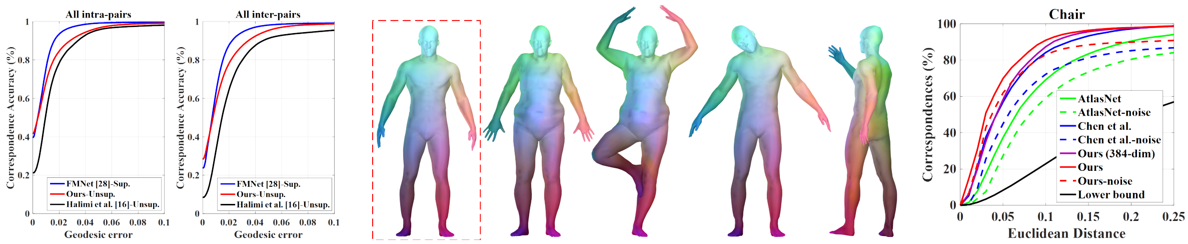

(a) (b) (c)

Figure R. 1: (a) Geodesic error comparison of ours (unsup.), FMNet [28] (sup.) and Halimi *et al.* [16] (unsup.) on FAUST. (b) One target shape and 4 pair-wise corresponded source shapes. (c) Additional semantic correspondence results for chair category in BHCP.

We thank the reviewers for their comments. All reviewers are positive about our *novelty* and *compelling results*. We address reviewer's concerns as follows, and will revise the paper accordingly, including missing citations, typos, etc.

Review1Q1. Computation time. Implemented as a custom C++/CUDA extension, our Chamfer distance calculation is efficient, *e.g.*, $\sim 42.8ms$ for two sets of point clouds $[B \times N \times 3]$, where the batch size B=10, the number of vertex N=4,096. Although the implicit function is point-wise, the K query 3D locations can be constructed as a 3D tensor $[B \times K \times 3]$ and passed to the fully connected layers. Our training on one category (500 samples) takes $\sim$6 hours (1, 1, and 4 hours for Stage 1, 2, 3 respectively) to converge with a GTX1080Ti GPU. In inference, the average runtime to pair two shapes (N=4,096) is $22.3ms$ including runtime of E, f, g networks, neighbour search and confidence calculation.

R1Q2. Dense correspondence on FAUST dataset. As suggested, we conduct additional experiments on FAUST humans dataset, with two SOTA baselines: supervised (FMNet [28]) and unsupervised (Halimi *et al.* [16]) methods. We follow the experimental protocol of [16, 28], and set N=8,192, d=256. Note that we use **real scans** in both training and testing. The ground-truth densely aligned shapes are used for evaluation. As shown in Fig. R.1(a), our method, as a unsupervised method, outperforms the unsupervised method [16]. Fig. R.1(b) visualizes the estimated correspondences.

R1Q3. Aligned vs. unaligned data. ***i)*** By only finding the closest points on aligned 3D shapes, we report its semantic correspondence accuracy as the black curve in Fig. R.1(c), following the setting of Sec. 4.1. Clearly, our accuracy is much higher than this 'lower bound', indicating our method doesn't rely much on the canonical orientation. ***ii)*** In fact, for practical applicability, we did train a rotation-invariant model for unaligned test setting (L226-228 and Fig. 5 in the main paper). Our method achieves competitive performance as optimization-based baselines.

R1Q4. Segmentation comparison. Unlike ours, the correspondence baselines cannot automatically predict segmentation. They require additional templates to transfer segmentation labels. Thus we only compared with the SOTA unsupervised segmentation method BAE-Net. Here, we report segmentation results (IoU: 78.5%) of Chen *et al.* [7] on the chair category by using a single template of 3 parts: back+seat, leg, arm, which is worse than ours (88.9%).

Review2Q1. Justifications on f and part embedding vector (PEV). ***i)*** On the test set of chair category in segmentation experiment, we measure the statistics of Cosine Similarity between the PEVs and their corresponding one-hot vectors: 0.972 ± 0.200 (BAE-Net), 0.966 ± 0.401 (ours). This shows PEVs are continuous and *approximately* one-hot vectors. We view this approximation not a bug, but a by-product of limited network capability, and leverage it to learn PEV. ***ii)*** Per your suggestion, we retrain a model by using the 384-dim feature (1 layer before $\mathbf{o}$ as Fig. 1(b) in **Supp.**) as point embedding. As shown in Fig. R.1(c), the correspondence accuracy (purple curve) is worse than ours. ***iii)*** Based on these results and t-SNE plots of PEVs in Fig. 7, we believe the PEVs of surface points can serve as point embedding.

R2Q2. CR loss for missing part. The CR loss is the key to ensure that corresponding points in two shapes share the *same* PEV (Eqn. 1). When defining CR loss for two shapes with a mis-match part, there are two strategies: 1) pursue correspondences for all points first, even on mis-match part; and detect mis-match part via the PEV-based confidences $\mathcal{C}$. 2) define CR loss "only" for highly-confident corresponded points while ignoring points on mis-match part. Strategy 2 requires confidence calculation, which is impossible when PEV has not yet satisfied Eqn. 1. We chose Strategy 1, as it optimizes PEV to satisfy Eqn. 1, during which points on mis-match part are likely outliers, and reflected in lower $\mathcal{C}$.

Review3Q1. Evaluation on real data. ***i)*** To validate on noisy real data, we evaluate on the BHCP (normalized) testing data with additive noise $\mathcal{N}(0, 0.02^2)$ and compare with Chen *et al.* [7] and AtlasNet [14]. As shown in Fig. R.1(c), the accuracy is slightly worse than testing on clean data. However, our method still outperforms baselines on noisy data. ***ii)*** In R1Q2, we use real scans in FAUST dataset for testing, which also illuminates our effectiveness on real data.

R3Q2. Train on aligned data. We will clarify the training data requirements early in Sec. 1. Please also refer to R1Q3.

R3Q3. EMD and biases issues. ***i)*** Please refer to R2Q2. ***ii)*** We agree that there are biases our evaluation due to biased annotation (only on salient points). However, our unsupervised framework encourages dense corresponding of *all* points, *without* using salient point supervision. Hence, our model itself has no semantic correspondence bias.

Review4Q1. Compare with AtlasNet on BHCP. Following the same experimental setting, we report the semantic correspondence accuracy of AtlasNet [14] in Fig. R.1(c). Our method outperforms AtlasNet.

R4Q2. Impact on shape representation. We did show our improved shape representation power in Sec. 4.4 (L267).