieldNet uses ReLU activation. We fix $\Lambda_1 = 1, \Lambda_2 = 500$, $\Lambda_3 = 50$, $\Lambda_4 = 5$, $\Lambda_5 = 20$, namely the coefficients in Equation 8. For a shape in a batch, we use 4,000 points for on-surface supervision Equation 1. We use 8,000 points for SDF regularization in Equation 4 and $\eta = 0.1$ after fitting all shapes within a unit-sphere. We provide additional pre-processing details in the Suppl.

4 Experiments

Overview: In this section we demonstrate the robustness of our method in computing correspondences under challenging scenarios through extensive benchmarking. We perform our experiments across 4 datasets namely, FAUST [59], SHREC'19 [44], SMAL [86] and CMU-Panoptic dataset [34]. The first three are

mesh based benchmarks and are well-studied in non-rigid shape correspondence literature. In addition, we introduce challenging point cloud variants of these benchmarks which will be detailed below. CMU-Panoptic dataset [34], on the other hand, consists of raw point clouds acquired from a 3D scanner.

For evaluation, we follow the Princeton benchmark protocol [36] to measure mean geodesic distortion of correspondence on meshes. We perform evaluation on our point cloud variants by composing the predicted map to the nearest vertex point and measure the mean geodesic distortion [36]. On the CMU-Panoptic dataset [34], we measure the error on established key-points. We stress that across all experiments, while the evaluations are performed under challenging scenarios, our model is trained on clean water-tight mesh *without any data-augmentation*. Across all tables, "*" denotes a method that requires a mesh structure and cannot be evaluated on point clouds. "**" refers to computational in-feasibility in evaluating a baseline.

Baselines: We compare our method against several shape correspondence methods which can be broadly categorized into four main classes - axiomatic, spectral learning, template based and point cloud learning (PC Learning). We use ZoomOut (ZO) [45], BCICP [59] and Smooth Shells (S-Shells) [20] as our axiomatic baselines. For spectral basis learning baselines, we use Geometric Functional Maps (GeoFM) [19] with the recent more powerful DiffusionNet [65] feature extractor and DeepShells (D-Shells) [22]. We use 3D-CODED (3DC) [26], Deformed Implicit Fields (DIF-Net) [17] and Deep Implicit Templates (DIT-Net) [84] as template based baselines. We use Diff-FMaps (Diff-FM) [42], DPC [38] and Cornet [83] as our point cloud learning baselines. For a fair evaluation, we identically pre-train them according to their category for different experimental settings as mentioned in the respective sections. We provide more details on the hyper-parameters used for baselines in the Supplementary.

4.1 FAUST

Dataset: FAUST [10] dataset consists of 100 shapes where evaluation is performed on the last 20 shapes. Recently, Ren. *et al.* [59] introduced a re-meshed version of this dataset and Marin *et al.* [42] proposed a non-isometric, noisy point cloud version. For our robustness discussion, we introduce two additional challenges on top of the aforementioned variants. First, complimentary to [42], we introduce a dense point cloud variant consisting of 45,000 points perturbed with Gaussian noise. Second, we introduce 10% clutter points by random sampling of points in space. In summary, we perform evaluation on (1) Re-meshed shapes [59], (2) Non-isometric noisy point cloud (NI-PC) [42], (3) Dense point clouds with noise (De-PC) and (4) Clutter.

Baselines: We train our model and all data-driven methods on the first 80 meshes of the FAUST dataset. All baseline methods are trained using the publicly available code, following the configuration stipulated by the respective authors.

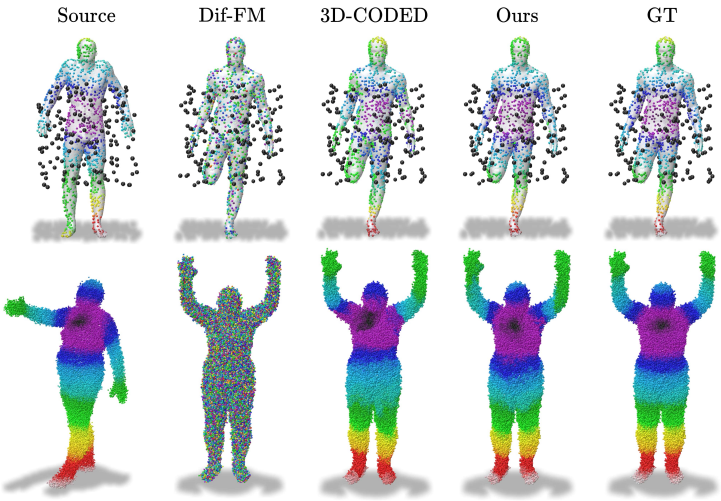

Fig. 3: Correspondence quality through color transfer on challenges we introduced to FAUST [59]. 1^{st} Row: Point clouds corrupted with 10% clutter shown in black. In contrast to baselines, our method shows strong resilience in the presence of clutter. 2^{nd} Row: Point cloud with 45k points and noise.

