DV-Matcher: Deformation-based Non-Rigid Point Cloud Matching Guided by Pre-trained Visual Features

Supplementary Material

In this supplementary material, we provide more technical details and experimental results, including 1) A detailed description of our Visual Encoding, LG-Network and ARAP loss in Sec. 1; 2) Detailed descriptions of datasets in Sec. 2; 3) Further qualitative results on matching heterogeneous shapes from SHREC'07-H and DT4D-H, quadruped shapes from SHREC'07-Fourleg and SHREC'20 in Sec. 3.1, as well as the full/partial registration results; 4) Quantitative results following the setting from [7, 16, 21], where train/test with the sparse point clouds of fixed 1024 points in Sec. 3.2; Besides, the qualitative result of garment datasets [23], SHREC'07-Fourleg and TOPKIDS are also presented in Sec. 3.2; 5) Robustness evaluation of our method with respect to several perturbations on input in Sec. 3.3; 6) More high-dimensional feature visualization and matching results of different dataset in Sec. 3.4; 7) Experimental setup, hyperparameter instruction in Sec. 3.5 and Sec. 3.6 respectively; 8) The balance between runtime and accuracy (G2) are detailed in Sec. 4; 9) Finally, the broader impacts are discussed in Sec. 5.

1. Technical Details

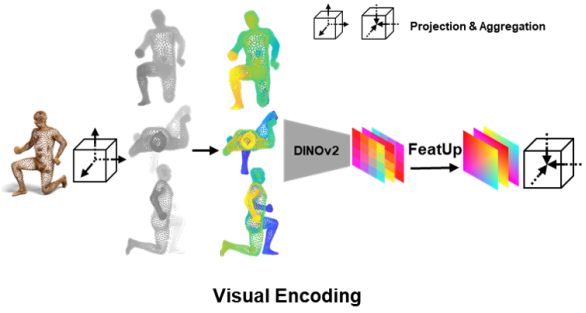

Figure 1. The schematic illustration of the proposed Visual Encoding.

Visual Encoding: Fig. 1 elucidates how we leverage the utilization of features from pre-trained vision models through a point-wise invertible projection from 3D point clouds to 2D images. Specifically, we get features through DINOv2 [28] and lift features via FeatUp [11], where the semantic features are then back-projected to their corresponding points, as described in Sec. 3.1 of the main text.

LG-Net: Fig. 2 illustrates the composition of LG-Net, which aims to refine the features learned from 2D pre-trained vision models, so that is robust to large deformations and

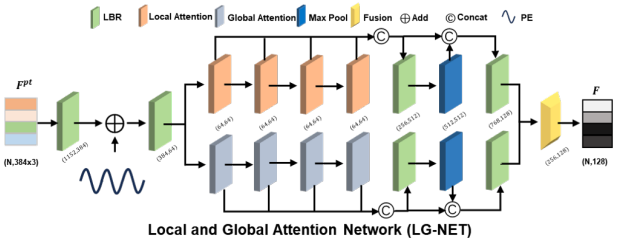

Figure 2. The schematic illustration of LG-Net.

generalized to the challenging partiality. Specifically, for the input representation $F^{pt}(P)$ derived from pre-trained vision models, we employ the LBR [15], which combines Linear, BatchNorm, and ReLU layers, to facilitate the feature dimension transformation into $F'_\Theta \in \mathbb{R}^{N\times 384}$. Following this, we apply position encoding from [27] to integrate 3D absolute position information, which is subsequently combined with the block-wise semantic features F'_Θ to yield a refined representation $F_\Theta \in \mathbb{R}^{N\times 384}$, denotes $F_\Theta = F'_\Theta + \gamma$, where γ is a mapping from $\mathbb{R}$ into a higher dimensional space $\mathbb{R}^{N\times 384}$. Later, our designed network is a dual-pathway architecture in parallel to refine F_Θ, comprising *Global Attention* and *Local Attention*. The two attention modules differ in the receptive field – given a point, the former abstracts features of the remaining points to achieve comprehensive global perceptual awareness, while the latter focuses on its nearest neighborhoods. After undergoing local and global attention mechanisms respectively, we further fuse both features at the end of the refined network to obtain a more comprehensive feature representation, where *Fusion* module consists of LBR and a three-layer stacked N2P [29] attention to merge features.

Network Details: Fig. 3 depicts from left to right the architecture diagrams of our *local attention block*, *global attention block*, and *fusion module*.

ARAP loss: The as-rigid-as-possible term is also incorporated following [14, 22], which reflects the deviation of estimated local surface deformations from rigid transformations:

$$\mathcal{L}_{\text{arap}}^{(\mathcal{S},\mathcal{T})} = ARAP(\mathbf{X}) \tag{1}$$

$$d_{h,l}(\mathbf{X}) = d_{h,l}(\Theta,\Delta) = R\left(\Theta_h\right)\left(g_l - g_h\right) + \Delta_k + g_k - \left(g_l + \Delta_l\right). \tag{2}$$

Here, $g \in \mathbb{R}^{H\times 3}$ are the original positions of the nodes in the deformation graph $\mathcal{DG}$, $\psi(h)$ denotes the 1-ring neighborhood of the $h-$ th deformation node. $R(\cdot)$ corresponds

