Towards Precise Completion of Deformable Shapes

Oshri Halimi*[1], Ido Imanuel*[1], Or Litany[2], Giovanni Trappolini[3], Emanuele Rodolà[3], Leonidas Guibas[2], and Ron Kimmel[1]

[1] Technion - Israel Institute of Technology
[2] Stanford University
[3] Sapienza University of Rome

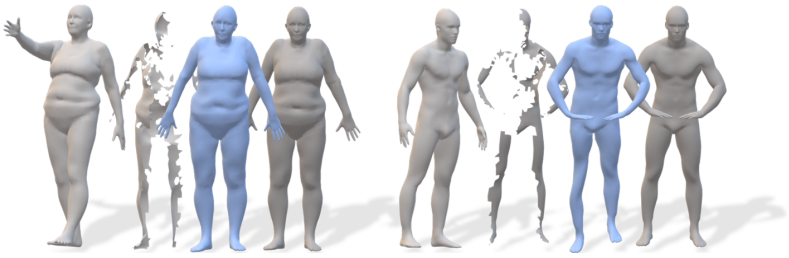

Fig. 1. Left to right: input reference shape, input part, output completion, and the ground truth full model.

Abstract. According to Aristotle, *"the whole is greater than the sum of its parts"*. This statement was adopted to explain human perception by the Gestalt psychology school of thought in the twentieth century. Here, we claim that when observing a part of an object which was previously acquired as a whole, one could deal with both partial correspondence and shape completion in a holistic manner. More specifically, given the geometry of a full, articulated object in a given pose, as well as a partial scan of the same object in a different pose, we address the *new* problem of matching the part to the whole while simultaneously reconstructing the new pose from its partial observation. Our approach is data-driven and takes the form of a Siamese autoencoder without the requirement of a consistent vertex labeling at inference time; as such, it can be used on unorganized point clouds as well as on triangle meshes. We demonstrate the practical effectiveness of our model in the applications of single-view deformable shape completion and dense shape correspondence, both on synthetic and real-world geometric data, where we outperform prior work by a large margin.

Keywords: Shape Completion · 3D Deep Learning · Shape Analysis

* equal contribution

1 Introduction

One of Aristotle's renowned sayings declares *"the whole is greater than the sum of its parts"*. This fundamental observation was narrowed down to human perception of planar shapes by the Gestalt psychology school of thought in the twentieth century. A guiding idea of Gestalt theory is the principle of *reification*, arguing that human perception contains more spatial information than can be extracted from the sensory stimulus, and thus giving rise to the view that the mind generates the additional information based on verbatim acquired patterns. Here, we adopt this line of thought in the context of non-rigid shape completion. Specifically, we argue that given access to a complete shape in one pose, one can accurately complete partial views of that shape at any other pose.

3D data acquisition using depth sensors is often done from a single view point, resulting in an incomplete point cloud. Many downstream applications require completing the partial observations and recovering the full shape. Based on this need, the task of shape completion has been extensively studied in the literature. The required fidelity of completion, however, is task dependent. In fact, in many cases even an approximate completion would be satisfying. For example, completing a car captured from one side by assuming the occluded side is symmetric would be perfectly acceptable for the purpose of obstacle avoidance in autonomous navigation, even if in reality the other side of that car has a large dent. In other cases, however, e.g. when capturing a person for telepresence or medical procedure purposes, it is crucial that the completion is exact, and no hallucination of shape details takes place. Clearly, this requirement is only viable given access to additional measurements or prior information. Here, we wish to focus on the latter case, which we coin as *precise shape completion*. In particular, provided a complete **non-rigid** shape in one pose, we require a solution for completing a partial view of the same shape in a different pose that is **accurate**, **fast**, and can handle **single-view partiality** resulting from self-occlusion. In this work we make a first attempt to address this specific setting, as opposed to the ubiquitous regime of precise *pose*-reconstruction, and as such, we focus solely on types of partialities that induce mild pose ambiguity.

Existing methods for rigid and non-rigid shape completion from partial scans fall largely into two categories: generative and alignment based. Generative methods have proven to be very powerful in completing shapes by learning to match the class distribution. However, they inherently aim at solving an ill-posed problem. Namely, they assume access only to the partial observation at inference time, and thus are incapable of performing *precise* shape completion of shapes unseen at train time. Non-rigid registration methods can take a full shape and align it to a partial observation and thus fit our prescribed setting. However, state of art methods are slow, and can usually handle only mild partiality. Here we propose a new method for precise completion of a partial non-rigid shape in an arbitrary (target) pose, given the full shape in a different (source) pose. Our method is fast, accurate and can handle severe partiality. Based on a deep neural network for point clouds, we learn a function that encodes the partial and full shapes, and outputs the complete shape at the target pose. By providing the

full shape, our completion achieves much higher accuracy than existing methods. Since completion in done a single feed-forward pass our solution is orders of magnitude faster than competing methods. In addition, our generated training set of rendered partial views and their corresponding complete shapes covers a broad range of plausible human poses, appearances and partialities which allows our method to gracefully generalize to unseen instances. Finally, our solution effortlessly recovers dense correspondences between the partial and full shapes that considerably improves state of the art performance on the FAUST projection benchmark.

Our main contributions can be summarized as follows:

1. We introduce a deep Siamese architecture to tackle *precise* non-rigid shape completion;
2. Our solution is significantly faster, more accurate and can handle more severe partialities than previous methods.
3. The recovered correspondences achieves state-of-the-art performance in partial shape correspondence.

2 Related work

Roughly speaking, there exist three approaches that address the challenge of reconstructing the geometry of an articulated shape from its partial scan, namely, partial non-rigid registration of surfaces, surface registration to a known skeleton, and shape completion of a given partial surface. While the first approach is the closest to our setting, none of these approaches has yet provided a good solution to the described application. Currently, state-of-the-art partial nonrigid shape registration/alignment [45,47,26] methods do not handle the significant partiality one often obtains when using commodity depth sensors, and their processing time even for a moderate-size point clouds of few thousand vertices vary between few minutes at best, to a few hours. In our experimental section, we compare with the most efficient methods belonging to this class.

The methods that can handle substantial partiality usually rely on some modification of the iterative closest point (ICP) algorithm and often have difficulties in handling large deformations between the full and the partial shape [66,50]. Another existing approach for the non-rigid alignment problem is to use an explicit deformation model such as skeleton rigging. These methods often completely ignore the detailed geometric and textural information in the actual scanned surface. Moreover, they rely on a rigged model for the full template, which is a limiting assumption when the full model is not restricted to a standard pose. The shape completion setting, as explained below, does not accommodate the full shape and therefore hallucinates details by construction, resulting in inferior completions, as shown by our results and ablation sections. We now turn to review some of the above approaches in more detail.

