Abstract:

1. Missing abstract

Introduction:

2. The challenges are not clearly stated in the second paragraph but are referred to in the third paragraph. Please revise.
3. The third paragraph talks about the ideas that address the challenges. Highlight the novelty of the ideas.
4. The use of words must be careful. Leveraging is often not novel in computer science publications. In an application paper, leveraging a new ML/CV technique to solve a domain-specific problem is sometimes considered novel. Again, revise the 3rd paragraph to highlight novelty.

Related work:

5. Since it is unclear what challenges (or research problems) are to be addressed in this paper, it is difficult to say if the related work is on target or not. After revising the introduction, make sure to reorganize the discussion around the problems stated in the introduction.
6. The last paragraph in this section needs revision. It needs to confirm and restate the problems stated in the introduction. In this version, it lists three open challenges: a) performance degradation in the presence of self-occlusions and incomplete observations, b) overfitting or failing to generalize (this point is not clearly stated and needs revision), and c) balancing between preservation of local geometric structures and efficiency.

Method:

7. Replace the titles of sections 3.2 and 3.3 according to the constraints they impose on the network optimization.

Results:

8. Reduce sections 4.1 and 4.2 to about half of their current length. Given the 8-page limit, the description of data and metrics is too long.
9. 4.3.1 and 4.3.2 can be merged.
10. 4.5 needs to be extended. This seems to be a major issue addressed in this paper. Also, this is similar to the partial point set problem. Consider merging sections 4.5 and 4.3.
11. Local structure preservation can be evaluated in the context of large deformation. Also, it seems that a balance of local structure and efficiency is needed.

Conclusion:

12. Again, given the 8-page limit, the conclusion section is too long. Consider reducing it to one paragraph of about 300 words.

UND-Net: Unsupervised Neural Deformation Network for Non-Rigid Registration of Articulated Human poses

Anonymous WACV Applications Track submission

Paper ID *****

1. Introduction

Articulated human pose registration is a fundamental task in computer vision, with applications in motion capture, virtual avatars, human-robot interaction, and medical analysis. Unlike generic non-rigid registration, human pose registration involves aligning complex articulated structures that exhibit large inter-limb motions, body-part specific deformations, and frequent self-occlusions. In dynamic capture scenarios, self-occlusion occurs when limbs occlude other parts, leading to incomplete, spatially disjoint point clouds [17, 27]. These occlusions, combined with viewpoint limitations and sensor noise, result in sparse partial overlaps between the two point clouds to be registered, making the establishment of dense point-to-point correspondences infeasible [18, 29].

Recent progress in non-rigid registration has spanned both optimization-based [1, 16, 23] and learning-based approaches [9, 22], many of which are not specifically tailored to the articulated nature of human bodies. Feature-based registration [2, 5] and global correspondence models [15, 32] often rely on stable keypoints or full surface coverage. Unfortunately, such assumptions do not hold in scenarios involving partial visibility or rapid human body motion, where significant occlusions and large deformations are common. Neural deformation field models [26] enable fine-grained alignment but commonly require dense supervision [8, 11], keypoint priors [24], or complete surface visibility [12]. In contrast, unsupervised approaches [19, 31] reduce dependence on supervision but still suffer from ambiguity in occluded regions and inter-limb correspondence errors.

To address these challenges, we propose an unsupervised neural deformation network explicitly designed for articulated human shapes. Our method learns a continuous deformation field, parameterized over spatial coordinates, to predict plausible non-linear body deformations in the absence of complete 3D observations. As illustrated in Fig. 1, the model adapts to partial visibility and motion-induced occlusion by leveraging local shape consistency and reciprocal alignment constraints, which preserve geometric fidelity even under severe articulation.

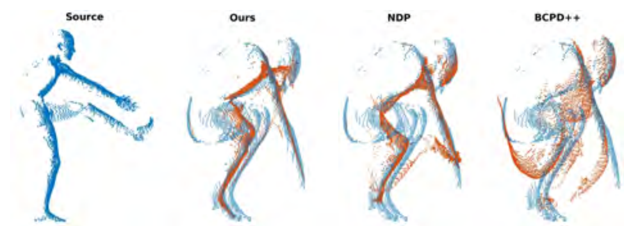

Figure 1. Example of registration of a complex 3D Human pose with Self-Occlusion

The rest of this paper is organized as follows: Sec. 2 reviews the non-rigid point cloud registration methods that address self-occlusion and noise issues. Sec. 3 presents our proposed method. Sec. 4 discusses the experimental results, including a comparison study and analysis of performance under self-occlusion, local deformation, and computational efficiency. Sec. 5 concludes this paper with a summary.

2. Related Work

Non-rigid point cloud registration has witnessed a surge of interest in learning-based methods that enable dense, correspondence-free registration through implicit field representations. Li and Harada [10] introduced the Neural Deformation Pyramid (NDP), a learning-based method that addresses deformations in a coarse-to-fine manner using a hierarchy of MLPs. The use of sinusoidal positional encoding allows the network to represent fine-grained deformations. However, NDP assumes that both source and target point clouds are relatively complete and well-overlapped. Liu et al. [13] proposed a learning-based deformation prior for the partial 3D human body, which employs a generative network to predict plausible completions and improve alignment under occlusion. Zhang et al. [28] introduced PartFusion, which addresses partial occlusions by learning part-aware deformation modules and integrating them with fusion guided by visible regions. To improve structure preservation during deformation, Ma et al. [14] proposed a structure-aware neural registration framework regularized with kinematic chains during deformation prediction. Xu et al. [25] presented a transformer-based architecture, DeformerNet++, for modeling long-range dependencies across partially visible surfaces.

In the context of occlusion-aware registration, Sun et al. [20] introduced PartialAligner, an unsupervised framework based on contrastive feature distillation, which guides the model to align mutually visible surfaces while ignoring spurious mismatches in occluded areas. Tang et al. [21] proposed OccReg, which jointly learns occlusion masks and deformation fields through a visibility-aware training pipeline. Chen et al. [4] developed a Nystrom-based deformation network that approximates dense displacement fields by sampling geodesic anchors.