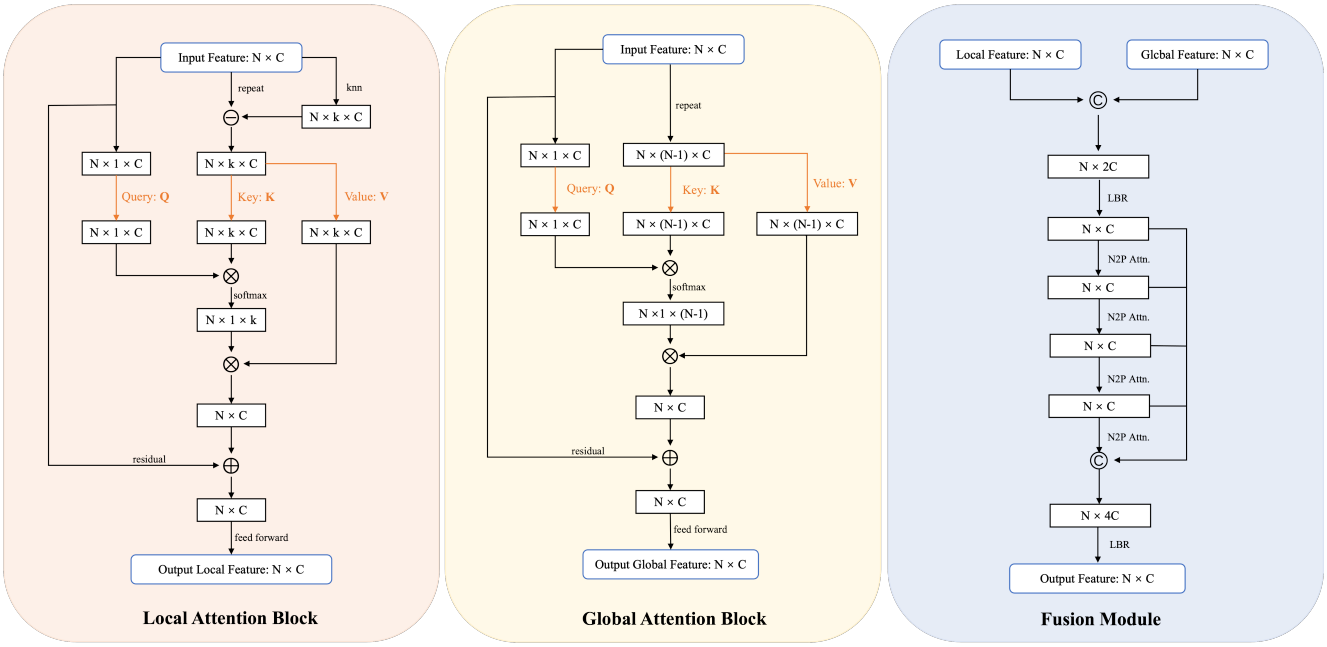

Figure 3. The schematic illustration of the main blocks of LG-Net.

to Rodrigues' rotation formula, which computes a rotation matrix from an axis-angle representation, and α is the weight of the smooth rotation regularization term.

2. Dataset Details

SCAPE_r: The remeshed version of the SCAPE dataset [2] comprises 71 human shapes. We split the first 51 shapes for training and the rest 20 shapes for testing. **FAUST_r:** The remeshed version of FAUST dataset [4] comprises 100 human shapes. We split the first 80 shapes for training and the rest 20 for testing. **SHREC'19_r:** The remehsed version of SHREC19 dataset [26] comprises 44 shapes. We pair them into 430 annotated examples provided by [26] for testing. **DT4D-H:** A dataset from [24] comprises 10 categories of heterogeneous humanoid shapes. Following [17], we use it solely in testing, and evaluating the inter-class maps split in [24]. **SHREC'07-H:** A subset of SHREC'07 dataset [12] comprises 20 heterogeneous human shapes. We use it solely in testing. **SHREC'07-Fourleg:** A subset of SHREC'07 dataset [12] comprises 20 heterogeneous fourleg animals. We use a total of 380 pairs for training. **SHREC'20:** A dataset [9] comprising highly non-isometric non-rigid quadruped shapes of 14 animals, encompassing 12 full shapes and 2 partial shapes. We use it solely for testing. **SURREAL:** It is the large-scale dataset from [13] comprises 230,000 training shapes, from which we take the first 2,000 shapes and use it solely for training. **TOSCA:** Dataset from [33] comprises 41 different shapes of various animal species. Following [7, 21], we pair these shapes to create both for training and evaluation, respectively. **SHREC'16:** Partial shape dataset SHREC'16 [6] includes two subsets, namely CUTS with 120 pairs and HOLES with 80 pairs. Following [3, 5], we train our method for each subset individually and evaluate it on the corresponding unseen test set (200 shapes for each subset). Moreover, we further conduct some practical experiments on partial real scan dataset processed from [19] and medical dataset from [1]. **SURREAL:** The large-scale dataset from [13] comprises 230,000 training shapes, from which we select the first 2,000 shapes and use them solely for training. **SMAL:** Large-scale dataset from [33], which includes parameterized animal models for generating shapes. We employ the model to generate 2000 instances of diverse poses for each animal category, resulting in a training dataset comprising 10000 shapes. **TOPKIDS:** A challenging dataset [20] consists of 26 shapes of a kid in different poses, which manifest significant topological perturbations. **GarmCap:** A dataset from [23], which contains textured 3D garment scans in various poses. We take 40 T-shirt shapes for training, and test on 10 unseen T-shirt shapes, along with 10 long coats, 10 thick coats, and 10 orange coats. **Spleen:** Following [1], we take 32 aligned medical spleens for training and 4 other shapes for testing. **Pancreas:** Following [1], we also take 216 aligned medical pancreases for training and 28 other shapes for testing.