Shape completion.

Recovering a complete shape from partial or noisy measurements is a long-standing research problem that comes in many flavors. In an early attempt to use

one pose in order to geometrically reconstruct another, Devir *et al.* [21] considered mapping a model shape in a given pose onto a noisy version of the shape in a different pose. Elad and Kimmel were the first to treat shapes as metric spaces [22,23]. They matched shapes by comparing second order moments of embedding the intrinsic metric into a Euclidean one via classical scaling. In the context of deformable shapes, early efforts focused on completion based on geometric priors [36] or reoccurring patterns [13,38,62,40]. These methods are not suited for severe partiality. For such cases model-based techniques are quite popular, for example, category-specific parametric morphable models that can be fitted to the partial data [5,24,44,1,65]. Model-based shape completion was demonstrated for key-points input [2], and was recently proven to be quite useful for recovering 3D body shapes from 2D images [71,70,28,78]. Parametric morphable models [5], coupled with axiomatic image formation models were used to train a network to reconstruct face geometry from images [58,57,64]. Still, much less attention has been given to the task of fitting a model to a partial 3D point cloud. Recently, Jiang *et al.* [34] tackled this problem using a skeleton-aware architecture. However, their approach works well when full coverage of the underlying shape is given. [63] proposed a real time solution based on a reinforcement learning agent controlled by a GAN network. [32] reconstructed a 3D completion by generating and back-projecting multi-view depth maps. [75] focused on the ambiguity in completion from a single view, and suggested to address it using adversarially learned shape priors. Finally, [67] suggested a weakly supervised approach and showed performance on realistic data.

Nonrigid Partial shape matching. Dense non-rigid shape correspondence [37,17,41,29,60,15,19] is a key challenge in 3D computer vision and graphics, and has been widely explored in the last few years. A particularly challenging setting arises whenever one of the two shapes has missing geometry. Bronstein *et al.*[10,11,13,9,12,14] dealt with partial matching of articulated objects in various scenarios, including pruning of the intrinsic structure while accounting for cuts. This setting has been tackled with moderate success in a few recent papers [59,42,56], however, it largely remains an open problem whenever the partial shape exhibits severe artifacts or large, irregular missing parts. In this paper we tackle precisely this setting, demonstrating unprecedented performance on a variety of real-world and synthetic datasets.

Deep learning of surfaces. Following the success of convolutional neural networks on images in the recent years, the geometry processing community has been rapidly adopting and designing computational modules suited for such data. The main challenge is that unlike images, geometric structures like surfaces come in many types of representations, and each requires a unique handling. Early efforts focused on a simple extension from a single image to multi-view representations [68,74]. Another natural extension are 3D CNNs on volumetric grids [76]. A host of techniques for mesh processing were developed as part of a research branch termed *geometric deep learning* [16]. These include graph-based methods [72,73,30], intrinsic patch extraction [48,8,49], and spectral techniques [41,29]. Point cloud networks [54,55] have recently gained much attention. Offering a

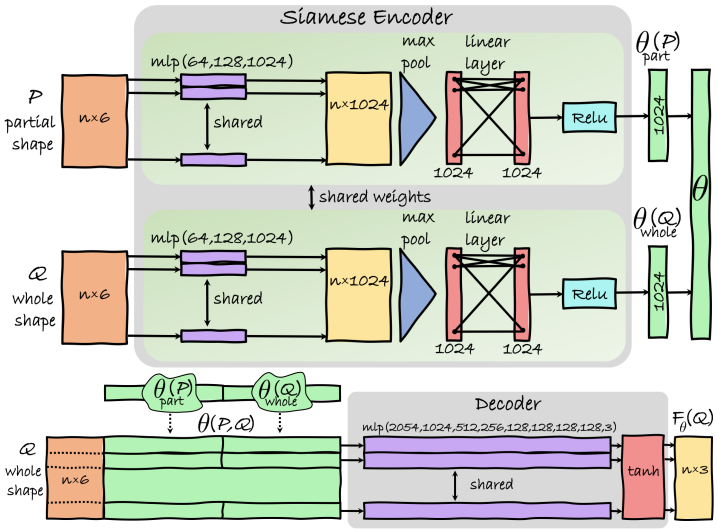

Fig. 2. Network Architecture. Siamese encoder architecture at the top, and the decoder (generator) architecture at the bottom. A shape is provided to the encoder as a list of 6D points, representing the spatial and unit normal coordinates. The latent codes of the input shapes $\theta_{part}(P)$ and $\theta_{whole}(Q)$ are concatenated to form a latent code θ representing the input pair. Based on this latent code, the decoder deforms the full shape by operating on each of its points with the same function. The result is the deformed full shape F_θ(Q).

light-weight computation restricted to sparse points with a sound geometric explanation [35], these networks have shown to provide a good compromise between complexity and accuracy, and are dominating the field of 3D object detection [53,77], semantic segmentation [25,3], and even temporal point cloud processing [18,43]. For generative methods, recent implicit and parametric methods have demonstrated promising results [27,51]. Following the success of encoding non-rigid shape deformations using a point cloud network [26], here, we also choose to use a point cloud representation. Importantly, while the approach presented in [26] predicts alignment of two shapes, it is not designed to handle severe partiality, and assumes a fixed template for the source shape. Instead, we show how to align arbitrary input shapes and focus on such a partiality.