Zhao et al. [30] proposed an unsupervised method that treats the source point set as a collection of centroids and the target as a set of cluster members. A soft clustering is optimized with l_1-norm regularization on a Laplacian kernel, ensuring smoothness and structural consistency. Furthermore, to address computational challenges, a clustering-enhanced Nyström method is used to approximate kernel matrices. Zhao et al. [31] introduced a method to learn a continuous deformation field, using unsupervised optimization guided by the Maximum Correntropy Criterion. The correntropy loss downweights high-error correspondences caused by occlusions, allowing the network to focus on deforming the visible regions. Additionally, a Locally Linear Reconstruction regularization is employed to enforce geometric coherence across neighboring points.

Despite significant progress, several open challenges persist in non-rigid point cloud registration. Existing methods have demonstrated impressive results on clean, complete, or synthetic data. However, their performance degrades in the presence of self-occlusions and incomplete observations, which are common in real-world, single-view depth sensing scenarios. Many neural deformation models either overfit to visible regions or fail to generalize without dense supervision or mesh-based priors. Furthermore, preserving local geometric structure during deformation while maintaining computational efficiency remains difficult, particularly for high-resolution human point clouds.

3. Unsupervised Neural Deformation Network

Let $X = \{x_i\}_{i=1}^{N} \subset \mathbb{R}^3$ denote the source point cloud and $Y = \{y_j\}_{j=1}^{M} \subset \mathbb{R}^3$ the target point cloud. We formulate non-rigid registration as the problem of learning a continuous deformation field $f_\theta : \mathbb{R}^3 \rightarrow \mathbb{R}^3$ that maps each source point x_i to a new position $x'_i = x_i + f_\theta(x_i)$ such that the deformed set $X' = \{x'_i\}$ aligns with the target point cloud Y.

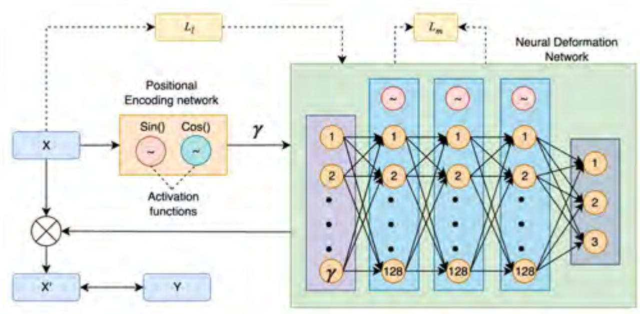

Figure 2. Network architecture of UND-Net.

Fig. 2 illustrates our network architecture that comprises three stages: (1) positional encoding of the input source point cloud, (2) neural deformation, and (3) deformation regularization through two loss terms: a correspondence free alignment loss using Maximum Correntropy Criterion (MCC) and a geometric structure preserving loss via Local Linear Reconstruction (LLR).

3.1. Positional Encoding

At the core of our registration framework lies a deformation network that predicts dense 3D displacements for each point in the source point cloud. The input to the deformation network is a single 3D point $x_i \in \mathbb{R}^3$. We map it into a higher-dimensional spectral domain using a learnable sinusoidal basis. Specifically, we apply a linear projection $B \in \mathbb{R}^{F \times 3}$ to the point, followed by sinusoidal nonlinearities, forming the positional encoding as follows:

$$\gamma(x_i) = [sin(2\pi B x_i), cos(2\pi B x_i)] \in \mathbb{R}^{2F}. \quad (1)$$

The encoded vector $\gamma(x_i) \in \mathbb{R}^{2F}$ consists of frequency band captures different levels of spatial detail. The matrix B is learned jointly with the deformation network during optimization. This enables the model to adapt its spectral decomposition to the underlying shape distribution and deformation complexity present in each registration instance.

$\gamma(x_i)$ is propagated through a series of fully connected layers, each followed by a sine activation function. The first layer employs a higher frequency coefficient ω_0, while the subsequent hidden layers use a shared frequency parameter ω. The forward computation of PE-SIREN is

$$h_1 = \sin(\omega_0 W_1 \gamma(x_i) + b_i), \quad (2)$$

$$h_l = \sin(\omega W_l h_{l-1} + b_l), \quad l = 2, \dots, L, \quad (3)$$

$$f_\theta(x_i) = W_{L+1} h_L + b_{L+1}. \quad (4)$$

The output $f_\theta(x_i) \in \mathbb{R}^3$ represents a displacement vector that is added to the original point, resulting in the deformed position $x'_i = x_i + f_\theta(x_i)$. The use of sinusoidal activations allows the network to capture high-frequency detail, which is particularly important in registering non-rigid shapes with fine articulations under occlusion.

Unlike the existing deformation models, our PE-SIREN is lightweight and highly expressive without relying on a large number of parameters. The deformation network acts as a function approximator that adapts to the geometry and visibility of shape pairs. It learns an instance-specific warping that aligns the source to the observed target shape with precision.

3.2. Local Linear Reconstruction Loss

While our deformation network can model arbitrary non-rigid transformations, unconstrained learning in the presence of occlusions or missing data can lead to degenerate solutions such as fold-overs, local distortions, or collapsed regions. We incorporate a Local Linear Reconstruction (LLR) loss to prevent such artifacts and maintain the geometric plausibility of the deformed surface.

This loss enforces that local neighborhoods in the source point cloud remain approximately consistent after deformation, effectively acting as a shape-preserving regularizer. Each point is assumed to lie approximately in the affine span of its nearest neighbors. For each point $x_i \in X$, we identify its k-nearest neighbors $\{x_{ij}\}_{j=1}^{k}$ using a fixed-radius or k-d tree-based search. We then solve for a set of barycentric weights $w_i = [w_{i1}, \ldots, w_{ik}]$ that satisfy

$$x_i \approx \sum_{j=1}^{k} w_{ij} x_{ij}, \tag{5}$$

where $\sum_{j=1}^{k} w_{ij} = 1$. These weights are obtained by minimizing the following reconstruction error

$$\| x_i - \sum_{j=1}^{k} w_{ij} x_{ij} \|_2^2 + \lambda_{reg} Tr(\mathbf{G}_i) \tag{6}$$

where $\mathbf{G}_i = (X_i - x_i)^T (X_i - x_i)$ is the local Gram matrix, and λ_{reg} is a coefficient added to the trace to ensure numerical stability. Once computed, these barycentric weights capture the intrinsic local geometry of the source surface around each point. To enforce deformation coherence, we reuse the weights to reconstruct the deformed x'_i from the deformed neighbors $\{x'_{ij}\}$, where $x'_{ij} = x_{ij} + f_\theta(x_{ij})$. The expected position after deformation becomes

$$\hat{x}'_i = \sum_{j=1}^{k} w_{ij} \hat{x}'_{ij}. \tag{7}$$

The LLR loss penalizes the deviation between this predicted position and the actual deformed point

$$\begin{aligned} L_l &= \sum_{i=1}^{N} \| x'_i - \hat{x}'_i \|_2 \qquad (8) \\ &= \sum_{i=1}^{N} \| x_i + f_\theta(x_i) - \sum_{j=1}^{k} w_{ij} \cdot (x_{i_j} + f_\theta(x_{i_j})) \|_2 \, . \end{aligned}$$