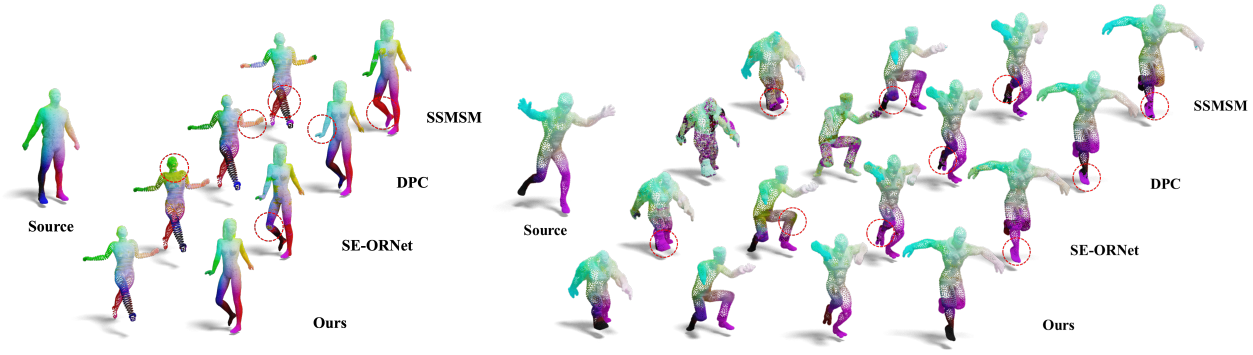

Figure 4. We estimate correspondences between heterogeneous shapes from SHREC'07-H and DT4D-H with DPC,SE-ORNET and one SSMSM, all trained on the SCAPE_r dataset. Our method outperforms the competing methods by a large margin.

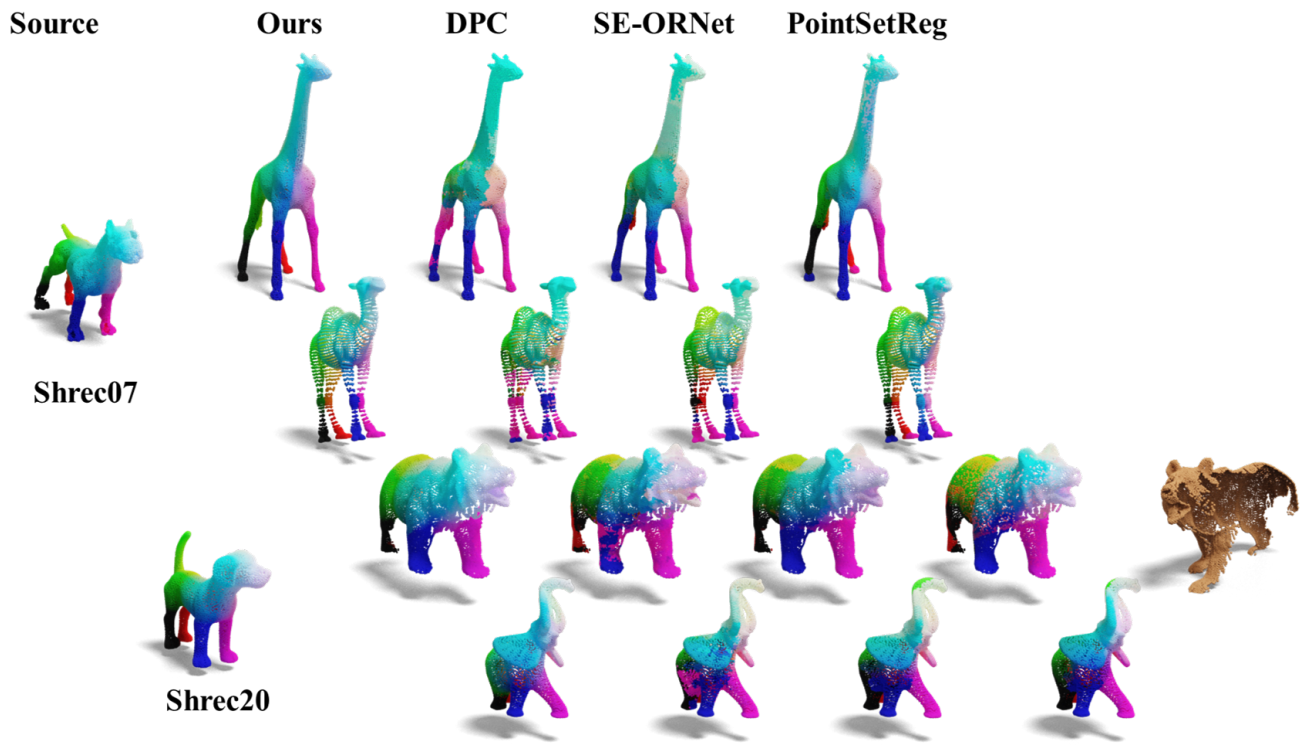

Figure 5. We estimate correspondences between highly non-isometric non-rigid quadruped shapes from SHREC'07-Fourleg and SHREC'20 with DPC,SE-ORNET and PointRegSet, all learning-based methods trained on the SHREC'07-Fourleg dataset. Our method outperforms the competing methods by a large margin. Note that the bear in the third row is incomplete.

3. Additional Experiments

3.1. Further Qualitative Results

Non-isometric human shapes Matching: In Fig. 4, we qualitatively visualize maps obtained by different methods tested in the SHREC'07-H and DT4D-H benchmark. It is obvious that our results outperform all the competing methods, showing superior generalization performance.