3 Method

3.1 Overview

We represent shapes as point clouds $S = \{s_i\}_{i=1}^{n_s}$ embedded in $\mathbb{R}^3$. Depending on the setting, each point may carry additional semantic or geometric information encoded as feature vectors in $\mathbb{R}^d$. For simplicity we will keep $d = 3$ in our formulation. Given a full shape $Q = \{q_i\}_{i=1}^{n_q}$ and its partial view in a different

pose $P = \{p_i\}_{i=1}^{n_p}$, our goal is to find a nonlinear function $F : \mathbb{R}^3 \to \mathbb{R}^3$ aligning Q to P[4]. If $R = \{r_i\}_{i=1}^{n_r}$ is the (unknown) full shape such that $P \subset R$, ideally we would like to ensure that $F(Q) = R$, where equality should be understood as same underlying surface. Thus, the deformed shape $F(Q)$ acts as a proxy to solve for the correspondence between the part P and the whole Q. By calculating for every vertex in P its nearest neighbor in $R \approx F(Q)$, we trivially obtain the mapping from P to Q. The deformation function F depends on the input pair of shapes (P, Q). We model this dependency by considering a parametric function $F_\theta : \mathbb{R}^3 \to \mathbb{R}^3$, where θ is a latent encoding of the input pair (P, Q). We implement this idea via an encoder-decoder neural network, and learn the space of parametric deformations from example pairs of partial and complete shapes, together with full uncropped versions of the partial shapes, serving as the ground truth completion. Our network is composed of an encoder E and a generator F_θ. The encoder takes as input the pair (P, Q) and embeds it into a latent code θ. To map points from Q to their new location, we feed them to the generator along with the latent code. Our network architecture shares a common factor with 3D-CODED architecture [26], namely the deformation of one shape based on the latent code of the another. However [26] uses a fixed template and is therefore only suited for no or mild partiality, as the template cannot make up for lost shape details in the part. Our pipeline on the other hand, is designed to merge two sources of information into the reconstructed model, resulting in an accurate reconstruction under extreme partiality. In the supplementary we perform an analysis where we train our network in a fixed-template setting, similar to 3D-CODED and demonstrate the advantage of our paradigm. In what follows we first describe each module, and then give details on the training procedure and the loss function. We refer to Figure 2 for a schematic illustration of our learning model.

3.2 Encoder

We encode P and Q using a Siamese pair of single-shape encoders, each producing a global shape descriptor (respectively θ_{part} and θ_{whole}). The two codes are then concatenated so as to encode the information of the specific pair of shapes, $\theta = [\theta_{part}, \theta_{whole}]$. Considering the specific architecture of the single-shape encoder, we think about the encoder network as a channel transforming geometric information to a vector representation. We would like to utilize architectures which have been empirically proven to encode the 3D surface with the least loss of information, thus enabling the decoder to convert the resulting latent code θ to an accurate spatial deformation F_θ. Encouraged by recent methods [27,26] that showed detailed reconstruction using PointNet [54], we also adopt it as our backbone encoder. We provide the encoder 6 input channels, representing the vertex location and the vertex normal field. We justify this design choice in section 3.6. Specifically, our

[4] In our setting, we assume that the pose can be inferred from the partial shape (*e.g.*, an entirely missing limb would make the prediction ambiguous), hence the deformation function F is well defined.

encoder passes all 6D points of the input shape through the same "mlp" units, of hidden dimensions $(64, 128, 1024)$. Here, the term "mlp", carries the same meaning as in PointNet, i.e. multi-layer perceptron, with ReLU activation, and batch normalization layers. After a max-pool operation over the input points, we receive a single 1024-dimensional vector. Finally, we apply a linear layer of size 1024 and a ReLu activation function. Hence, each shape in the input pair is represented by a latent code $\theta_{whole}, \theta_{part}$ of size 1024 respectively. We concatenate these to a joint representation θ of size 2048.

3.3 Generator

Given the code θ, representing the partial and full shapes, the generator has to predict the deformation function F_θ to be applied to the full shape Q. We realize F_θ as a Multi-Layer Perceptron (MLP) that maps an input point q_i on the full shape Q, to its corresponding output point r_i on the ground truth completed shape. The MLP operates pointwise on the tuple (q_i, θ), with θ kept fixed. The result is the destination location $F_\theta(q_i) \in \mathbb{R}^3$, for each input point of the full shape Q. This generator architecture allows, in principle, to calculate the output reconstruction in a flexible resolution, by providing the generator a full shape with some desired output resolution. In detail, the generator consists of 9 layers of hidden dimensions $(2054, 1024, 512, 256, 128, 128, 128, 128, 3)$, followed by a hyperbolic tangent activation function. The output of the decoder is the 3D coordinates. In addition, we can compute a normal field based on the vertex coordinates, making the overall output of the decoder a 6D point. The normal is calculated using the known connectivity of the full shape Q, from our training dataset. Thus, the reconstruction loss in the below section could be generalized and defined using the normal output channel, as well. In the implementation section, we ablate this design choice, and show it leads to a performance improvement.

3.4 Loss function

The loss definition should reflect the visual plausibility of the reconstructed shape. Measuring such a quality analytically is a challenging problem worth studying on itself. Yet, in this paper we adopt a naive measurement of the Euclidean proximity between the ground-truth and the reconstruction. Formally, we define the loss as,

$$\mathcal{L}(P, Q, R) = \sum_{i=0}^{n_q} \left\| F_{\theta(P,Q)}(q_i) - r_i \right\|^2, \tag{1}$$

where $r_i = \pi^*(q_i) \in R$ is the matched point of $q_i \in Q$, given by the ground-truth mapping $\pi^* : Q \to R$.

3.5 Training Procedure

We train our model using samples from datasets of human shapes. These contain 3D models of different subjects in various poses. The datasets are described in

detail in Section 4.1. Each training sample is a triplet (P, Q, R) of a partial shape P, a full shape in a different pose Q and a ground truth completion R. The shapes Q and R are sampled from the same subject in two different poses. To receive P we render a depth map of R, at a viewpoint of zero elevation and a random azimuth angle in the range $0°$ and $360°$. These projections approximate the typical partiality pattern of depth sensors. Note that despite the large missing region, these projections largely retain the pose, making the reconstruction task well-defined. We also analysed different types of projections, such as projections from different elevation angles. This analysis is provided in the supplementary. The training examples $(P_n, Q_n, R_n)_{n=1}^N$ were provided in batches to the Siamese Network, where N is the size of the train set. Each input pair is fed to the encoder to receive the latent code $\theta(P_n, Q_n)$ and the reconstruction $F_{\theta(P_n,Q_n)}(Q_n)$ is determined by the generator. This reconstruction is subsequently compared against the ground-truth reconstruction R_n using the loss in Eq. (1).

3.6 Implementation considerations

Our implementation is available at `https://github.com/OshriHalimi/precise_shape_completion`. The network was trained using the PyTorch [52] ADAM optimizer with a learning rate of 0.001 and a momentum of 0.9. Each training batch contained 10 triplet examples (P, Q, R). The network was trained for 50 epochs, each containing 1000 batches. The input shapes, Q and R, are were centered, such that their center of mass lies at the origin.