This loss encourages the deformation field to be locally smooth and shape-consistent, even in regions where the target point cloud provides no direct guidance. Unlike global rigidity constraints, LLR enforces preservation of local affine structure, which is well-suited to capturing articulations and non-isometric motions of humans. By coupling LLR with the deformation network, we ensure that even the unconstrained, self-occluded parts of the source are warped plausibly, maintaining consistent edge lengths and curvature distributions across the surface. This enhances the robustness of our model under extreme occlusion.

In our implementation, we compute LLR loss on GPUs. All local linear systems for barycentric weight computation are batched and solved using PyTorch's batched matrix operations. This enables us to recompute neighbors, Gram matrices, and reconstructions at every training step without performance bottlenecks. The neighbor graph is built once per iteration using either GPU-accelerated kNN or efficient spatial hash structures. The LLR regularization thus scales efficiently with point cloud size and remains differentiable and stable throughout optimization.

3.3. Maximum Correntropy Criterion Loss

A key challenge in non-rigid registration under self-occlusion is the presence of large, unmatched regions between the source and target point clouds. We adopt a correntropy-based formulation that incorporates robust statistical matching by selectively downweighting unreliable correspondences. For each deformed point x'_i, we compute the squared distance to its nearest neighbor in Y

$$d_i = \min_j \| x'_i - y_j \|^2 \, . \tag{9}$$

Likewise, for each y_j, we compute,

$$\tilde{d}_j = \min_i \| y_j - x'_i \|^2 \, . \tag{10}$$

Each distance term is reweighted using a Gaussian kernel, yielding the MCC loss

$$L_m = -\sum_{i=1}^{N} \exp\left(-\frac{d_i}{\sigma^2}\right) - \sum_{j=1}^{M} \exp\left(-\frac{\tilde{d}_i}{\sigma^2}\right), \tag{11}$$

where σ^2 controls the sensitivity of the kernel. This probabilistic weighting ensures that well-aligned point pairs contribute strongly to the loss, while occluded regions or distant mismatches contribute minimally. Unlike hard correspondence-based metrics, our method is unsupervised and resilient to partial visibility, making it well-suited for scenarios where the source and target differ significantly due to occlusion.

To mitigate misleading gradients from severely mismatched regions, we employ a distance-based truncation strategy. Specifically, all distances d_i and $\tilde{d}_j$ exceeding r^2 are discarded by setting them to zero before applying the kernel. This serves as a hard geometric gate that focuses the optimization on overlapping, well-matched regions and improves stability.

The nearest neighbor computations required for d_i and $\tilde{d}_i$ are implemented using pytorch3d's GPU-accelerated kNN, enabling scalability to point clouds with thousands of points during training. When combined with LLR loss, the MCC loss guides the deformation network towards learning data-consistent and geometrically plausible deformations, even under challenging occlusions or missing data conditions.

The loss function of our method is a weighted sum of LLR and MCC as follows:

$$L = \alpha_l L_l + \alpha_m L_m. \tag{12}$$

4. Results and Discussion

4.1. Datasets and Settings

We conduct experiments using a publicly available MPI-FAUST dataset [3]. MPI-FAUST contains high-resolution 3D scans of 10 subjects captured in various articulated poses. Each point cloud consists of over 170,000 points approximately. Each point cloud is normalized to zero-mean and unit-variance before the registration. This process ensures numerical stability during training and promotes scale-invariance, which is critical for learning shape deformations. After deformation, the registered point cloud is scaled back to its original size.

In our experiments, we select 50 source-target pairs representing significant pose variations, including inter-subject and intra-subject combinations. To simulate real-world sensor limitations, we create self-occluded point clouds by applying a virtual viewpoint-based projection to the full body mesh. Specifically, we generate synthetic partial point clouds by rendering each point cloud from a single viewpoint and removing obstructed points, mimicking single-view RGB-D capture as shown in Fig. 3. This allows us to conduct a controllable evaluation of the registration under self-occlusion. For the occluded point clouds, the number of points is variable, depending on the visible surface area, typically ranging between 3000 and 6000 points per shape.

Training and evaluation are performed in a pairwise unsupervised manner on 50 test pairs selected from the MPI-FAUST dataset. To evaluate the robustness of our method

Figure 3. Generation of synthetic data with occlusion. Left: the original point cloud; middle: the down-sampled point cloud; right: the perspective-view of the point cloud by removing the occluded points to mimic occlusion.

under challenging deformations, we first selected all complex human poses from the MPI-FAUST test set, including examples such as deep bending, twisted torsos, and asymmetric limb placements. These poses are known to include significant non-rigid transformations and are considered a representative subset for dynamic motion. From this curated set of challenging poses, we randomly sampled 50 pairwise combinations of source and target scans to form our evaluation set. This strategy ensures both diversity and deformation complexity in the registration tasks while maintaining controlled experimental conditions.

For each pair, a fresh instance of the deformation network is trained from scratch using the Adam optimizer with an initial learning rate of 10^{-4}, decayed via a ReduceLROnPlateau scheduler. Training proceeds for 200 gradient descent steps. Experiments are conducted on a workstation running Ubuntu 20.04.6 LTS, equipped with an Intel Core i7-11700 CPU (2.50 GHz, 16 threads), 31 GB of RAM, and an NVIDIA GeForce RTX 3060 GPU (12 GB VRAM). Our model is implemented in Python using PyTorch 1.13 and CUDA 11.6.