Category	Axiomatic			PC Learning			Spectral Learning		Template Based			
Method	BCICP [59]	ZO [45]	S-Shells [20]	Dif-FM [42]	DPC [38]	CorrNet [83]	D-Shells [22]	GeoFM [19]	3DC [26]	DIF-Net [17]	DIT-Net [84]	Ours
Remesh [59]	10.5	6.0	2.5	34.0	27.1	28.1	**1.7**	2.7	2.5	21.0	20.1	2.6
NI-PC + Noise [42]	11.5	8.7	*	6.6	8.4	25.2	*	31.3	7.3	14.6	13.6	**3.1**
De-PC + Noise	*	*	*	31.8	**	27.9	*	53.7	9.1	18.1	18.0	**4.1**
Clutter	*	*	*	17.7	50.0	51.1	*	52.2	22.1	14.7	14.3	**8.1**

Table 1: Quantitative results on FAUST-Remesh dataset and its variants reported as mean geodesic error (in cm) scaled by shape diameter.

Discussion: Our main quantitative results are summarized in Table 1. On the re-meshed shapes [59], our method demonstrates comparable performance with existing state-of-the-art methods. However, as we decrease the perfection of data, our method shows compelling resilience towards artifacts and consistently outperforms all the other baselines by a noticeable margin. It is also worthy to remark that among all baselines that we compare with, our method is the only one that is capable of providing reasonable (less than 10cm) correspondence in the presence of clutter points. We also show two qualitative examples on our newly introduced variant in Figure 3.

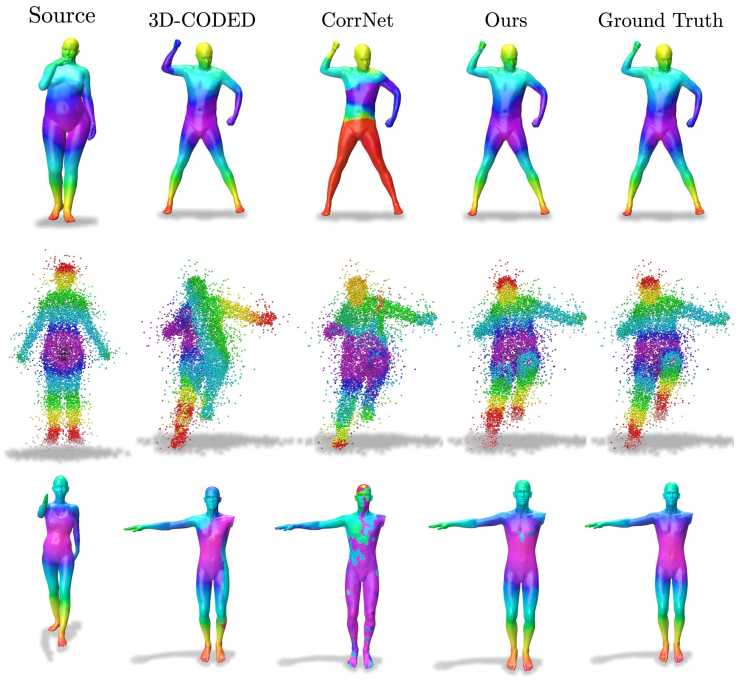

Fig. 4: Correspondence quality on SHREC'19 [44] and its variants. 1^{st} Row: Meshes. 2^{nd} Row: Point clouds with noise and outliers. 3^{rd} Row: Missing parts. Compared to baselines, our method exhibits strong resilience to artifacts.

4.2 SHREC'19

Dataset: SHREC'19 [44] is a challenging shape correspondence benchmark due to significant variations in mesh sampling, connectivity and presence of multiple connected components. It consists of 44 shapes and a total of 430 evaluation pairs. In addition, we introduce 3 challenging scenarios with different data imperfections. **Scenario 1:** We compare the meshes provided by Melzi *et al.* [44]. **Scenario 2:** We subsample the meshes to 10,000 points and introduce 20% outliers. **Scenario 3:** We further corrupt the surface information in Scenario 2 using Gaussian noise. **Scenario 4:** We introduce partiality in the form of missing parts, to a subset for a part-to-whole evaluation scheme [60].

Baseline: We pre-train all template based and point cloud learning baselines on 2,000 SURREAL shapes [75] including 10% humans in bent poses [26]. For our spectral basis learning baselines, we pre-train them on the training set of FAUST+SCAPE [5], consisting of $\binom{80}{2} + \binom{51}{2}$ shape pairs, a setting which is demonstrated to be best suited for them [22,19]. We use Partial Functional Map (PFM) [60] as an additional axiomatic baseline for Scenario 4.