Surace Normals In practice we found it helpful to include normal vectors as additional input features, making each input point 6D. The normal vector field is especially helpful for disambiguating contact points of the surface allowing prevention of contradicting requirements of the estimated deformation function. The input normals were computed using the connectivity for the mesh inputs, and approximated using Hoppe's method [31] for point clouds, as described in the experimental section. We note that at training, we always had access to the mesh connectivity and Hoppe's method was only applied on real scans, at inference time. Additionally, in the loss evaluation, we found that by considering also surface normals, in addition to point coordinates, fine details are better preserved. Therefore, in equation (1), we defined r_i as the concatenation of the coordinates and unit normal vector at each point: $(\vec{x}_{ri}, \alpha\vec{n}_{ri}) \in \mathbb{R}^6$. We used a scale factor of $\alpha = 0.1$, for the normal vector. To conclude, we used the surface normals in two places: (A) as additional channels for the input shapes, and (B) in the loss definition. To quantify the contribution of each design choice we ran all 4 configurations on FAUST dataset [6]. The relative improvement w.r.t not using normals at all is as follows: A+\B-: 4.6%; A-\B+: -3.3%; A+\B+ (as in the paper): 13%. Our experiments indicate that setting A+ is consistently helpful in disambiguating contact points, and that the chosen setting A+\B+, is the best performing.

ICP Refinement Empirically, the network reconstruction is often slightly shifted from the source partial scan. To recover the partial correspondence via a nearest neighbor query, it is crucial that the alignment be as exact as possible, and

therefore we apply a rigid Iterative Closest Point algorithm [4], as refinement, choosing the moving input as the partial shape, and the fixed input as the network reconstruction. Since the initial alignment is already adequate, this step is both stable and fast.

Activation Function Studying the displacement field statistics between all pose pairs in our training datasets, we observed that the maximal coordinate displacement is bounded by 1.804m, and relatively symmetric. Accordingly, in the generator module, we used the activation 2*tanh(x) - a symmetric function, bounded in the range [-2,2], akin to [26].

4 Experiments

The proposed method tackles two important tasks in nonrigid shape analysis: shape completion and partial shape matching. We emphasize the graceful handling of severe partiality resulting from range scans. In contrast, prior efforts either addressed one of these tasks or attempted to address both at mild partiality conditions. Here, we describe the different datasets used and then evaluate our method on both tasks. Finally, we show performance on real scanned data.

4.1 Datasets

We utilize two datasets of human shapes for training and evaluation, FAUST [6] and AMASS [46]. In addition we use raw scans from Dynamic FAUST [7] for testing purposes only. FAUST was generated by fitting SMPL parametric body model [44] to raw scans. It is a relatively small set of 10 subjects posing at 10 poses each. Following training and evaluation protocols from previous works (e.g. [41]), we kept the same train/test split, and for each of these sets, we generated 10 projected views per model, using pyRender [33]. AMASS, on the other hand, is currently the largest and most diverse dataset of human shapes designed specifically for deep learning applications. It was generated by unifying 15 archived datasets of marker-based optical motion capture (mocap) data. Each mocap sequence was converted to a sequence of rigged meshes using SMPL+H model [61]. Consequently, AMASS provides a richer resource for evaluating generalization. We generated a large set of single-view projections by sampling every 100th frame of all provided sequences. We then used pyRender [33] to render each shape from 10 equally spaced azimuth angles, keeping elevation at zero. Keeping the data splits prescribed by [46], our dataset comprises a total of $110K$, $10K$, and $1K$ full shapes for train, validation and test, respectively; and 10 times that in partial shapes. Note that at train time we randomly mix and match full shapes and their parts which drastically increases the effective set size.

4.2 Methods in comparison

The problem of deformable shape completion was recently studied by Litany *et al.* [39]. In their work, completion is achieved via optimization in a learned shape

space. Different from us, their task is completion from a partial view without explicit access to a full model. This is an important distinction as it means missing parts can only be hallucinated. In contrast, we assume the shape details are provided but are not in the correct pose. Moreover, their solution requires a preliminary step of solving partial matching to a template model, which by itself is a hard problem. Here, we solve for it jointly with the alignment. The optimization at inference time also makes their solution quite slow. Instead we output a result in a single feed forward fashion. 3D-CODED [26] performs template alignment to an input shape in two stages: fast inference and slow refinement. It is designed for inputs which are either full or has mild partiality. Here we evaluate the performance of their network predictions under significant partiality. In the refinement step we use directional Chamfer distance, as suggested by the authors in the partial case. FARM [47] is another alignment-based solution that has shown impressive results on shape completion and dense correspondences. It builds on the SMPL [44] human body model due to its compact parameterization, yet, we found it to be very slow to converge (up to 30 min for a single shape) and prone to getting trapped in local minima. We also tried to compare with a recent nonrigid registration method [45] that aligns a given full source point cloud to a partial target point cloud. However, this method didn't converge on our moderate size point clouds (< 7000 vertices) even within 48 hours, therefore we do not report on this method. 3D-EPN [20] is a rigid shape completion method. Based on a 3D-CNN, it accepts a voxelized signed distance field as input, and outputs that of a completed shape. Results are then converted to a mesh via computation of an isosurface. Comparison with classic Poisson reconstruction [36] is also provided. It serves as a naïve baseline as it has access only to the partial input. Lacking a single good measure of completion quality, we provide 5 different ones (see tables 1 and 2). Each measurement highlights a different aspect of the predicted completion. We report the root mean square error (RMSE) of the Euclidean distance between each point on the reconstructed shape and its ground truth mapping. We report this measure for predictions with well defined correspondence to the true reconstruction. We also report the RMSE of two directional Chamfer distances: ground-truth to prediction, and vice versa. The former measures coverage of the target shape by the prediction and the later penalizes prediction outliers. We report the sum of both as full the Chamfer distance. Finally, we report volumetric error as the absolute volume difference divided by the ground truth volume. Please note that the results reported in [39] as "Euclidean distance error" are reported differently in our Table 1 and 2. We confirm that the column named "Euclidean Distance Error" in [39] is, in fact, a directional Chamfer distance from GT to reconstruction. We, therefore, reported that error in the appropriate column and added a computation of the Euclidean distance.

4.3 Single view completion

We evaluate our method on the task of deformable shape completion on FAUST and AMASS.

	Euclidean distance	Volumetric err.	Chamfer GT → Recon.	Chamfer Recon. → GT	Full Chamfer
Poisson [36]	–	24.8 ± 23.2	7.3	3.64	10.94
3D-EPN [20]	–	89.7 ± 33.8	4.52	4.87	9.39
3D-CODED [26]	35.50	21.8 ± 0.3	11.15	38.49	49.64
FARM [47]	35.77	43.08 ± 20.4	9.5	3.9	13.4
Litany *et al.* [39]	7.07	9.24 ± 8.62	2.84	2.9	5.74
Ours	**2.94**	$\mathbf{7.05 \pm 3.45}$	**2.42**	**1.95**	**4.37**

Table 1. FAUST Shape Completion. Comparison of different methods with respect to errors in vertex position and shape volume.