Tab. 1 reports the accuracy A_s of our method using different values for α_l and α_m. This summarizes the sensitivity of our method to different combinations of the parameters. The results indicate that accuracy remains poor when both parameters are set to relatively low or unbalanced values, with accuracy staying below 30%. A clear trend emerges where the method benefits from larger values of α_m, particularly when α_l is set to 10^2. In this setting, accuracy reaches its maximum of 83.65% at $(\alpha_l, \alpha_m) = (10^2, 10^4)$, significantly outperforming all other choices. This suggests that stronger weighting of the α_m-term, balanced by a moderate α_l, provides the most stable optimization and best alignment quality. Therefore, we adopt $\alpha_l = 10^2$ and $\alpha_m = 10^4$ as the constants for all subsequent experiments.

4.2. Evaluation Metrics

Because the registered point cloud X' and the target point cloud Y may have different cardinalities, we compute all

Table 1. Accuracy A_s of our method using different values for α_l and α_m.

α_l \ α_m	10^1	10^2	10^3	10^4
10^1	12.59	29.80	60.88	63.59
10^2	10.13	11.23	27.73	**83.65**
10^3	10.06	11.12	12.47	27.11
10^4	10.32	7.46	11.57	13.40

evaluation metrics using the nearest neighbor from X' and P^{gt}. Let $x'_i \in X'$ denote a point in the registered point cloud, and let $y_i = NN(x'_i)$ be its nearest neighbor in the ground truth cloud. The evaluation metrics include End-Point Error (EPE), 3D Accuracy (Strict and Relaxed), and Outlier Ratio (OR). End-Point Error, denoted with $\mathscr{E}$, computes the average Euclidean distance between registered points and their closest ground truth counterparts as follows:

$$\mathscr{E} = \frac{1}{|\hat{P}|} \sum_{\hat{p}_i \in \hat{P}} ||\hat{p}_i - p_i^{gt}||_2. \tag{13}$$

3D Accuracy is the fraction of registered points whose relative distance to the nearest ground truth point is within a threshold τ. Strict 3D accuracy, denoted with $\mathscr{A}_s$, is computed when τ is 2.5, and relaxed 3D accuracy, denoted with $\mathscr{A}_r$, is computed when τ is 5 for relaxed accuracy as follows:

$$\mathscr{A}_{s/r} = \frac{100}{|\hat{P}|} \sum_{\hat{p}_i \in \hat{P}} \left(\frac{||\hat{p}_i - p_i^{gt}||_2}{||p_i^{gt}||_2} < \tau \right). \tag{14}$$

Outlier Ratio, denoted with $\mathscr{O}$, is the proportion of registered points whose relative distance to the nearest ground truth point exceeds 30% and is computed as follows:

$$\mathscr{O} = \frac{100}{|\hat{P}|} \sum_{\hat{p}_i \in \hat{P}} \left(\frac{||\hat{p}_i - p_i^{gt}||_2}{||p_i^{gt}||_2} > 0.30 \right). \tag{15}$$

This formulation ensures compatibility between point sets of different sizes while reflecting realistic use cases where dense-to-sparse or partial-to-full registration is required.

4.3. Self-Occlusion

4.3.1. Occlusion in Both Point Clouds

Table 2. Evaluating Self-Occluded Point Cloud Registration where both source and target point clouds are incomplete.

Method	EPE (10^{-2})	A_s	A_r	OR
OAR	6.39 (2.86)	30.61 (16.32)	55.13 (17.16)	16.09 (13.18)
ClusterReg	3.54 (1.02)	56.17 (11.09)	69.68 (11.35)	16.46 (9.08)
BCPD++	2.96 (1.15)	65.16 (19.85)	81.96 (12.98)	4.23 (4.73)
BCPD	2.68 (1.15)	68.50 (19.46)	84.50 (13.07)	3.50 (3.88)
NDP	1.82 (0.60)	80.19 (12.69)	94.27 (5.29)	1.39 (2.13)
Ours	**1.46** (0.41)	**87.33** (8.53)	**96.76** (3.39)	**1.07** (1.46)

We report the quantitative comparison of our proposed method against several state-of-the-art non-rigid registration baselines, namely BCPD [6], BCPD++ [7], NDP [10], CluReg [30], and OAR [31] in Tab. 2. All metrics are evaluated under partial-to-partial scenarios, where both source and target point clouds exhibit self-occlusions and incomplete observations. Our method achieves the lowest EPE of 1.46×10^{-2} with a standard deviation of 0.41×10^{-2}, indicating high precision in correspondence estimation despite occlusion-induced sparsity. In comparison, the next best performer, NDP, records an EPE of 1.82×10^{-2} with a variance of 0.60×10^{-2}. Accuracy metrics also show that our method performs better than other methods, as we obtain 87.33 and 96.76 under strict (A_s) and relaxed (A_r) thresholds, respectively. This indicates it outperforms NDP (80.19, 94.27) and also classical methods like BCPD (68.50, 84.50). Notably, our standard deviation is the lowest, reflecting consistent performance across multiple test executions. Furthermore, our OR is the lowest among all methods at 1.07 with a small deviation of 1.46, demonstrating robustness to spurious regions typically induced by missing surface regions or occluded limbs. Competing approaches such as OAR and ClusterReg exhibit higher OR, highlighting their vulnerability to partial visibility. Overall, the empirical evidence suggests that our framework improves registration accuracy under severe partiality and occlusion, making it suitable for real-time 3D reconstruction applications. The consistently high performance across all metrics underlines the efficacy of our design choices.

Fig. 4 illustrates qualitative results for partial-to-partial non-rigid point cloud registration under large pose variation and severe self-occlusions. The source and target point clouds, depicted in blue, represent partial observations (with an average overlap ratio of 0.35 on a scale of 0-1, where 0 represents no overlap and 1 represents complete overlap) with significant self-occlusions and articulated deformations. The overlaid orange points show the registered source. BCPD and BCPD++ model the registration process via probabilistic inference over Gaussian Mixture Models, which provides robustness against moderate noise and sparsity but struggles under large non-overlapping regions, often resulting in visible misalignment in occluded limbs and joints. ClusterReg leverages fuzzy clustering to enforce structure preservation, but lacks a strong mechanism for handling severe occlusions, leading to over-smoothed deformations or loss of articulation. NDP and OAR offer greater tolerance to partial observations, but their performance degrades in highly articulated regions, as seen in misaligned arms and legs across multiple examples. In contrast, our method achieves consistently accurate and anatomically plausible registration across all test cases.