Discussion: Quantitative results across 4 scenarios are summarized in Table 2. Our method demonstrates state-of-the-art performance across all variants

of the SHREC'19 dataset and remains inert to imperfections in the data. While Smooth-Shells [20], is comparable to our approach in Scenario 1, it cannot be evaluated in other scenarios due to its strong dependence on spectral information. Moreover, even among template based methods, it is important to note that the supervised learning baseline 3D-CODED [26] demonstrates significant decline in performance in the presence of outliers and noise. We posit that a well defined shape embedding, obtained by learning a volumetric mapping, plays a crucial role in our method's performance. Even among methods that construct a shape space through an auto-decoder framework, DIF-Net [17] and DIT [84], are not reliable when presented with non-rigid shapes. Among point cloud learning methods, while DPC [38] shows comparable performance to our approach in Scenario 2, their performance declines in Scenario 3, when surface information is corrupted by noise. Furthermore, since DPC [38] depends on input point cloud resolution, it is infeasible to be evaluated in Scenarios 1 and 4. Finally, despite training on clean meshes with no missing components, the performance of our approach is unaffected by the partiality introduced in Scenario 4. We attribute our learning of *volumetric alignment* coupled with off-surface regularization to be the reason behind robustness to missing components. We summarize this discussion by qualitatively depicting Scenarios 1, 3 and 4 in Figure 4, wherein, despite subsequently increasing artifacts, our method shows compelling resilience. Additional qualitative results in different poses are provided in the Supplementary.

Category	Axiomatic		PC Learning			Spectral Learning		Template Based			
Method	S-Shells [20]	PFM [60]	CorrNet [83]	DPC [38]	Diff-FM [42]	GeoFM [19]	D-Shells [22]	3DC [26]	DIF-Net [17]	DIT-Net [84]	Ours
Scenario: 1 (Meshes) [44]	7.6	N/A	13.4	**	29.6	11.7	15.2	9.2	14.9	41.4	**6.5**
Scenario: 2 (Outliers)	*	N/A	35.9	8.5	17.1	26.1	*	12.2	12.4	12.6	**7.4**
Scenario: 3 (Outliers + Noise)	*	N/A	36.0	11.5	16.7	27.8	*	14.4	36.2	12.5	**7.7**
Scenario: 4 (Missing parts)	*	52.4	23.5	**	26.3	48.6	23.8	6.0	11.9	41.1	**4.3**

Table 2: Quantitative results on 430 test set pairs of SHREC'19 dataset reported as mean geodesic error (in cm), scaled by shape diameter.

4.3 SMAL

Dataset: In this section, we show the generalization ability to *inter-class* non-rigid shape correspondence among to *non-human* shapes. To this end, we use the SMAL dataset [86], a parametric model that consists of 5 main categories of animals. We construct the training set by sampling 100 animals per each category. For correspondence evaluation, we generate 20 new shapes consisting of 4 animals per category, resulting in 180 *inter-class* evaluation pairs. We relax the degrees of freedom for selected joints while generating the test-set to introduce

new poses, unseen in the training set. In addition, we introduce partiality to this dataset in the form of multiple connected components.

Baseline Settings: We train all template based methods, including ours, on the aforementioned 500 training shapes. For our method and 3D-Coded, which are supervised template based methods, we share the same animal template. Since spectral basis learning baselines learn correspondence pairwise, we train all data-driven spectral methods on $\binom{100}{2}$ shapes with 20 animals per-category.

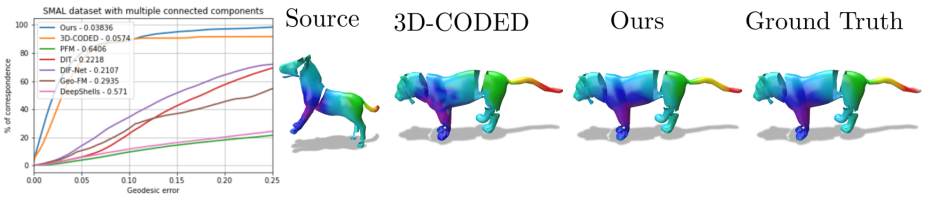

Fig. 5: Quantitative and qualitative inter-class correspondence on SMAL[86] dataset. Our approach produces a smooth map, unaffected by partiality.

Discussion: Our main quantitative and qualitative results are summarized in Figure 5. We observe that Geo-FM[19,65] that is a representation agnostic method and Partial Functional Maps, an approach built to tackle partial non-rigid shape correspondence methods fail to establish reasonable correspondence. Our approach on the other hand, remains agnostic to shape connectivity arising from inter-class non-isometry and introduced partiality. Finally, our method surpasses the template-based baseline method, 3D-Coded by a considerable margin.