Non-isometric quadruped shapes Matching: We also conducted training on the quadruped dataset – SHREC'07-Fourleg, and subsequently tested on challenging SHREC'07 and SHREC'20, respectively. Fig. 5 illustrates several of the most highly non-isometric shapes, where our method significantly outperforms other baselines. Specifically, as we

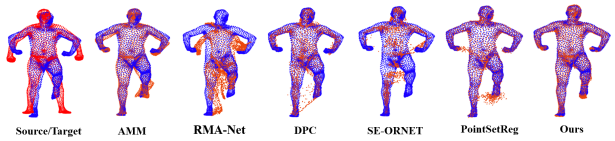

Figure 6. The figure illustrates the registration results of various baselines, along with our proposed deformer.

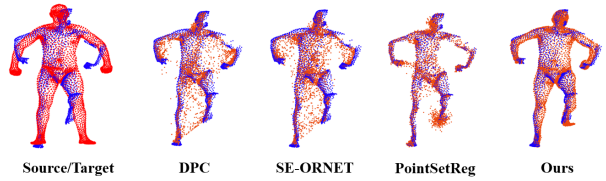

Figure 7. The figure illustrates the partial registration results of various baselines, along with our proposed deformer.

leverage the semantic information extracted from pre-trained vision models and the formulation of geometric information, our approach exhibits promising performance across even challenging heterogeneous shapes.

Full Registration Results: Fig. 6 illustrates the registration results of different methods on full point clouds, where all learning-based methods were trained on SCAPE_r dataset. The results indicate that axiomatic non-learning-based methods, whether AMM [30] or the recent PointSetReg [32], all exhibit errors in the vicinity of the foot area; whereas, learning-based reconstruction methods – DPC [21], SE-ORNet [7] reconstruct the point clouds with substantial noise; RMA-Net [10], which also employs projected 2D images as a prior, but fails to deform effectively to the target shape as well. In contrast, our deformer achieves efficient, high-quality, and smooth deformed point clouds quickly without optimizing iterations.

Partial Registration Results: Fig. 7 presents more challenging cases, namely registering the full point cloud to the partial point cloud, where all learning-based methods were trained on SCAPE-PV dataset. The results show that all other baselines fail to maintain the complete source shape after registration, collapsing into partial, and both learning-based methods [7, 21] and recent axiomatic non-learning-based method [32] result in significant noise post-registration. This further underscores the robustness of our method and our ability to handle partial cases effectively.

3.2. Further Quantitative Results

For the benchmarks involving downsampled point clouds from original shapes, which results in the absence of the complete mesh structure. Thus, we replace the geodesic distance with the Euclidean distance for our evaluation, as defined in Eq. 3. This substitution is detailed in Tab. 1 and Tab. 2 in Supp. Mat., also Tab. 3 within the main text.

Table 1. Quantitative results on human and animals datasets. Acc signifies correspondence accuracy at 0.01 error tolerance, and err denotes average correspondence error ($err \times 1000$). The **best** results in each column are highlighted.

Train	SHREC'19		SURREAL		TOSCA		SMAL	
Test	SHREC'19				TOSCA			
	acc ↑	err ↓	acc ↑	err ↓	acc ↑	err ↓	acc ↑	err ↓
3D-CODED[S] [13]	\	\	2.1%	8.1	\	\	0.5%	19.2
Elementary[S] [8]	\	\	0.5%	13.7	\	\	2.3%	7.6
CorrNet3D[U] [31]	0.4%	33.8	6.0%	6.9	0.3%	32.7	5.3%	9.8
RMA-Net[U] [10]	4.5%	6.0	\	\	2.2%	29.4	\	\
DPC[U] [21]	15.3%	5.6	17.7%	6.1	34.7%	2.8	33.2%	5.8
SE-ORNet[U] [7]	17.5%	5.1	21.5%	4.6	38.3%	2.7	36.4%	3.9
HSTR[U] [16]	19.3%	4.9	19.4%	5.6	52.3%	1.2	33.9%	5.6
Ours [U]	**23.9%**	**4.3**	**27.1%**	**4.0**	**56.2%**	**0.9**	**39.5%**	**3.3**

Sparse Humans/Animals Benchmarks: Following the prior works [7, 16, 21], we conduct the experiments with a consistent sampling point number of $n = 1024$. Specifically, for a pair of source and target shapes $(\mathcal{S}, \mathcal{T})$, the correspondence error is defined as:

$$err = \frac{1}{N} \sum_{x_i \in \mathcal{S}} \left\| f\left(x_i\right) - y_{gt} \right\|_2, \tag{3}$$

where $y_{gt} \in \mathcal{T}$ is the ground truth corresponding point to x_i. Additionally, we measure the correspondence accuracy, defined as:

$$acc(\epsilon) = \frac{1}{N} \sum_{x_i \in \mathcal{S}} \mathbb{I}\left(\left\| f\left(x_i\right) - y_{gt} \right\|_2 < \epsilon d \right), \tag{4}$$

where $\mathbb{I}(\cdot)$ is the indicator function, d is the maximal Euclidean distance between points in $\mathcal{T}$, and $\epsilon \in [0, 1]$ is an error tolerance. We evaluate the accuracy at 1% tolerance following [21].

We train on the SURREAL and SHREC'19 dataset respectively, and then test on the SHREC'19 dataset. Similarly, we train respectively on SMAL and TOSCA dataset, and then test on the TOSCA dataset. As shown in Tab. 1, unlike HSTR[16], which achieves the best performance on its intra-dataset but lags behind SE-ORNet[7] on cross-dataset generalization, our approach excels in both intra-dataset and cross-dataset tests, surpassing all existing methods by over **12%** (**4.3** vs. 4.9). This also complements Tab. 1 in the main text, demonstrating that our method yields robust results whether trained/tested on dense or sparse point clouds.