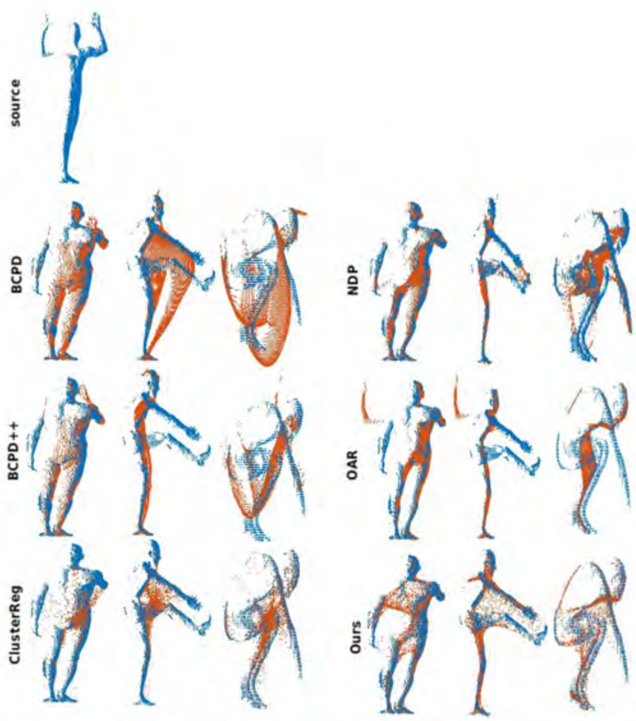

Figure 4. Qualitative results for registration of point clouds with self-occlusion in both source and target point clouds.

4.3.2. Occlusion in One Point Cloud

Tab. 3 presents the quantitative evaluation of our method for the full-to-partial 3D point cloud registration task. In

Table 3. Evaluating Self-Occluded Point Cloud Registration where only the source point cloud is incomplete.

Method	EPE (10^{-2})	A_s	A_r	OR
OAR	6.43 (2.87)	30.10 (15.63)	54.79 (17.13)	16.27 (13.40)
ClusterReg	3.52 (1.02)	56.39 (11.07)	69.84 (11.34)	16.33 (9.00)
BCPD++	2.93 (1.08)	65.92 (18.93)	82.53 (11.64)	4.78 (6.91)
BCPD	2.65 (1.05)	68.78 (18.16)	85.06 (10.88)	3.36 (3.97)
NDP	1.80 (0.53)	80.76 (11.47)	94.61 (4.40)	1.30 (1.86)
Ours	**1.48** (0.43)	**87.08** (8.95)	**96.58** (3.52)	**1.19** (1.84)

terms of EPE, our method achieves a value of 1.48×10^{-2}, outperforming the second-best method, NDP, which has an error of 1.80×10^{-2}. Despite all methods struggling due to the source-side incompleteness, our approach maintains the lowest error and variance, indicating accurate and stable matching under occlusions. On 3D accuracy, we obtain 87.08 strict A_s and 96.58 relaxed A_r thresholds, which surpass all other methods, with NDP achieving 80.76 and 94.61, respectively. Notably, methods like ClusterReg and BCPD++ drop significantly under strict evaluation, highlighting their sensitivity to occluded inputs. We also observe significant improvements in the Outlier Ratio (OR). Our method achieves 1.19, the lowest among all competitors, closely followed by NDP at 1.30. This demonstrates the effectiveness of our occlusion-aware reconstruction in avoiding spurious matches in non-overlapping or missing regions. Overall, these results confirm the robustness of our framework to challenging scenarios involving self-occluded, partial source inputs, where traditional baselines degrade in accuracy or exhibit large variance.

4.4. Large Degree of Deformation

To quantitatively assess the degree of non-rigid deformation between two 3D point clouds, we adopt a patch-wise LLR-based metric. This method estimates the extent to which local geometric structures from the target point cloud can reconstruct the source, thereby reflecting the amount of local deformation. To formally describe our quantitative measure, let us consider $\mathscr{X} = \{x_i\}_{i=1}^{N} \subset \mathbb{R}^3$ denotes the source point cloud and $\mathscr{Y} = \{y_j\}_{j=1}^{M} \subset \mathbb{R}^3$ denotes the target point cloud. For each point $x_i \in \mathscr{X}$, we identify its k-nearest neighbors $\mathscr{N}_i = \{y_{j_1}, \ldots, y_{j_k}\} \subset \mathscr{Y}$ using a spatial k-NN search. We construct a local patch centered at the mean of the neighbors, $\bar{y}_i = \frac{1}{k}\sum_{j \in \mathscr{N}_i} y_j, \tilde{y}_j = y_j - \bar{y}_i$. The reconstruction of x_i is then modeled as a linear combination of the neighbor offsets $\hat{x}_i = \bar{y}_i + \sum_{j \in \mathscr{N}_i} w_{ij}\tilde{y}_j$ where w_{ij} are the optimal weights solving the least-squares problem as,

$$\mathbf{w}_i^* = \arg\min_{\mathbf{w}_i} \| \sum_{j \in \mathscr{N}_i} w_{ij}\tilde{y}_j - (x_i - \bar{y}_i) \|_2^2, \tag{16}$$

which is equivalent to solving,

$$\mathbf{w}_i^* = \arg\min_{\mathbf{w}_i} \| A_i \mathbf{w}_i - \mathbf{b}_i \|_2^2, \tag{17}$$

where $A_i \in \mathbb{R}^{3 \times k}$ is the matrix formed by the centered neighbors $\tilde{y}_j$ and $\mathbf{b}_i = x_i - \bar{y}_i$. Once $\mathbf{w}_i^*$ is obtained, the reconstruction error is computed as $e_i = \| \hat{x}_i - x_i \|_2$ the LLR deformation score between $\mathscr{X}$ and $\mathscr{Y}$ is the average reconstruction error across all points in the source cloud

$$\mathscr{E}_{LLR}(\mathscr{X}, \mathscr{Y}) = \frac{1}{N}\sum_{i=1}^{N} e_i \tag{18}$$

This metric captures local geometric compatibility between point clouds. A low $\mathscr{E}_{LLR}$ implies that the local regional structure of the source can be faithfully reconstructed from the target using linear approximation, indicating low deformation. Conversely, a high error suggests significant structural deviation and hence large local deformation. Unlike global measures such as Chamfer distance or ICP alignment error, the LLR-based score is sensitive to localized geometric differences, making it particularly useful for assessing non-rigid deformations in articulated or self-occluded human poses. Fig. 5 shows the different deformation levels along with the LLR mean value for each pair of point clouds.