4.4 CMU-Panoptic Dataset

Dataset: In this section, we demonstrate the generalization ability of our approach to real-world sensor data. To that end, we use the CMU Panoptic [34] dataset, which consists of 3+hrs footage of 8 subjects in frequently occurring social postures captured using the Kinect RGB+D sensor. This dataset consists of point clouds with noise, outliers, self-occlusions and clutter, allowing to evaluate correspondence methods on real-world data. We sample 200 shape pairs consisting of 3 distinct humans in 7 distinct poses. We measure non-rigid correspondence accuracy using the sparsely annotated anatomical landmark keypoints. More specifically, for each keypoint in the source, we consider 32 neighbors points and measure the disparity (as Euclidean distance, in cm) between their closest keypoint in the target and source.

Discussion: In order for a fair evaluation of generalizability, we test all approaches using the trained model elaborated in Section 4.2. Quantitative results of keypoint errors are summarized in Table 3. Our approach shows convincing performance in comparison to baselines, and more noticeably, it outperforms the

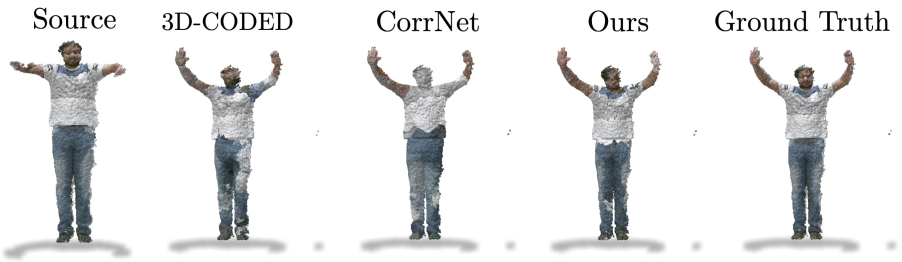

Fig. 6: Qualitative comparison using texture transfer on noisy point clouds from the CMU-Panoptic [34] dataset.

Method	Dif-FM[42]	GeoFM[19]	3D-CODED[26]	DIF-Net[17]	CorrNet[83]	DPC [38]	Ours
Keypoint Error	39.4	23.9	17.1	15.3	14.8	**	**8.5**

Table 3: Avg. Euclidean keypoint error (cm) for 200 test pairs, scaled by shape diameter.

conceptually closest supervised baseline, 3D-CODED [26] by a twofold margin. We also show a qualitative example through texture transfer in Figure 6, highlighting the efficacy of our approach in comparison to existing approaches on real-world data.

5 Conclusion, Limitations and Future work

We presented a novel approach for robust non-rigid shape correspondence based on the auto-decoder framework. Leveraging its strong expressive power, we demonstrated the ability of our approach in exhibiting strong resilience to practical artifacts like noise, outliers, clutter, partiality and occlusion across multiple benchmarks. To the best of our knowledge, our approach is the first to successfully demonstrate the use of Neural Fields, which predominantly are used as generative models, to the field of non-rigid shape correspondence, generalizable to arbitrary shape categories.

Despite various merits, we see multiple avenues for improvement and possible future work. Firstly, our current framework of joint learning of latent spaces by continuous functions opens possibilities for local descriptor learning alongside purely extrinsic information. This can potentially lead to an unsupervised pipeline in contrast to our existing supervised method. Also, auto-decoder style learning approaches are not rotation invariant and conventional techniques like data-augmentation can prove costly in terms of training effort. Making Neural Fields rotational invariant is also an interesting future direction.