Garment Dataset: We choose T-shirt (GarmCap_1) to train our DV-Matcher and other baselines, then evaluate on all four sequences of garment dataset [23]. As shown in Tab. 2, our method outperforms the second best over **35%** relative error reduction (**5.24** vs. 8.09).

SHREC'07-Fourleg Dataset: We conducted further validation on challenging heterogeneous quadrupeds. We selected all 20 shapes and uniformly sampled (including upsampling and downsampling) to 5,000 points for training, and tested

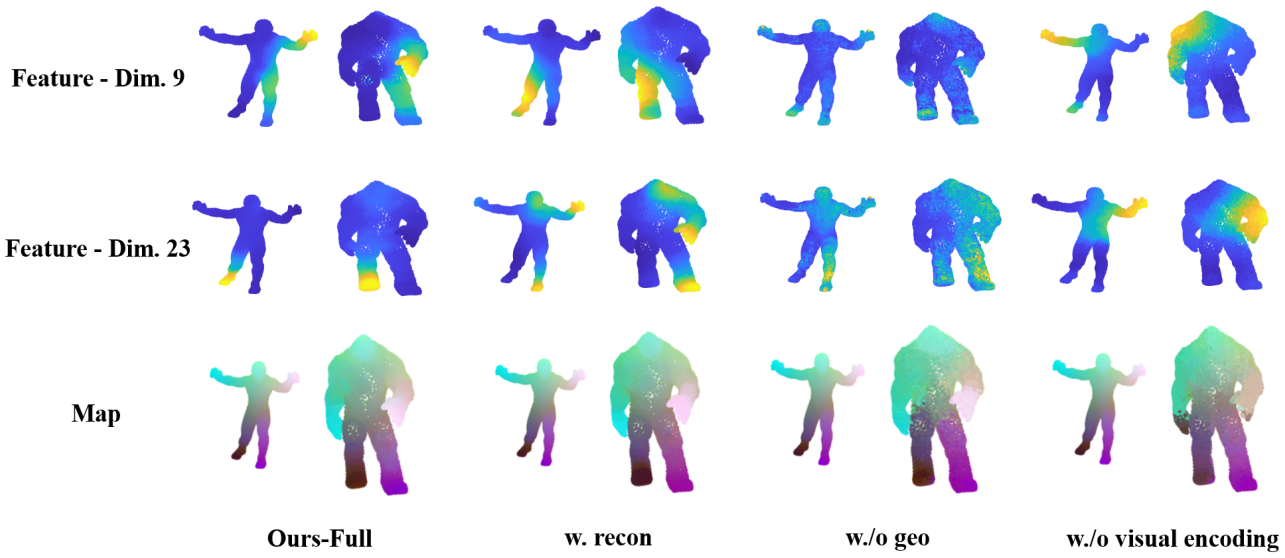

Figure 8. Visualization of different feature dimensions and mapping. **Dim.i** denotes the features of the $i-th$ dimension, where $i \leq 128$.

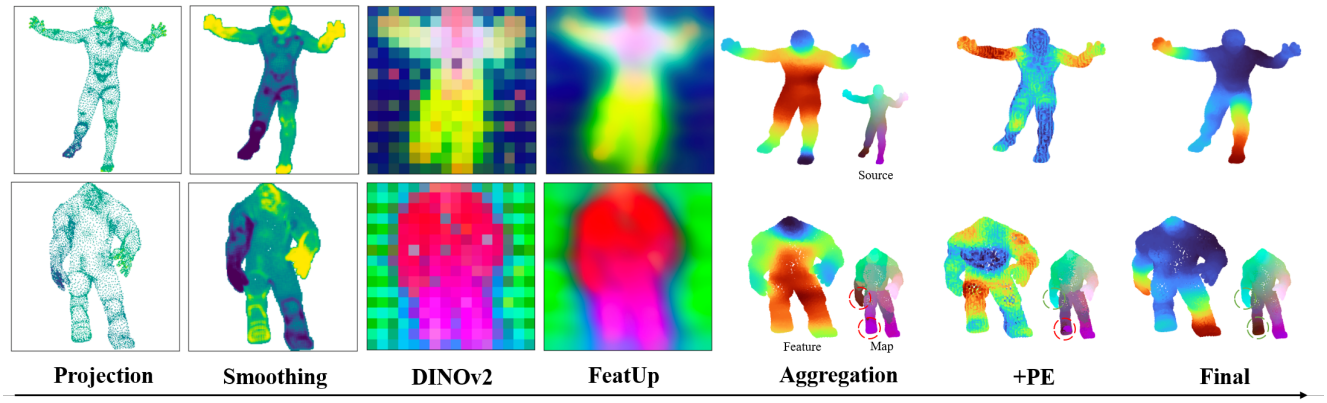

Figure 9. The arrows illustrate the entire process, with features visualized comprehensively. 2D features are visualized using PCA, while 3D shows one channel's features (left) and its map (right).

on 380 pairs of original point clouds. As shown in Tab. 3, our method outperforms **49%** over past approaches (**6.19** vs. 12.37), whether they are learning-based [7, 21] or axiomatic [32].

Topological Noise: We compare PointSegReg, DPC, SE-ORNet and Ours on TOPKIDS [20] (trained on SCAPE_r), achieving respectively error: $20.3, 20.9, 20.2, \mathbf{12.9}$, where ours outperforms the rest by a significant margin of **36%**.