Table 4. Evaluation of different levels of deformations. * All values of EPE are of the order 10^{-2}

Method	Low Deformation				Medium Deformation				High Deformation			
	EPE*	A_s	A_r	OR	EPE*	A_s	A_r	OR	EPE*	A_s	A_r	OR
OAR	2.40 (0.57)	71.74 (10.96)	90.41 (5.42)	2.15 (4.41)	9.34 (7.21)	31.99 (21.61)	51.55 (28.76)	26.18 (25.09)	6.20 (1.25)	21.47 (6.38)	46.56 (10.50)	12.07 (8.01)
ClusterReg	3.48 (0.70)	59.87 (4.71)	70.02 (5.39)	21.41 (6.78)	3.47 (0.88)	59.16 (7.75)	70.39 (9.25)	23.36 (7.47)	3.87 (0.84)	48.34 (6.88)	66.45 (9.55)	5.37 (7.48)
BCPD++	1.36 (0.36)	92.73 (4.19)	95.94 (3.36)	0.86 (1.40)	2.22 (0.49)	76.70 (11.07)	90.47 (4.29)	2.78 (2.04)	3.53 (0.68)	54.01 (9.49)	75.89 (6.63)	3.35 (4.09)
BCPD	1.54 (1.16)	89.28 (18.04)	93.59 (14.69)	1.74 (4.40)	1.93 (0.45)	79.82 (10.02)	93.16 (3.81)	1.75 (1.83)	2.90 (0.81)	62.69 (12.72)	82.47 (8.29)	2.55 (3.78)
NDP	1.05 (0.16)	95.90 (2.76)	98.84 (1.49)	0.31 (0.65)	1.76 (0.56)	82.22 (12.41)	94.69 (4.59)	1.61 (2.00)	2.09 (0.58)	73.93 (12.11)	92.20 (5.49)	1.20 (1.93)
Ours	**0.93** (0.11)	**97.41** (1.74)	**99.43** (0.78)	**0.22** (0.34)	**1.55** (0.66)	**86.33** (12.86)	**95.56** (5.44)	**1.28** (1.86)	**1.54** (0.31)	**84.93** (6.17)	**96.36** (2.66)	**1.02** (1.95)

Figure 5. Deformation levels. Top row: a low level of deformation; Middle row: a medium level of deformation; Bottom row: a high level of deformation.

To assess the robustness of each method, we divide the evaluation into three regimes based on the degree of deformation, as shown in Tab. 4. Under low deformation settings, most methods perform reasonably well, but there are distinctions in precision and robustness. Our method achieves the lowest EPE and the highest strict and relaxed accuracies, demonstrating precise alignment under mild shape variation. NDP is the second best in terms of EPE and accuracy, although with slightly higher OR than ours. BCPD++ performs well with decent accuracy and relatively low OR, outperforming traditional BCPD. CluetrReg and OAR show comparatively weaker performance, with higher EPE values and notably larger outlier ratios, indicating limited adaptability in low-deformation settings.

As the degree of deformation increases, performance gaps, our method continues to outperform other methods with the lowest EPE and highest A_s and A_r, showcasing resilience to moderate non-rigid transformations. NDP remains competitive with strong accuracy but with higher variation, as seen in the standard deviations. BCPD and BCPD++ offer a balance between EPE and accuracy, although they are less effective than ours and NDP under modern deformation. OAR underperforms in this regime with high EPE and low accuracy, likely due to its limited ability to model intermediate deformations. ClusterReg, although consistent in accuracy, suffers from a relatively higher OR, reflecting difficulties in rejecting mismatched regions.

This setting presents the most challenging deformation cases, where structural variations between source and target shapes are severe. Our method demonstrates superior robustness with the lowest EPE and accuracy, alongside the lowest OR. This indicates the model's strong ability to preserve point cloud integrity even under large shape changes. NDP again provides competitive performance, though it trails slightly behind ours in accuracy and OR. BCPD offers moderate performance but shows increasing sensitivity to large deformation, as reflected in its rising OR and variance. ClusterReg exhibits a reduced accuracy and an increasing OR, suggesting poor generalization under large transformations. OAR fails to maintain performance in this setting, with EPE and OR substantially higher than other methods, confirming its limited deformation modeling capacity. Across all three deformation regimes, our proposed method consistently yields the lowest EPE, highest accuracy, and lowest OR, which establishes the robustness and precision in handling varying degrees of non-rigid deformation. Particularly in high deformation scenarios, the advantage of our method becomes more pronounced, demonstrating effective shape alignment even under severe transformations where other methods struggle. Despite having robust quantitative results, our method shows room for improvement when there are complex human poses involved in the registration process, that is, when the arms and feet are deformed largely but stay very close to the main body. This can be noticed in the qualitative results shown in Fig. 6.

4.5. Different Levels of Occlusion

In this section, we evaluate registration between the same person in various occluded poses provided in the MPI-FAUST dataset. We synthetically introduce different levels of occlusion into the MPI-FAUST dataset and attempt to register between other poses. Tab. 5 presents a comparative analysis of registration performance on the MPI-FAUST dataset involving occluded intra-person pose pairs, that is, registration between the same subject in different poses.

4.6. Computational Efficiency

Table 6 reports the computation time across registration scenarios, including self-occlusion in both point clouds, self-occlusion in one point cloud, and no occlusion. Our method consistently achieves the lowest average runtime, complet-

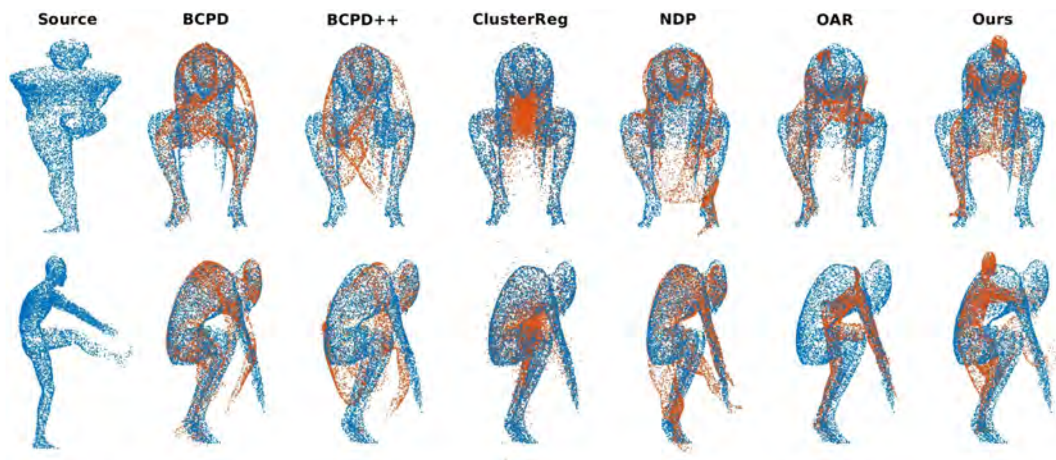

Figure 6. Results of registering point clouds of complex poses without self-occlusion.