References

1. Adams, B., Ovsjanikov, M., Wand, M., Seidel, H.P., Guibas, L.J.: Meshless modeling of deformable shapes and their motion. In: Proceedings of the 2008 ACM SIGGRAPH/Eurographics Symposium on Computer Animation. p. 77–86. SCA '08, Eurographics Association, Goslar, DEU (2008) 7
2. Aflalo, Y., Kimmel, R.: Spectral multidimensional scaling. Proceedings of the National Academy of Sciences **110**(45), 18052–18057 (2013) 3
3. Allen, B., Curless, B., Popović, Z.: Articulated body deformation from range scan data. ACM Trans. Graph. **21**(3), 612–619 (jul 2002). https://doi.org/10.1145/566654.566626, https://doi.org/10.1145/566654.566626 3
4. Allen, B., Curless, B., Popović, Z.: The space of human body shapes: Reconstruction and parameterization from range scans. ACM Trans. Graph. **22**(3), 587–594 (jul 2003). https://doi.org/10.1145/882262.882311, https://doi.org/10.1145/882262.882311 3
5. Anguelov, D., Srinivasan, P., Koller, D., Thrun, S., Rodgers, J., Davis, J.: Scape: Shape completion and animation of people. ACM Trans. Graph. **24**(3), 408–416 (jul 2005). https://doi.org/10.1145/1073204.1073207, https://doi.org/10.1145/1073204.1073207 11
6. Atzmon, M., Lipman, Y.: Sal: Sign agnostic learning of shapes from raw data. In: IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) (June 2020) 4
7. Atzmon, M., Novotny, D., Vedaldi, A., Lipman, Y.: Augmenting implicit neural shape representations with explicit deformation fields. arXiv preprint arXiv:2108.08931 (2021) 2
8. Bhatnagar, B.L., Sminchisescu, C., Theobalt, C., Pons-Moll, G.: Loopreg: Self-supervised learning of implicit surface correspondences, pose and shape for 3d human mesh registration. In: Larochelle, H., Ranzato, M., Hadsell, R., Balcan, M.F., Lin, H. (eds.) Advances in Neural Information Processing Systems. vol. 33, pp. 12909–12922. Curran Associates, Inc. (2020), https://proceedings.neurips.cc/paper/2020/file/970af30e481057c48f87e101b61e6994-Paper.pdf 4
9. Biasotti, S., Cerri, A., Bronstein, A., Bronstein, M.: Recent trends, applications, and perspectives in 3d shape similarity assessment. Computer Graphics Forum **35**(6), 87–119 (2016) 3
10. Bogo, F., Romero, J., Loper, M., Black, M.J.: FAUST: Dataset and evaluation for 3D mesh registration. In: Proceedings IEEE Conf. on Computer Vision and Pattern Recognition (CVPR). IEEE, Piscataway, NJ, USA (Jun 2014) 4, 9
11. Bogo, F., Romero, J., Pons-Moll, G., Black, M.J.: Dynamic faust: Registering human bodies in motion. In: 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). pp. 5573–5582 (2017). https://doi.org/10.1109/CVPR.2017.591 4
12. Boscaini, D., Masci, J., Rodolà, E., Bronstein, M.: Learning shape correspondence with anisotropic convolutional neural networks. In: Lee, D., Sugiyama, M., Luxburg, U., Guyon, I., Garnett, R. (eds.) Advances in Neural Information Processing Systems. vol. 29. Curran Associates, Inc. (2016), https://proceedings.neurips.cc/paper/2016/file/228499b55310264a8ea0e27b6e7c6ab6-Paper.pdf 3
13. Burghard, O., Dieckmann, A., Klein, R.: Embedding shapes with green's functions for global shape matching. Computers & Graphics **68**, 1–10 (2017) 3

14. Chen, Z., Zhang, H.: Learning implicit fields for generative shape modeling. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 5939–5948 (2019) 1
15. Chen, Z., Zhang, H.: Learning implicit fields for generative shape modeling. 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) pp. 5932–5941 (2019) 4
16. Corona, E., Pumarola, A., Alenyà, G., Pons-Moll, G., Moreno-Noguer, F.: Smplicit: Topology-aware generative model for clothed people. In: CVPR (2021) 4
17. Deng, Y., Yang, J., Tong, X.: Deformed implicit field: Modeling 3d shapes with learned dense correspondence. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 10286–10296 (2021) 2, 4, 9, 10, 12, 14
18. Deprelle, T., Groueix, T., Fisher, M., Kim, V.G., Russell, B.C., Aubry, M.: Learning elementary structures for 3d shape generation and matching. In: NeurIPS (2019) 4, 5
19. Donati, N., Sharma, A., Ovsjanikov, M.: Deep geometric functional maps: Robust feature learning for shape correspondence. In: 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). IEEE (Jun 2020). https://doi.org/10.1109/cvpr42600.2020.00862, https://doi.org/10.1109/cvpr42600.2020.00862 3, 9, 10, 11, 12, 13, 14
20. Eisenberger, M., Lähner, Z., Cremers, D.: Smooth shells: Multi-scale shape registration with functional maps. 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) pp. 12262–12271 (2020) 9, 10, 12
21. Eisenberger, M., Novotny, D., Kerchenbaum, G., Labatut, P., Neverova, N., Cremers, D., Vedaldi, A.: Neuromorph: Unsupervised shape interpolation and correspondence in one go. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 7473–7483 (2021) 1, 3
22. Eisenberger, M., Toker, A., Leal-Taixé, L., Cremers, D.: Deep shells: Unsupervised shape correspondence with optimal transport. arXiv preprint arXiv:2010.15261 (2020) 3, 9, 10, 11, 12
23. Ezuz, D., Ben-Chen, M.: Deblurring and denoising of maps between shapes. In: Computer Graphics Forum. vol. 36, pp. 165–174. Wiley Online Library (2017) 3
24. Gao, L., Yang, J., Wu, T., Yuan, Y.J., Fu, H., Lai, Y.K., Zhang, H.: Sdm-net: Deep generative network for structured deformable mesh. ACM Trans. Graph. **38**(6) (nov 2019). https://doi.org/10.1145/3355089.3356488, https://doi.org/10.1145/3355089.3356488 4
25. Genova, K., Cole, F., Vlasic, D., Sarna, A., Freeman, W.T., Funkhouser, T.A.: Learning shape templates with structured implicit functions. 2019 IEEE/CVF International Conference on Computer Vision (ICCV) pp. 7153–7163 (2019) 4
26. Groueix, T., Fisher, M., Kim, V.G., Russell, B., Aubry, M.: 3d-coded : 3d correspondences by deep deformation. In: ECCV (2018) 3, 4, 5, 8, 9, 10, 11, 12, 14
27. Groueix, T., Fisher, M., Kim, V., Russell, B., Aubry, M.: Unsupervised cycle-consistent deformation for shape matching. In: Symposium on Geometry Processing (SGP) (2019) 4
28. Güler, R.A., Neverova, N., Kokkinos, I.: Densepose: Dense human pose estimation in the wild. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 7297–7306 (2018) 3
29. Ha, D., Dai, A., Le, Q.V.: Hypernetworks (2016) 4, 6