FAUST projections We follow the evaluation protocol proposed in [39] and summarize the completion results of our method and prior art in Table 1. As can be seen, our network generates a much more accurate completion. Contrary to optimization-based methods [39,26,47] which are very slow at inference, our feed-forward network performs inference in less than a second. To better appreciate the quality of our reconstructions, in Figure 4 we visualize completions predicted by various methods. Note how our method accurately preserves fine details that were lost in previous methods. In the supplementary, we analyse the reconstruction error as a function of proximity between the source and the target pose, as well as provide additional completion results.

AMASS projections Using our test set of partial shapes from AMASS (generated as described in 4.1), we compare our method with two recent methods based on shape alignment: 3D-CODED [26], and FARM [47]. As described in 4.2, 3D-CODED is a learning-based method that uses a fixed template and is not designed to handle severe partiality. FARM, on the other hand, is an optimization method built for the same setting as ours. We summarize the results in Table 2. As can be seen, our method outperforms the two baselines by a large margin in all reported metrics. Note that on some of the examples (about 30%) FARM crashed during the optimization. We therefore only report the errors on its successful runs. Visualizations of several completions are shown in Figure 3. Additional completions are visualized in the supplementary.

	Euclidean distance	Volumetric err.	Chamfer GT → Recon.	Chamfer Recon. → GT	Full Chamfer
3D-CODED [26]	36.14	–	13.65	35.35	49
FARM [47]	27.75	49.42 ± 29.12	11.17	5.14	16.31
Ours	**6.58**	$\mathbf{27.62 \pm 15.27}$	**4.86**	**3.06**	**7.92**

Table 2. AMASS Shape Completion. Comparison of different methods with respect to errors in vertex position and shape volume.

4.4 Non-rigid partial correspondences

Finding dense correspondences between a full shape and its deformed parts is still an active research topic. Here we propose a solution in the form of alignment

between the full shape and the partial shape, allowing for the recovery of the correspondence by a simple nearest neighbor search. As before, we evaluate this task on both FAUST and AMASS.

FAUST projections On the FAUST projections dataset, we compare with two alignment-based methods, FARM and 3D-CODED. We also compare with 3 methods designed to only recover correspondences, that is, without performing shape completion: MoNet [49], and two 3-layered Euclidean CNN baselines, trained on either SHOT [69] descriptors or depth maps. Results are reported in Figure 5. As in the single view completion experiment, the test set consists of 200 shapes: 2 subjects at 10 different poses with 10 projected views each. The direct matching baselines solve a labeling problem, assigning each input vertex a matching index in a fixed template shape. Differently, 3D-CODED deforms a fixed template and recovers correspondence by a nearest neighbor query for each input vertex using a one-sided Chamfer distance, as suggested in [26]. Our method and FARM both require a complete shape as input, which we chose as the null pose of each of the test examples. Due to slow convergence and unstable behavior of FARM we only kept 20 useful matching results on which we report the performance. As seen in Figure 5, our method outperforms prior art by a significant margin. This result is particularly interesting since it demonstrates that even though we solve an alignment problem, which is strictly harder than correspondence, we receive better results than methods that specialize in the latter. At the same time, looking at the poor performance demonstrated by the other alignment-based methods, we conclude that simply solving an alignment problem is not enough and the details of our method and training scheme allow for a substantial difference. Qualitative correspondence results are visualized in the supplementary.

AMASS Projections As FAUST is limited in variability, we further test our method on the recently published AMASS dataset. On the task of partial correspondence, we compare with FARM [47] and 3D-CODED [26] for which code was available online. We report the correspondence error graphs in Figure 6. For evaluation we used 200 pairs of partial and full shapes chosen randomly (but consistently between different methods). Specifically, for each of the 4 subjects in AMASS test set we randomized 50 pairs of full poses: one was taken as the full shape Q and one was projected to obtain the partial shape P, using the full unprojected version as the ground truth completion R. As with FAUST, we report the error curve of FARM taking the average of only the successful runs. As can be observed, our method outperforms both methods by a large margin. Qualitative correspondence results are visualized in the supplementary.

4.5 Real scans

To evaluate our method in real-world conditions, we test it on raw measurements taken during the preparation of the Dynamic FAUST [7] dataset. This use case nicely matches our setting: these are partial scans of a subject for which we have a complete reference shape at a different pose. As preprocessing we compute

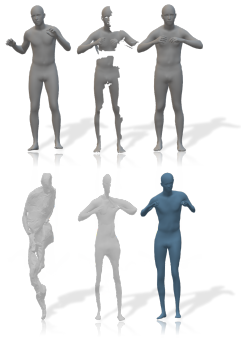

Fig. 3. AMASS Shape Completion. At the top from left to right: full shape Q, partial shape P, ground truth completion R. At the bottom from left to right: reconstructions of FARM [47], 3D-CODED [26] and ours.

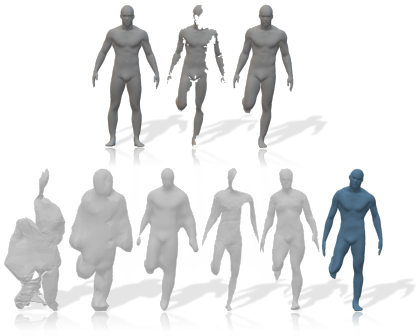

Fig. 4. FAUST Shape Completion. At the top from left to right: full shape Q, partial shape P, ground truth completion R. At the bottom from left to right: reconstructions from FARM [47], 3D-EPN [20], Poisson [36], 3D-CODED [26], Litany *et al* [39] and ours.

point normals for the input scan using the method presented in [31]. The point cloud and the reference shape are subsequently inserted into a network pretrained on FAUST. The template, raw scan, and our reconstruction are shown, from left to right, in Figure 7. We show our result both as the recovered point cloud as well as the recovered mesh using the template triangulation. As apparent from the figure, this is a challenging test case as it introduces several properties not seen at test time: a point cloud without connectivity leads to noisier normals, scanner noise, different point density and extreme partiality (note the missing bottom half of the shapes). Despite all these, the proposed network was able to recover the input quite elegantly, preserving shape details and mimicking the desired pose. In the rightmost column, we report a comparison with Litany *et al.* [39]. Note that while [39] was trained on Dynamic FAUST, our network was trained on FAUST which is severely constrained in its pose variability. The result highlights that our method captures appearance details while pose accuracy is limited by the variability of the training set.