Table 5. Evaluation of different levels of occlusions. * All values of EPE are of the order 10^{-3}

	20% occlusion				30% occlusion				40% occlusion			
Method	EPE*	A_s	A_r	OR	EPE*	A_s	A_r	OR	EPE*	A_s	A_r	OR
BCPD	23.39	71.23	85.73	2.94	25.20	68.17	84.85	2.92	27.36	64.92	83.54	3.45
	12.89	18.25	14.07	3.41	11.62	15.74	13.37	3.24	9.06	11.30	9.75	3.64
BCPD++	22.61	72.38	87.29	2.86	24.80	69.69	85.09	2.72	28.08	64.96	83.11	5.61
	9.34	13.96	8.23	3.31	13.55	17.28	16.36	3.03	12.92	13.07	13.98	14.15
GBCPD	14.98	82.53	92.90	1.18	14.81	82.73	93.66	1.32	18.66	77.71	89.98	3.38
	6.46	9.46	7.19	1.46	2.93	4.40	2.57	1.12	11.74	12.55	13.49	14.01
GBCPD++	173.04	27.18	41.55	41.09	177.26	26.85	42.06	41.05	177.25	26.66	41.69	41.09
	173.91	25.22	27.05	29.31	179.24	24.15	27.29	29.95	176.11	23.47	26.61	29.32
NDP	14.20	83.59	95.08	1.11	14.77	82.65	95.21	0.83	16.72	78.73	94.07	1.30
	6.56	11.85	4.88	1.49	5.31	10.03	4.13	1.00	5.23	10.56	4.10	1.53
OAR	**10.44**	**90.58**	97.47	0.89	15.06	89.47	96.30	2.05	22.03	87.38	95.30	2.98
	4.13	6.72	2.47	1.39	18.42	7.49	5.17	4.42	41.68	8.95	6.99	6.53
Ours	11.34	90.71	**98.21**	**0.58**	**11.61**	**90.30**	**97.98**	**0.69**	**12.49**	**89.21**	**97.78**	**0.76**
	2.83	5.61	1.58	0.84	2.52	5.11	1.68	0.88	1.93	4.24	1.50	0.86

Table 6. Computational time. column a represents self-occlusion (X, Y), column b represents self-occlusion (X), and column c represents no occlusion.

Method	occlusion in both	occlusion in one	no-occlusion
Ours	1.99 (0.17)	**2.11 (0.20)**	**2.32 (0.04)**
BCPD	2.35 (1.13)	2.45 (1.07)	2.44 (0.95)
GBCPD	**1.82 (0.75)**	11.18 (6.65)	9.59 (4.50)
NDP	13.35 (3.11)	12.94 (3.07)	11.08 (1.80)
ClusterReg	70.63 (38.19)	69.34 (2.12)	68.28 (0.08)
OAR	64.96 (5.98)	70.10 (7.84)	79.13 (0.67)

ing all registrations in under 2.5 seconds, even when handling challenging full-to-partial and highly deformed full-to-full cases. Compared to traditional optimization-based methods such as ClusterReg and OAR, which require over 60 seconds per pair, our approach offers more than 30 times speedup. Even when compared to the more recent learning-based methods such as NDP and BCPD variants, our method demonstrates better efficiency, underscoring its practicality for real-time or large-scale deployment.

In summary, our method achieves state-of-the-art accuracy and robustness while maintaining a fast runtime in full-to-partial registration settings. The improvements stem from our occlusion-aware deformation architecture, which enables effective matching under severe visibility gaps and structural incompleteness. This efficiency stems from our use of batched LLR computation, which reduces both memory and computational complexity without compromising accuracy.

5. Conclusion

This paper presents an unsupervised neural deformation framework for non-rigid 3D point cloud registration. Our method leverages a learnable positional encoding integrated into a PE-SIREN architecture to capture high-frequency spatial details crucial for resolving fine-grained deformation. To address the ambiguity introduced by partial visibility, we introduced two complementary regularization terms: MCC loss to align visible regions robustly and LLR loss to preserve local geometric consistency in both observed and occluded regions. Extensive experiments were conducted on the synthetically occluded subset of the MPI-FAUST dataset, demonstrating that our method outperforms state-of-the-art techniques in terms of accuracy, robustness, and runtime. In all registration scenarios, our model achieves near-perfect quantitative results, with minimal outlier ratios and constantly low end-point error. Moreover, our framework remains computationally efficient, enabling rapid inference with minimal overhead, making it practical for large-scale or real-time applications. By removing the dependency on ground-truth correspondences, our method paves the way for more robust non-rigid registration in real-world scenarios involving incomplete 3D data.

Although our method achieves robust performance under self-occlusion and varying deformation levels, it remains challenging to handle complex human poses, e.g., limbs close to the torso. In cases such as crossed arms or legs pressed against the torso, local neighborhoods may overlap or merge, making it difficult to preserve structural distinctness during deformation. Future research will focus on integrating part-aware regularization and shape disentanglement strategies to better separate adjacent anatomical regions. Moreover, enhancing the model with global structural cues could improve its ability to maintain fidelity in such highly articulated and near-contact configurations.