3.3. Robustness

Moreover, we evaluate the robustness of our model with respect to noise and rotation perturbation and report in Tab. 4. More specifically, we perturb the point clouds by: 1) Adding per-point Gaussian noise with i.i.d $\mathcal{N}(0, 0.02)$ along the normal direction on each point; 2) Randomly rotating ± 30 degree along some randomly sampled direction. We perform 3 rounds of test, and report both mean error and the standard deviation in parentheses. Our pipeline delivers the most robust performance among all the other baselines (**0.1** vs. 0.11, **0.25** vs. 0.41), including SE-ORNET[7] which is designed for rotational robustness.

Table 2. Quantitative results on four different garments from GarmCap in terms of Euclidean distance error ($err \times 100$). The **best** is highlighted.

Method	GarmCap_1	GarmCap_2	GarmCap_3	GarmCap_4
PointSetReg[A] [32]	8.95	8.53	9.27	8.96
DPC[U] [21]	7.24	10.12	10.19	9.03
SE-ORNET[U] [7]	7.11	8.09	10.08	8.86
RMA-Net[U] [10]	7.02	10.92	9.55	9.67
Ours[U]	**4.92**	**5.24**	**5.85**	**5.62**

Table 3. Quantitative results on SHREC'07-Fourleg in terms of mean geodesic distance errors ($\times 100$). The **best** is highlighted.

Method	SHREC'07-Fourleg
DPC[U] [21]	20.82
SE-ORNET[U] [7]	17.44
PointSetReg[A] [32]	12.37
Ours[U]	**6.19**

Table 4. Mean geodesic errors ($\times 100$) on under different perturbations. Noisy PC means the input point clouds are perturbed by Gaussian noise. Rotated PC means the input point clouds are randomly rotated within ±30 degrees. The standard deviation value is shown in parentheses.

Method		Unperturbed	Noisy PC	Rotated PC
DiffFMaps[S] [25]	Mesh Required	12.0	14.9(2.57)	26.5(3.35)
NIE[U] [18]		11.0	11.5(0.32)	19.9(1.29)
SSMSM[U] [5]		4.1	5.4(0.11)	9.2(1.01)
DPC[U] [21]	Pure PCD	17.3	18.2(0.80)	22.1(0.72)
SE-ORNet[U] [7]		24.6	24.7(0.15)	27.2(0.41)
Ours [U]		**6.2**	**6.4(0.10)**	**7.0(0.25)**

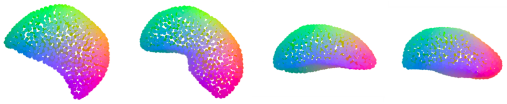

Figure 10. Our matching result of the spleen dataset from [1].

3.4. More Visualizations

High-dimensional feature visualization: To further validate the characteristics of the representations learned by our method, we present a set of more comprehensive visualizations of the features. As shown in Fig. 8, our feature distribution is more clean and localized. However, upon losing geometric or semantic information, the features across different dimensions become divergent, resulting in the loss of regular fine-grained representation at various levels.

We also provide feature visualization after each modules in Fig. 9. Even on humanoid shapes, the visual features tend to be coarse and symmetric, which is not good enough for high-quality dense matching. With our operations, the features become more sharp but also accurate.

Matching results of medical dataset: To supplement Tab. 5 of the main text, we further visualize the matching results on the Spleen in Fig.10, where excellent mapping is achieved regardless of whether the spleen exhibits various shapes or is positioned at different angles.

More qualitative results: We further visualize the results of TOSCA, DT4D, SHREC'07 and SCAPE-PV, which respectively serve as qualitative validation supplements for learning sparse point clouds in Tab. 1, the generalization capability in Tab. 1 of the main text, and the adaptability to partial shapes in Tab. 2 of the main text. The training and testing procedures align with the methods described in the aforementioned table, with quantitative supplements presented respectively in Fig. 11, Fig. 12, Fig. 13 and Fig. 14, respectively. Furthermore, to supplement Fig. 5 and Tab. 3,

Figure 11. More qualitative results of TOSCA. All horse shapes from the dataset have been showcased.

Figure 12. More qualitative results of DT4D. Our method demonstrates a notable improvement over other baselines.

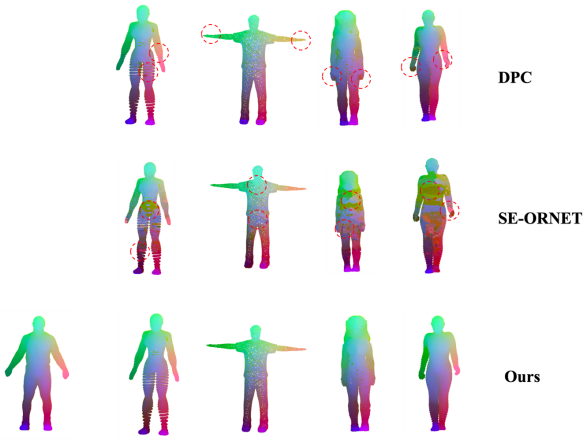

Figure 13. More qualitative results of SHREC'07. Our approach significantly outperforms other baselines.

we further visualized the quantitative performance of our method on SHREC'07-Fourleg and SHREC'20 in Fig. 15.

3.5. Experimental Setup

We perform all the experiments on a machine with NVIDIA A100-SMX4 80GB and Intel(R) Xeon(R) CPU E5-2680 v4 @ 2.40GHz using the PyTorch 2.2.0 framework.

Table 5. Hyper-parameters. The tables details the hyperparameter values that we used for the training of SCAPE_r.