5 Conclusions

We proposed an alignment-based solution to the problem of shape completion from range scans. Different from most previous works, we focus on the setting where a complete shape is given, but is at a different pose than that of the scan. Our data-driven solution is based on learning the space of distortions, linking scans at various poses to whole shapes in other poses. As a result, at test time we can accurately align unseen pairs of parts and whole shapes at different poses.

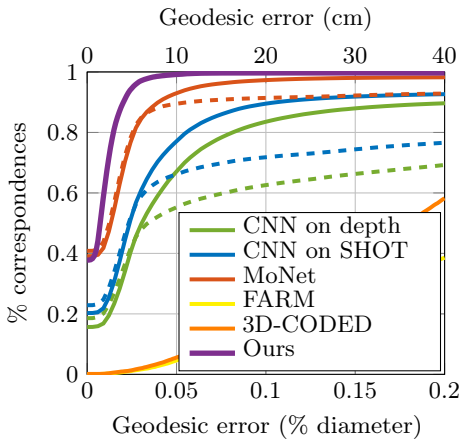

Fig. 5. Partial correspondence error, FAUST dataset. Same color dashed and solid lines indicate performance before and after refinement, respectively.

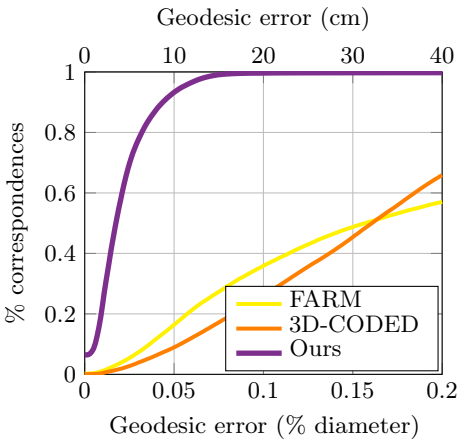

Fig. 6. Partial correspondence error, AMASS dataset.

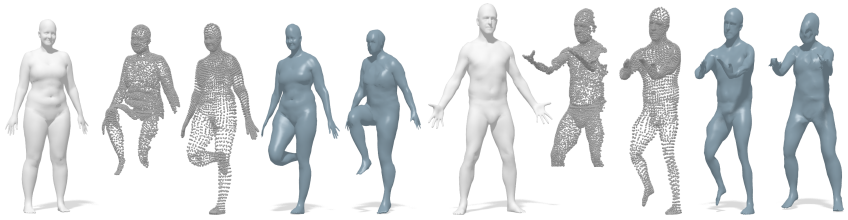

Fig. 7. Completion from real scans from the Dynamic Faust dataset [7]. From left to right: Input reference shape; input raw scan; our completed shape as a point cloud; and as mesh; completion from Litany *et al.* [39].

Acknowledgements. We gratefully thank Rotem Cohen for contributing to the article visualizations. This work was supported by the Israel Ministry of Science and Technology grant number 3-14719, the Technion Hiroshi Fujiwara Cyber Security Research Center and the Israel Cyber Directorate, the Vannevar Bush Faculty Fellowship, the SAIL-Toyota Center for AI Research, and by Amazon Web Services. Giovanni Trappolini and Emanuele Rodolà are supported by the ERC Starting Grant No. 802554 (SPECGEO) and the MIUR under grant "Dipartimenti di eccellenza 2018-2022" of the Department of Computer Science of Sapienza University.

References

1. Allen, B., Curless, B., Popović, Z., Hertzmann, A.: Learning a correlated model of identity and pose-dependent body shape variation for real-time synthesis. In: Proceedings of the 2006 ACM SIGGRAPH/Eurographics symposium on Computer animation. pp. 147–156. Eurographics Association (2006)
2. Anguelov, D., Srinivasan, P., Koller, D., Thrun, S., Rodgers, J., Davis, J.: Scape: shape completion and animation of people **24**(3), 408–416 (2005)
3. Ben-Shabat, Y., Lindenbaum, M., Fischer, A.: 3D point cloud classification and segmentation using 3D modified fisher vector representation for convolutional neural networks. arXiv preprint arXiv:1711.08241 (2017)
4. Besl, P.J., McKay, N.D.: Method for registration of 3-d shapes. In: Sensor fusion IV: control paradigms and data structures. vol. 1611, pp. 586–606. International Society for Optics and Photonics (1992)
5. Blanz, V., Vetter, T.: A morphable model for the synthesis of 3D faces. In: Proc. Computer Graphics and Interactive Techniques. pp. 187–194 (1999)
6. Bogo, F., Romero, J., Loper, M., Black, M.J.: FAUST: Dataset and Evaluation for 3d Mesh Registration. In: Proc. CVPR (2014)
7. Bogo, F., Romero, J., Pons-Moll, G., Black, M.J.: Dynamic FAUST: Registering human bodies in motion. In: IEEE Conf. on Computer Vision and Pattern Recognition (CVPR) (Jul 2017)
8. Boscaini, D., Masci, J., Rodolà, E., Bronstein, M.: Learning shape correspondence with anisotropic convolutional neural networks. In: Advances in Neural Information Processing Systems. pp. 3189–3197 (2016)
9. Bronstein, A.M., Bronstein, M.M., Bruckstein, A., Kimmel, R.: Matching two-dimensional articulated shapes using generalized multidimensional scaling. In: Proc. of Articulated Motion and Deformable Objects (AMDO) (2006)
10. Bronstein, A.M., Bronstein, M.M., Kimmel, R.: Expression-invariant 3d face recognition. In: Proc. Audio & Video-based Biometric Person Authentication (AVBPA), Lecture Notes in Comp. Science 2688, Springer (2003)
11. Bronstein, A.M., Bronstein, M.M., Kimmel, R.: Three-dimensional face recognition. International Journal of Computer Vision **64**(1), 5–30 (2005)
12. Bronstein, A.M., Bronstein, M.M., Kimmel, R.: Face2face: an isometric model for facial animation. In: Conf. on Articulated Motion and Deformable Objects (AMDO) (2006)
13. Bronstein, A.M., Bronstein, M.M., Kimmel, R.: Robust expression-invariant face recognition from partially missing data. In: Proc. ECCV, Graz, Austria (May 2006)
14. Bronstein, A.M., Bronstein, M.M., Kimmel, R.: Expression-invariant representations of faces. IEEE Trans. Image Processing **16**(1), 188–197 (2007)
15. Bronstein, A.M., Bronstein, M.M., Kimmel, R.: Generalized multidimensional scaling: a framework for isometry-invariant partial surface matching. PNAS **103**(5), 1168–1172 (2006)
16. Bronstein, M.M., Bruna, J., LeCun, Y., Szlam, A., Vandergheynst, P.: Geometric deep learning: going beyond euclidean data. IEEE Signal Processing Magazine **34**(4), 18–42 (2017)
17. Chen, Q., Koltun, V.: Robust nonrigid registration by convex optimization. In: Proc. ICCV (2015)
18. Choy, C., Gwak, J., Savarese, S.: 4d spatio-temporal convnets: Minkowski convolutional neural networks. arXiv preprint arXiv:1904.08755 (2019)