30. Halimi, O., Litany, O., Rodola, E., Bronstein, A.M., Kimmel, R.: Unsupervised learning of dense shape correspondence. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 4370–4379 (2019) 3
31. Hao, Z., Averbuch-Elor, H., Snavely, N., Belongie, S.: Dualsdf: Semantic shape manipulation using a two-level representation. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (2020) 4
32. Hardy, R.L.: Multiquadric equations of topography and other irregular surfaces. J. Geophys. Res. **76**(8), 1905–1915 (mar 1971). https://doi.org/10.1029/jb076i008p01905, https://doi.org/10.10292Fjb076i008p01905 6
33. Jiang, C.M., Huang, J., Tagliasacchi, A., Guibas, L.J.: Shapeflow: Learnable deformations among 3d shapes. ArXiv **abs/2006.07982** (2020) 2, 4
34. Joo, H.e.a.: Panoptic studio: A massively multiview system for social interaction capture. TPAMI (2017) 8, 9, 13, 14
35. Kanazawa, A., Black, M.J., Jacobs, D.W., Malik, J.: End-to-end recovery of human shape and pose. In: Computer Vision and Pattern Regognition (CVPR) (2018) 3
36. Kim, V.G., Lipman, Y., Funkhouser, T.: Blended intrinsic maps. ACM Transactions on Graphics **30**(4), 1–12 (Jul 2011). https://doi.org/10.1145/2010324.1964974, https://doi.org/10.1145/2010324.1964974 9
37. Kovnatsky, A., Bronstein, M.M., Bronstein, A.M., Glashoff, K., Kimmel, R.: Coupled quasi-harmonic bases. In: Computer Graphics Forum. vol. 32, pp. 439–448. Wiley Online Library (2013) 3
38. Lang, I., Ginzburg, D., Avidan, S., Raviv, D.: DPC: Unsupervised Deep Point Correspondence via Cross and Self Construction. In: Proceedings of the International Conference on 3D Vision (3DV). pp. 1442–1451 (2021) 3, 9, 10, 12, 14
39. Litany, O., Remez, T., Rodola, E., Bronstein, A., Bronstein, M.: Deep functional maps: Structured prediction for dense shape correspondence. In: Proceedings of the IEEE international conference on computer vision. pp. 5659–5667 (2017) 3
40. Liu, S., Zhang, Y., Peng, S., Shi, B., Pollefeys, M., Cui, Z.: Dist: Rendering deep implicit signed distance function with differentiable sphere tracing. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 2019–2028 (2020) 2
41. Loper, M., Mahmood, N., Romero, J., Pons-Moll, G., Black, M.J.: SMPL: A skinned multi-person linear model. ACM Trans. Graphics (Proc. SIGGRAPH Asia) **34**(6), 248:1–248:16 (Oct 2015) 3
42. Marin, R., Rakotosaona, M.J., Melzi, S., Ovsjanikov, M.: Correspondence learning via linearly-invariant embedding. Proc. NeurIPS (2020) 9, 10, 12, 14
43. Masci, J., Boscaini, D., Bronstein, M., Vandergheynst, P.: Geodesic convolutional neural networks on riemannian manifolds. In: Proceedings of the IEEE international conference on computer vision workshops. pp. 37–45 (2015) 3
44. Melzi, S., Marin, R., Rodolà, E., Castellani, U., Ren, J., Poulenard, A., Wonka, P., Ovsjanikov, M.: Shrec 2019: Matching humans with different connectivity. In: Eurographics Workshop on 3D Object Retrieval. vol. 7 (2019) 8, 11, 12
45. Melzi, S., Ren, J., Rodolà, E., Sharma, A., Wonka, P., Ovsjanikov, M.: Zoomout: Spectral upsampling for efficient shape correspondence. ACM Trans. Graph. **38**(6) (nov 2019). https://doi.org/10.1145/3355089.3356524, https://doi.org/10.1145/3355089.3356524 9, 10
46. Mescheder, L., Oechsle, M., Niemeyer, M., Nowozin, S., Geiger, A.: Occupancy networks: Learning 3d reconstruction in function space. In: Proceedings of the

IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 4460–4470 (2019) 1
47. Mezghanni, M., Boulkenafed, M., Lieutier, A., Ovsjanikov, M.: Physically-aware generative network for 3d shape modeling. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). pp. 9330–9341 (June 2021) 4
48. Mildenhall, B., Srinivasan, P.P., Tancik, M., Barron, J.T., Ramamoorthi, R., Ng, R.: Nerf: Representing scenes as neural radiance fields for view synthesis. In: ECCV (2020) 2, 4
49. Monti, F., Boscaini, D., Masci, J., Rodola, E., Svoboda, J., Bronstein, M.M.: Geometric deep learning on graphs and manifolds using mixture model cnns. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 5115–5124 (2017) 3
50. Niemeyer, M., Geiger, A.: Giraffe: Representing scenes as compositional generative neural feature fields. In: Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR) (2021) 2, 4
51. Niemeyer, M., Mescheder, L., Oechsle, M., Geiger, A.: Occupancy flow: 4d reconstruction by learning particle dynamics. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 5379–5389 (2019) 2, 4
52. Omran, M., Lassner, C., Pons-Moll, G., Gehler, P.V., Schiele, B.: Neural body fitting: Unifying deep learning and model-based human pose and shape estimation. Verona, Italy (2018) 3
53. Ovsjanikov, M., Ben-Chen, M., Solomon, J., Butscher, A., Guibas, L.: Functional maps: a flexible representation of maps between shapes. ACM Transactions on Graphics (TOG) **31**(4), 1–11 (2012) 3
54. Park, J.J., Florence, P., Straub, J., Newcombe, R., Lovegrove, S.: Deepsdf: Learning continuous signed distance functions for shape representation. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 165–174 (2019) 1, 2, 4, 5
55. Paschalidou, D., Gool, L.V., Geiger, A.: Learning unsupervised hierarchical part decomposition of 3d objects from a single rgb image. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 1060–1070 (2020) 2
56. Pons-Moll, G., Pujades, S., Hu, S., Black, M.: Clothcap: Seamless 4d clothing capture and retargeting. ACM Transactions on Graphics, (Proc. SIGGRAPH) **36**(4) (2017), http://dx.doi.org/10.1145/3072959.3073711, two first authors contributed equally 4
57. Pons-Moll, G., Romero, J., Mahmood, N., Black, M.J.: Dyna: A model of dynamic human shape in motion **34**(4) (jul 2015). https://doi.org/10.1145/2766993, https://doi.org/10.1145/2766993 4
58. Poulenard, A., Ovsjanikov, M.: Multi-directional geodesic neural networks via equivariant convolution. ACM Transactions on Graphics (TOG) **37**(6), 1–14 (2018) 3
59. Ren, J., Poulenard, A., Wonka, P., Ovsjanikov, M.: Continuous and orientation-preserving correspondences via functional maps. ACM Trans. Graph. **37**(6) (dec 2018). https://doi.org/10.1145/3272127.3275040, https://doi.org/10.1145/3272127.3275040 8, 9, 10
60. Rodolà, E., Cosmo, L., Bronstein, M.M., Torsello, A., Cremers, D.: Partial functional correspondence. Computer Graphics Forum **36**(1), 222–236 (Feb 2016). https://doi.org/10.1111/cgf.12797, https://doi.org/10.1111/cgf.12797 11, 12