Symbol	Description	Value
k_dist	The nearest number for computing geometrically similarity loss.	500
N_dist	The number of points sampled to calculate the geometrically similarity loss.	1000
k_deform	The number of neighborhood features gathered in the graph convolutional network of our Deformer.	10
k_attn	The number for searching latent nearest features in our Local Attention Block.	40
C	The dimention of output feature.	128
decay_factor	The multiplicative factor by which the learning rate is reduced during each decay.	0.5
decay_iter	The epoch interval at which the learning rate is decayed.	10
alpha	The temperature parameter in the Softmax function. It is a dynamically increasing value with epochs. As alpha increases, Pi becomes harder.	10-100
TEs	Training epochs.	20
H,W	The size of our projected image.	224,224

Figure 14. More qualitative results of SCPAE-PV. Our approach achieves superior performance over other baselines across various partial views.

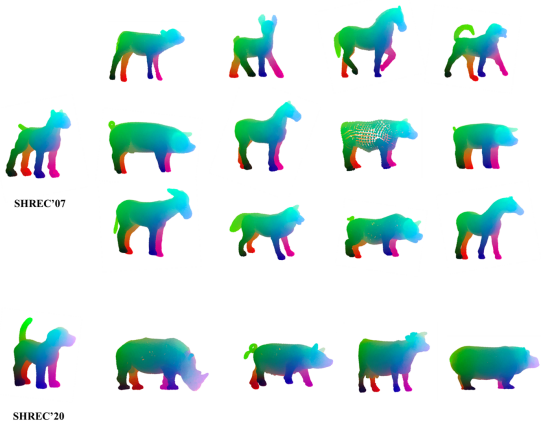

Figure 15. More qualitative results of SHREC'07-Fourleg and SHREC'20.

3.6. Additional Hyper-parameter Details

For a comprehensive understanding of the specific hyper-parameter configurations, please refer to Tab. 5.

4. Runtime and Accuracy Balance

Our method achieves a balance between accuracy and efficiency, which is illustrated in Fig. 16 by comparing with the point-based learning method. DFR in Tab.7 of the main text is a mesh-based method, thus ignored here.

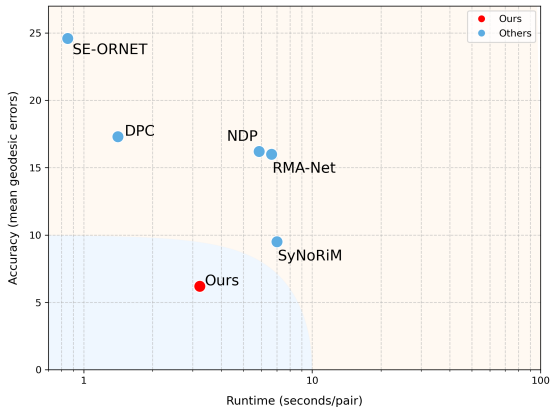

Figure 16. Runtime-Accuracy scatter.

5. Broader Impacts

We fail to see any immediate ethical issue with the proposed method. On the other hand, since our method is extensively evaluated in matching human shapes and achieves excellent results, one potential misuse can be surveillance, which may pose negative societal impact.

References

[1] Jadie Adams and Shireen Elhabian. Point2ssm: Learning morphological variations of anatomies from point cloud. *arXiv preprint arXiv:2305.14486*, 2023. 2, 6

[2] Dragomir Anguelov, Praveen Srinivasan, Daphne Koller, Sebastian Thrun, Jim Rodgers, and James Davis. SCAPE: Shape Completion and Animation of People. 2005. 2

[3] Souhaib Attaiki, Gautam Pai, and Maks Ovsjanikov. Dpfm: Deep partial functional maps. In *2021 International Conference on 3D Vision (3DV)*, pages 175–185. IEEE, 2021. 2

[4] Federica Bogo, Javier Romero, Matthew Loper, and Michael J Black. Faust: Dataset and evaluation for 3d mesh registration. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3794–3801, 2014. 2

[5] Dongliang Cao and Florian Bernard. Self-supervised learning

for multimodal non-rigid shape matching. In *CVPR*, 2023. 2, 6

[6] Luca Cosmo, Emanuele Rodola, Michael M Bronstein, Andrea Torsello, Daniel Cremers, Y Sahillioğlu, et al. Shrec'16: Partial matching of deformable shapes. In *Eurographics Workshop on 3D Object Retrieval, EG 3DOR*, pages 61–67. Eurographics Association, 2016. 2

[7] Jiacheng Deng, Chuxin Wang, Jiahao Lu, Jianfeng He, Tianzhu Zhang, Jiyang Yu, and Zhe Zhang. Se-ornet: Self-ensembling orientation-aware network for unsupervised point cloud shape correspondence. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 5364–5373, 2023. 1, 2, 4, 5, 6

[8] Theo Deprelle, Thibault Groueix, Matthew Fisher, Vladimir G Kim, Bryan C Russell, and Mathieu Aubry. Learning elementary structures for 3d shape generation and matching. *arXiv preprint arXiv:1908.04725*, 2019. 4

[9] Roberto M. Dyke, Yu-Kun Lai, Paul L. Rosin, Stefano Zappalà, Seana Dykes, Daoliang Guo, Kun Li, Riccardo Marin, Simone Melzi, and Jingyu Yang. SHREC'20: Shape correspondence with non-isometric deformations. *Computers & Graphics*, 92:28–43, 2020. 2

[10] Wanquan Feng, Juyong Zhang, Hongrui Cai, Haofei Xu, Junhui Hou, and Hujun Bao. Recurrent multi-view alignment network for unsupervised surface registration. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2021. 4, 5

[11] Stephanie Fu, Mark Hamilton, Laura E. Brandt, Axel Feldmann, Zhoutong Zhang, and William T. Freeman. Featup: A model-agnostic framework for features at any resolution. In *The Twelfth International Conference on Learning Representations*, 2024. 1