19. Cosmo, L., Panine, M., Rampini, A., Ovsjanikov, M., Bronstein, M.M., Rodolà, E.: Isospectralization, or how to hear shape, style, and correspondence. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 7529–7538 (2019)
20. Dai, A., Qi, C.R., Nießner, M.: Shape completion using 3D-encoder-predictor cnns and shape synthesis. arXiv:1612.00101 (2016)
21. Devir, Y., Rosman, G., Bronstein, A. M. Bronstein, M.M., Kimmel, R.: On reconstruction of non-rigid shapes with intrinsic regularization. In: Proc. of Workshop on Nonrigid Shape Analysis and Deformable Image Alignment (NORDIA) (2009)
22. Elad, A., Kimmel, R.: Bending invariant representations for surfaces. In: Proc. of CVPR'01, Hawaii (December 2001)
23. Elad, A., Kimmel, R.: On bending invariant signatures for surfaces. IEEE Trans. on Pattern Analysis and Machine Intelligence (PAMI) **25**(10), 1285–1295 (2003)
24. Gerig, T., Morel-Forster, A., Blumer, C., Egger, B., Lüthi, M., Schönborn, S., Vetter, T.: Morphable face models-an open framework. arXiv preprint arXiv:1709.08398 (2017)
25. Graham, B., Engelcke, M., van der Maaten, L.: 3d semantic segmentation with submanifold sparse convolutional networks. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 9224–9232 (2018)
26. Groueix, T., Fisher, M., Kim, V.G., Russell, B.C., Aubry, M.: 3D-coded: 3D correspondences by deep deformation. In: Proceedings of the European Conference on Computer Vision (ECCV). pp. 230–246 (2018)
27. Groueix, T., Fisher, M., Kim, V.G., Russell, B.C., Aubry, M.: Atlasnet: A papier-Mâchè approach to learning 3D surface generation. arXiv preprint arXiv:1802.05384 (2018)
28. Guler, R.A., Kokkinos, I.: Holopose: Holistic 3d human reconstruction in-the-wild. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 10884–10894 (2019)
29. Halimi, O., Litany, O., Rodolà, E., Bronstein, A.M., Kimmel, R.: Unsupervised learning of dense shape correspondence. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 4370–4379 (2019)
30. Hanocka, R., Hertz, A., Fish, N., Giryes, R., Fleishman, S., Cohen-Or, D.: Meshcnn: A network with an edge. ACM Transactions on Graphics (TOG) **38**(4), 90 (2019)
31. Hoppe, H., DeRose, T., Duchamp, T., McDonald, J., Stuetzle, W.: Surface reconstruction from unorganized points, vol. 26. ACM (1992)
32. Hu, T., Han, Z., Shrivastava, A., Zwicker, M.: Render4completion: Synthesizing multi-view depth maps for 3d shape completion. In: Proceedings of the IEEE International Conference on Computer Vision Workshops. pp. 0–0 (2019)
33. Huang, J., Zhou, Y., Funkhouser, T., Guibas, L.: Framenet: Learning local canonical frames of 3d surfaces from a single rgb image. arXiv preprint arXiv:1903.12305 (2019)
34. Jiang, H., Cai, J., Zheng, J.: Skeleton-aware 3d human shape reconstruction from point clouds. In: Proceedings of the IEEE International Conference on Computer Vision. pp. 5431–5441 (2019)
35. Joseph-Rivlin, M., Zvirin, A., Kimmel, R.: Mom^{e}Net: Flavor the moments in learning to classify shapes. In: Proc. of IEEE Int. Conference on Computer Vision (CVPR) Workshops (2019)
36. Kazhdan, M., Hoppe, H.: Screened Poisson surface reconstruction. TOG **32**(3), 29 (2013)
37. Kim, V.G., Lipman, Y., Funkhouser, T.A.: Blended intrinsic maps. Trans. Graphics **30**(4) (2011)