References

[1] Brian Amberg, Sami Romdhani, and Thomas Vetter. Optimal step nonrigid icp algorithms for surface registration. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1–8, 2007. 1

[2] Xiaoyang Bai, Zeyu Luo, Lei Zhou, Hongbo Fu, Long Quan, and Chen Change Loy. D3feat: Joint learning of dense detection and description of 3d local features. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 6359–6367, 2020. 1

[3] Federica Bogo, Javier Romero, Matthew Loper, and Michael J. Black. FAUST: Dataset and evaluation for 3D mesh registration. In *Proceedings IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, Piscataway, NJ, USA, 2014. IEEE. 4

[4] Liyuan Chen, Wei Zhao, and Yuwei Li. Fastdeform: Efficient non-rigid registration via geodesic anchor compression. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, 2023. 2

[5] Milad Fathi, Or Litany, Emanuele Rodola, and Alex M Bronstein. Deep functional maps: Structured prediction for dense shape correspondence. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 3689–3698, 2022. 1

[6] Osamu Hirose. A bayesian formulation of coherent point drift. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 43(7):2269–2286, 2021. 5

[7] Osamu Hirose. Acceleration of non-rigid point set registration with downsampling and gaussian process regression. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 43(8):2858–2865, 2021. 5

[8] Minghan Jiang, Xinyu Han, Zhengqi Li, Yinda Wang, and Thomas Funkhouser. Sirenwarp: Learning continuous deformation fields with siren-based implicit templates. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2024. 1

[9] Ruiqi Li, Xiaoyu Yang, Yuxin Tong, Ligang Liu, and Yufeng Guo. Learning deformation fields for point cloud matching. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 4262–4271, 2021. 1

[10] Yang Li and Tatsuya Harada. Non-rigid point cloud registration with neural deformation pyramid. In *Proceedings of the 36th International Conference on Neural Information Processing Systems*, 2022. 1, 5

[11] Yuan Li, Shunwang Wang, Nanyang Yu, and Zhiqiang Cao. Georeg: Geometry-aware registration via deformation-aware features. In *Proceedings of the European Conference on Computer Vision (ECCV)*, 2024. 1

[12] Yuhua Lin, Hsien-Yi Liu, Gao Huang, and Zerong Wang. Partial-nr: Partial-to-partial non-rigid registration via region-level correspondence. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2024. 1

[13] Jingyuan Liu, Yuwei Li, Wei Zhang, Shuai Ma, and Yebin Liu. Learning deformation priors for partial 3d human point clouds. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 4434–4444, 2023. 1

[14] Shuai Ma and Yebin Liu. Structure-aware non-rigid registration with kinematic priors. In *International Conference on Computer Vision (ICCV)*, 2023. 2

[15] Shuai Ma, Qixing Zhou, Ruofei Wang, Qiang Ye, and Yebin Liu. Reliable non-rigid registration under partial visibility. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 12906–12915, 2021. 1

[16] Amar Maharjan and Xiaohui Yuan. Point set registration of large deformation using auxiliary landmarks. In *International Conference on Urban Intelligence and Applications*, pages 86–98. Springer Singapore Singapore, 2020. 1

[17] Amar Maharjan and Xiaohui Yuan. Registration of human point set using automatic key point detection and region-aware features. In *Proceedings of the IEEE/CVF winter conference on applications of computer vision*, pages 741–749, 2022. 1

[18] A. Maharjan, X. Yuan, Q. Lu, Y. Fan, and T. Chen. Non-rigid registration of point clouds using landmarks and stochastic neighbor embedding. *Journal of Electronic Imaging*, 30(3): 031202–1, 2021. 1

[19] Minsoo Park, Sungho Kim, and Jaesik Lee. Neuralcorr: Learning correspondences from partial and occluded point clouds. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2025. 1

[20] Liang Sun, Shijie Li, and Hui Zhou. Partialaligner: Contrastive unsupervised registration of occluded 3d shapes. In *European Conference on Computer Vision (ECCV)*, 2023. 2

[21] Yiran Tang and Qixing Huang. Occreg: Visibility-aware learning for occlusion-robust point cloud registration. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2024. 2

[22] Dong Wang, Shuai Ma, Yuwei Li, and Yebin Liu. Hidden transformation fields: Unsupervised non-rigid point cloud registration with rigidity constraints. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 13606–13615, 2022. 1

[23] Linjie Wei, Qixing Huang, Duygu Ceylan, Etienne Vouga, and Hao Li. Dense human body correspondences using convolutional networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1544–1553, 2016. 1

[24] Cheng Wu, Ying Zhang, Gao Huang, Yiqun Wang, and Wenqiang Sun. Keypointformer: Unsupervised keypoint discovery and correspondence estimation with transformers. In *NeurIPS*, 2023. 1

[25] Tiancheng Xu, Jun Zhao, Dong Wang, and Yebin Liu. Deformernet++: Occlusion-robust non-rigid registration via transformer-based global matching. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2024. 2

[26] Xin Yan, Huajian Xu, Zerong Wang, Zhiqin Ma, Sifei Liu, Xiaolong Wang, Song-Chun Zhu, and Yixin Jiang. Deformingthings4d: Dynamic 4d human-object interactions. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2024. 1

[27] Xiaohui Yuan and Amar Man Maharjan. Non-rigid point set registration: recent trends and challenges. *Artificial Intelligence Review*, 56(6):4859–4891, 2023. 1

[28] Kai Zhang, Yinan Lin, Zekun Wu, and Xiaolong Xu. Partfusion: Occlusion-aware deformation via part-level fusion networks. In *Neural Information Processing Systems (NeurIPS)*, 2023. 1

[29] Wei Zhang, Yuwei Li, Shuai Ma, and Yebin Liu. Learning occlusion-aware correspondences for partial-to-full non-rigid registration. In *European Conference on Computer Vision (ECCV)*, pages 1–18, 2022. 1

[30] Mingyang Zhao, Jingen Jiang, Lei Ma, Shiqing Xin, Gaofeng Meng, and Dong-Ming Yan. Correspondence-free non-rigid point set registration using unsupervised clustering analysis. In *2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 21199–21208, 2024. 2, 5

[31] Mingyang Zhao, Gaofeng Meng, and Dong ming Yan. Occlusion-aware non-rigid point cloud registration via unsupervised neural deformation correntropy. In *The Thirteenth International Conference on Learning Representations*, 2025. 1, 2, 5

[32] Keyu Zhou, Xuyang Zhang, Or Litany, and Yaron Lipman. Unsupervised learning for non-rigid shape matching via soft functional maps. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 8097–8106, 2022. 1