[12] Daniela Giorgi, Silvia Biasotti, and Laura Paraboschi. Shape retrieval contest 2007: Watertight models track. *SHREC competition*, 8(7):7, 2007. 2

[13] Thibault Groueix, Matthew Fisher, Vladimir G Kim, Bryan C Russell, and Mathieu Aubry. 3d-coded: 3d correspondences by deep deformation. In *ECCV*, 2018. 2, 4

[14] Chen Guo, Xu Chen, Jie Song, and Otmar Hilliges. Human performance capture from monocular video in the wild. In *3DV*, 2021. 1

[15] Meng-Hao Guo, Jun-Xiong Cai, Zheng-Ning Liu, Tai-Jiang Mu, Ralph R Martin, and Shi-Min Hu. Pct: Point cloud transformer. *Computational Visual Media*, 7:187–199, 2021. 1

[16] Jianfeng He, Jiacheng Deng, Tianzhu Zhang, Zhe Zhang, and Yongdong Zhang. Hierarchical shape-consistent transformer for unsupervised point cloud shape correspondence. *IEEE Transactions on Image Processing*, 2023. 1, 4

[17] Puhua Jiang, Mingze Sun, and Ruqi Huang. Non-rigid shape registration via deep functional maps prior. In *NeurIPS*, 2023. 2

[18] Puhua Jiang, Mingze Sun, and Ruqi Huang. Neural intrinsic embedding for non-rigid point matching. In *CVPR*, 2023. 6

[19] Hanbyul Joo, Tomas Simon, Xulong Li, Hao Liu, Lei Tan, Lin Gui, Sean Banerjee, Timothy Scott Godisart, Bart Nabbe, Iain Matthews, Takeo Kanade, Shohei Nobuhara, and Yaser Sheikh. Panoptic studio: A massively multiview system for social interaction capture. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2017. 2

[20] Zorah Lähner, Emanuele Rodolà, Michael M Bronstein, Daniel Cremers, Oliver Burghard, Luca Cosmo, Andreas Dieckmann, Reinhard Klein, and Yusuf Sahillioglu. Shrec'16: Matching of deformable shapes with topological noise. *Proc. 3DOR*, 2(10.2312), 2016. 2, 5

[21] Itai Lang, Dvir Ginzburg, Shai Avidan, and Dan Raviv. Dpc: Unsupervised deep point correspondence via cross and self construction. In *2021 International Conference on 3D Vision (3DV)*, pages 1442–1451. IEEE, 2021. 1, 2, 4, 5, 6

[22] Zohar Levi and Craig Gotsman. Smooth rotation enhanced as-rigid-as-possible mesh animation. *IEEE transactions on visualization and computer graphics*, 21(2):264–277, 2014. 1

[23] Siyou Lin, Boyao Zhou, Zerong Zheng, Hongwen Zhang, and Yebin Liu. Leveraging intrinsic properties for non-rigid garment alignment. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 14485–14496, 2023. 1, 2, 4

[24] Robin Magnet, Jing Ren, Olga Sorkine-Hornung, and Maks Ovsjanikov. Smooth non-rigid shape matching via effective dirichlet energy optimization. In *2022 International Conference on 3D Vision (3DV)*, pages 495–504. IEEE, 2022. 2

[25] Riccardo Marin, Marie-Julie Rakotosaona, Simone Melzi, and Maks Ovsjanikov. Correspondence learning via linearly-invariant embedding. *Advances in Neural Information Processing Systems*, 33:1608–1620, 2020. 6

[26] Simone Melzi, Riccardo Marin, Emanuele Rodolà, Umberto Castellani, Jing Ren, Adrien Poulenard, et al. Shrec'19: matching humans with different connectivity. In *Eurographics Workshop on 3D Object Retrieval*. The Eurographics Association, 2019. 2

[27] Ben Mildenhall, Pratul P Srinivasan, Matthew Tancik, Jonathan T Barron, Ravi Ramamoorthi, and Ren Ng. Nerf: Representing scenes as neural radiance fields for view synthesis. *Communications of the ACM*, 65(1):99–106, 2021. 1

[28] Maxime Oquab, Timothée Darcet, Theo Moutakanni, Huy V. Vo, Marc Szafraniec, Vasil Khalidov, Pierre Fernandez, Daniel Haziza, Francisco Massa, Alaaeldin El-Nouby, Russell Howes, Po-Yao Huang, Hu Xu, Vasu Sharma, Shang-Wen Li, Wojciech Galuba, Mike Rabbat, Mido Assran, Nicolas Ballas, Gabriel Synnaeve, Ishan Misra, Herve Jegou, Julien Mairal, Patrick Labatut, Armand Joulin, and Piotr Bojanowski. Dinov2: Learning robust visual features without supervision, 2023. 1

[29] Chengzhi Wu, Junwei Zheng, Julius Pfrommer, and Jürgen Beyerer. Attention-based point cloud edge sampling. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 5333–5343, 2023. 1

[30] Yuxin Yao, Bailin Deng, Weiwei Xu, and Juyong Zhang. Fast and robust non-rigid registration using accelerated majorization-minimization. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2023. 4

[31] Yiming Zeng, Yue Qian, Zhiyu Zhu, Junhui Hou, Hui Yuan, and Ying He. Corrnet3d: Unsupervised end-to-end learning

of dense correspondence for 3d point clouds. In *CVPR*, pages 6052–6061, 2021. 4

[32] 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 *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 21199–21208, 2024. 4, 5, 6

[33] Silvia Zuffi, Angjoo Kanazawa, David W Jacobs, and Michael J Black. 3d menagerie: Modeling the 3d shape and pose of animals. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 6365–6373, 2017. 2