61. Rodolà, E., Cosmo, L., Bronstein, M.M., Torsello, A., Cremers, D.: Partial functional correspondence. In: Computer Graphics Forum. vol. 36, pp. 222–236. Wiley Online Library (2017) 3
62. Roufosse, J.M., Sharma, A., Ovsjanikov, M.: Unsupervised deep learning for structured shape matching. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 1617–1627 (2019) 3
63. Sahillioğlu, Y.: Recent advances in shape correspondence. The Visual Computer **36**(8), 1705–1721 (2020) 3
64. Schwarz, K., Liao, Y., Niemeyer, M., Geiger, A.: Graf: Generative radiance fields for 3d-aware image synthesis. In: Advances in Neural Information Processing Systems (NeurIPS) (2020) 4
65. Sharp, N., Attaiki, S., Crane, K., Ovsjanikov, M.: Diffusionnet: Discretization agnostic learning on surfaces (2021) 1, 3, 9, 13
66. Sitzmann, V., Martel, J., Bergman, A., Lindell, D., Wetzstein, G.: Implicit neural representations with periodic activation functions. Advances in Neural Information Processing Systems **33** (2020) 2, 4, 6, 7, 8
67. Sitzmann, V., Zollhöfer, M., Wetzstein, G.: Scene representation networks: Continuous 3d-structure-aware neural scene representations. arXiv preprint arXiv:1906.01618 (2019) 1, 4
68. Takikawa, T., Litalien, J., Yin, K., Kreis, K., Loop, C., Nowrouzezahrai, D., Jacobson, A., McGuire, M., Fidler, S.: Neural geometric level of detail: Real-time rendering with implicit 3d shapes. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 11358–11367 (2021) 2
69. Takikawa, T., Litalien, J., Yin, K., Kreis, K., Loop, C., Nowrouzezahrai, D., Jacobson, A., McGuire, M., Fidler, S.: Neural geometric level of detail: Real-time rendering with implicit 3D shapes (2021) 4
70. Tam, G.K., Cheng, Z.Q., Lai, Y.K., Langbein, F.C., Liu, Y., Marshall, D., Martin, R.R., Sun, X.F., Rosin, P.L.: Registration of 3d point clouds and meshes: A survey from rigid to nonrigid. IEEE transactions on visualization and computer graphics **19**(7), 1199–1217 (2012) 3
71. Tancik, M., Srinivasan, P.P., Mildenhall, B., Fridovich-Keil, S., Raghavan, N., Singhal, U., Ramamoorthi, R., Barron, J.T., Ng, R.: Fourier features let networks learn high frequency functions in low dimensional domains. NeurIPS (2020) 2, 4
72. Tiwari, G., Sarafianos, N., Tung, T., Pons-Moll, G.: Neural-gif: Neural generalized implicit functions for animating people in clothing. ArXiv **abs/2108.08807** (2021) 4
73. Van Kaick, O., Zhang, H., Hamarneh, G., Cohen-Or, D.: A survey on shape correspondence. Computer Graphics Forum **30**(6), 1681–1707 (2011) 3
74. Varol, G., Laptev, I., Schmid, C., Zisserman, A.: Synthetic humans for action recognition from unseen viewpoints **129**(7), 2264–2287 (May 2021). https://doi.org/10.1007/s11263-021-01467-7, https://doi.org/10.1007/s11263-021-01467-7 3
75. Varol, G., Romero, J., Martin, X., Mahmood, N., Black, M.J., Laptev, I., Schmid, C.: Learning from synthetic humans. In: CVPR (2017) 3, 11
76. Vasu, S., Talabot, N., Lukoianov, A., Baqué, P., Donier, J., Fua, P.: Hybridsdf: Combining free form shapes and geometric primitives for effective shape manipulation. ArXiv **abs/2109.10767** (2021) 4
77. Wang, W., Ceylan, D., Mech, R., Neumann, U.: 3dn: 3d deformation network. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) (June 2019) 4

78. Wiersma, R., Eisemann, E., Hildebrandt, K.: Cnns on surfaces using rotation-equivariant features. ACM Transactions on Graphics (TOG) **39**(4), 92–1 (2020) 3
79. Wu, R., Zhuang, Y., Xu, K., Zhang, H., Chen, B.: Pq-net: A generative part seq2seq network for 3d shapes. In: IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) (June 2020) 4
80. Xie, Y., Takikawa, T., Saito, S., Litany, O., Yan, S., Khan, N., Tombari, F., Tompkin, J., Sitzmann, V., Sridhar, S.: Neural fields in visual computing and beyond (2021), https://neuralfields.cs.brown.edu/ 1
81. Yu, T., Zheng, Z., Zhong, Y., Zhao, J., Dai, Q., Pons-Moll, G., Liu, Y.: Simulcap : Single-view human performance capture with cloth simulation. 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) pp. 5499–5509 (2019) 4
82. Zadeh, A., Lim, Y.C., Liang, P.P., Morency, L.P.: Variational auto-decoder. ArXiv **abs/1903.00840** (2019) 4
83. Zeng, Y., Qian, Y., Zhu, Z., Hou, J., Yuan, H., He, Y.: Corrnet3d: Unsupervised end-to-end learning of dense correspondence for 3d point clouds. In: IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) (2021) 3, 9, 10, 12, 14
84. Zheng, Z., Yu, T., Dai, Q., Liu, Y.: Deep implicit templates for 3d shape representation. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 1429–1439 (2021) 2, 4, 9, 10, 12
85. Zuffi, S., Kanazawa, A., Black, M.J.: Lions and tigers and bears: Capturing non-rigid, 3D, articulated shape from images. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR). IEEE Computer Society (2018) 3
86. Zuffi, S., Kanazawa, A., Jacobs, D., Black, M.J.: 3D menagerie: Modeling the 3D shape and pose of animals. In: IEEE Conf. on Computer Vision and Pattern Recognition (CVPR) (Jul 2017) 3, 8, 12, 13