38. Korman, S., Ofek, E., Avidan, S.: Peeking template matching for depth extension. In: Proc. CVPR (2015)
39. Litany, O., Bronstein, A., Bronstein, M., Makadia, A.: Deformable shape completion with graph convolutional autoencoders. CVPR (2018)
40. Litany, O., Remez, T., Bronstein, A.: Cloud dictionary: Sparse coding and modeling for point clouds. arXiv:1612.04956 (2016)
41. Litany, O., Remez, T., Rodolà, E., Bronstein, A.M., Bronstein, M.M.: Deep functional maps: Structured prediction for dense shape correspondence. In: Proc. ICCV. vol. 2, p. 8 (2017)
42. Litany, O., Rodolà, E., Bronstein, A.M., Bronstein, M.M.: Fully spectral partial shape matching. Computer Graphics Forum **36**(2), 247–258 (2017)
43. Liu, X., Yan, M., Bohg, J.: Meteornet: Deep learning on dynamic 3d point cloud sequences. In: Proceedings of the IEEE International Conference on Computer Vision. pp. 9246–9255 (2019)
44. Loper, M., Mahmood, N., Romero, J., Pons-Moll, G., Black, M.J.: Smpl: A skinned multi-person linear model. ACM transactions on graphics (TOG) **34**(6), 248 (2015)
45. Ma, J., Wu, J., Zhao, J., Jiang, J., Zhou, H., Sheng, Q.Z.: Nonrigid point set registration with robust transformation learning under manifold regularization. IEEE transactions on neural networks and learning systems **30**(12), 3584–3597 (2018)
46. Mahmood, N., Ghorbani, N., Troje, N.F., Pons-Moll, G., Black, M.J.: Amass: Archive of motion capture as surface shapes. In: The IEEE International Conference on Computer Vision (ICCV) (Oct 2019), `https://amass.is.tue.mpg.de`
47. Marin, R., Melzi, S., Rodolà, E., Castellani, U.: Farm: Functional automatic registration method for 3d human bodies. In: Computer Graphics Forum. Wiley Online Library (2018)
48. 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)
49. Monti, F., Boscaini, D., Masci, J., Rodolà, E., Svoboda, J., Bronstein, M.M.: Geometric deep learning on graphs and manifolds using mixture model cnns. In: Computer Vision and Pattern Recognition (CVPR), 2017 IEEE Conference on. pp. 5425–5434. IEEE (2017)
50. Newcombe, R.A., Fox, D., Seitz, S.M.: Dynamicfusion: Reconstruction and tracking of non-rigid scenes in real-time. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 343–352 (2015)
51. Park, J.J., Florence, P., Straub, J., Newcombe, R., Lovegrove, S.: Deepsdf: Learning continuous signed distance functions for shape representation. arXiv preprint arXiv:1901.05103 (2019)
52. Paszke, A., Gross, S., Chintala, S., Chanan, G., Yang, E., DeVito, Z., Lin, Z., Desmaison, A., Antiga, L., Lerer, A.: Automatic differentiation in pytorch (2017)
53. Qi, C.R., Litany, O., He, K., Guibas, L.J.: Deep hough voting for 3d object detection in point clouds. arXiv preprint arXiv:1904.09664 (2019)
54. Qi, C.R., Su, H., Mo, K., Guibas, L.J.: Pointnet: Deep learning on point sets for 3d classification and segmentation. In: Proc. CVPR (2017)
55. Qi, C.R., Yi, L., Su, H., Guibas, L.J.: Pointnet++: Deep hierarchical feature learning on point sets in a metric space. arXiv:1706.02413 (2017)
56. Rampini, A., Tallini, I., Ovsjanikov, M., Bronstein, A.M., Rodolà, E.: Correspondence-free region localization for partial shape similarity via hamiltonian spectrum alignment. arXiv preprint arXiv:1906.06226 (2019)

57. Richardson, E., Sela, M., Or-El, R., Ron, K.: Learning detailed face reconstruction from a single image. In: IEEE Conf. on Computer Vision and Pattern Recognition (CVPR), Hawaii, Honolulu (2017)
58. Richardson, E., Sela, M., Ron, K.: 3D face reconstruction by learning from synthetic data. In: 4th Int. Conf. on 3D Vision (3DV) Stanford University, CA, USA (2016)
59. Rodolà, E., Cosmo, L., Bronstein, M.M., Torsello, A., Cremers, D.: Partial functional correspondence **36**(1), 222–236 (2017)
60. Rodolà, E., Rota Bulo, S., Windheuser, T., Vestner, M., Cremers, D.: Dense non-rigid shape correspondence using random forests. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 4177–4184 (2014)
61. Romero, J., Tzionas, D., Black, M.J.: Embodied hands: Modeling and capturing hands and bodies together. ACM Transactions on Graphics, (Proc. SIGGRAPH Asia) **36**(6) (Nov 2017)
62. Sarkar, K., Varanasi, K., Stricker, D.: Learning quadrangulated patches for 3D shape parameterization and completion. arXiv:1709.06868 (2017)
63. Sarmad, M., Lee, H.J., Kim, Y.M.: Rl-gan-net: A reinforcement learning agent controlled gan network for real-time point cloud shape completion. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 5898–5907 (2019)
64. Sela, M., Richardson, E., Kimmel, R.: Unrestricted facial geometry reconstruction using image-to-image translation. In: Int. Conf. Comp. Vision (ICCV), Venice, Italy (2017)
65. Shtern, A., Sela, M., Kimmel, R.: Fast blended transformations for partial shape registration. Journal of Mathematical Imaging and Vision **60**(6), 913–928 (2018)
66. Slavcheva, M., Baust, M., Cremers, D., Ilic, S.: Killingfusion: Non-rigid 3d reconstruction without correspondences. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 1386–1395 (2017)
67. Stutz, D., Geiger, A.: Learning 3d shape completion from laser scan data with weak supervision. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 1955–1964 (2018)
68. Su, H., Maji, S., Kalogerakis, E., Learned-Miller, E.: Multi-view convolutional neural networks for 3d shape recognition. In: Proc. CVPR (2015)
69. Tombari, F., Salti, S., Di Stefano, L.: Unique signatures of histograms for local surface description. In: International Conference on Computer Vision (ICCV). pp. 356–369 (2010)
70. Varol, G., Ceylan, D., Russell, B., Yang, J., Yumer, E., Laptev, I., Schmid, C.: Bodynet: Volumetric inference of 3d human body shapes. In: Proceedings of the European Conference on Computer Vision (ECCV). pp. 20–36 (2018)
71. Varol, G., Romero, J., Martin, X., Mahmood, N., Black, M.J., Laptev, I., Schmid, C.: Learning from synthetic humans. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 109–117 (2017)
72. Verma, N., Boyer, E., Verbeek, J.: Dynamic filters in graph convolutional networks. arXiv:1706.05206 (2017), `http://arxiv.org/abs/1706.05206`
73. Wang, Y., Sun, Y., Liu, Z., Sarma, S.E., Bronstein, M.M., Solomon, J.M.: Dynamic graph cnn for learning on point clouds. ACM Transactions on Graphics (TOG) **38**(5), 146 (2019)
74. Wei, L., Huang, Q., Ceylan, D., Vouga, E., Li, H.: Dense human body correspondences using convolutional networks. In: Proc. CVPR (2016)
75. Wu, J., Zhang, C., Zhang, X., Zhang, Z., Freeman, W.T., Tenenbaum, J.B.: Learning shape priors for single-view 3d completion and reconstruction. In: Proceedings of the European Conference on Computer Vision (ECCV). pp. 646–662 (2018)

76. Wu, Z., Song, S., Khosla, A., Yu, F., Zhang, L., Tang, X., Xiao, J.: 3D shapenets: A deep representation for volumetric shapes. In: Proc. CVPR (2015)
77. Xu, D., Anguelov, D., Jain, A.: Pointfusion: Deep sensor fusion for 3d bounding box estimation. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 244–253 (2018)
78. Zanfir, A., Marinoiu, E., Sminchisescu, C.: Monocular 3d pose and shape estimation of multiple people in natural scenes-the importance of multiple scene constraints. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 2148–2157 (2018)