Supplementary Material for Unsupervised Learning of Category-Specific Symmetric 3D Keypoints from Point Sets

Clara Fernandez-Labrador[1,2,3], Ajad Chhatkuli[3], Danda Pani Paudel[3], Jose J. Guerrero[1], Cédric Demonceaux[2], and Luc Van Gool[3,4]

[1] I3A, University of Zaragoza, Spain
[2] VIBOT ERL CNRS 6000, ImViA, Université de Bourgogne Franche-Comté, France
[3] Computer Vision Lab, ETH Zürich, Switzerland
[4] VISICS, ESAT/PSI, KU Leuven, Belgium
{cfernandez,josechu.guerrero}@unizar.es,
cedric.demonceaux@u-bourgogne.fr,
{ajad.chhatkuli,paudel,vangool}@vision.ee.ethz.ch

Abstract. In this **supplementary document**, we provide more insights regarding symmetry, including the proof of Proposition 1. Furthermore, an experiment showing the generalization of our method on real data is included, as well as some results for the segmentation label transfer task. Finally additional qualitative results are presented on the four datasets evaluated in the main paper at the end of the document.

1 Symmetry

1.1 Symmetric deformation space.

Proof. The two linear spaces due to the two basis $\mathcal{B}_{\mathcal{C}\frac{1}{2}}$ and $\mathcal{B}'_{\mathcal{C}\frac{1}{2}}$ are symmetric by Definition 1 as $\mathcal{B}_{\mathcal{C}\frac{1}{2}}$ is symmetric to $\mathcal{B}'_{\mathcal{C}\frac{1}{2}}$ for any $K \in \mathbb{Z}$. Let $\mathsf{c}_i \in \mathcal{L}$ and $\mathsf{c}'_j \in \mathcal{L}'$ represent the respective half coefficients for any two shape instances i and j, where $\mathcal{L}$ and $\mathcal{L}'$ defines the spaces of the predicted half coefficient vectors. Consequently, the actual deformation spaces are symmetric to one another if $\mathcal{L}$ and $\mathcal{L}'$ are equal. We define $p : p(\mathsf{c}_i)$ as the probability distribution of c_i and $q : q(\mathsf{c}'_j)$ as the probability distribution of c'_j. If p and q come from the same distribution, we approach $p = q$. Then we have:

$$\begin{aligned} &\text{if } \mathsf{c}_i = \mathsf{c}'_j, \\ &\text{either, } p(\mathsf{c}_i) = q(\mathsf{c}'_j) = 0, \\ &\text{or, } p(\mathsf{c}_i) > 0 \text{ and } q(\mathsf{c}'_j) > 0 \\ &\text{for all, } \mathsf{c}_i \in \mathcal{L}, \mathsf{c}'_j \in \mathcal{L}'. \end{aligned} \tag{1}$$

Condition (1) guarantees that $\mathcal{L} = \mathcal{L}'$ and thus we obtain a symmetric deformation space. □

Note that for condition (1) to be true, we do not require the two distributions to be equal, however, it is sufficient and desirable to have so. Therefore, Proposition 1 in the main text highlights such sufficient and desirable case. It is particularly meaningful when we are learning to predict the coefficients through stochastic methods such as a neural network training. In our network architecture indeed one can expect the distributions of these two vectors to be similar given the data exhibits such a symmetric deformation space, since the prediction branches of c_i and c'_i are very similar. Alternatively, one may also try to enforce the condition using a KL divergence loss.

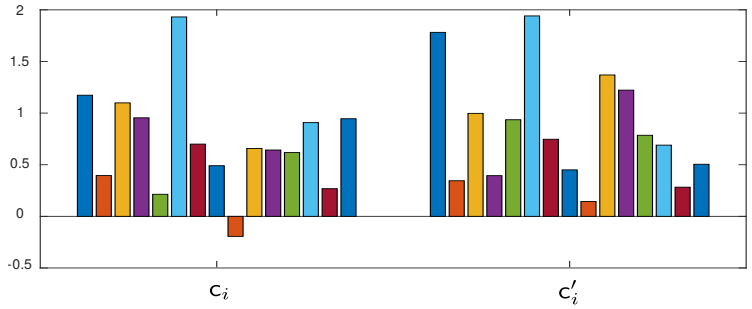

Fig. 1: **Coefficients distribution**. Mean values of c_i components (left) and c'_i components (right) for the Dynamic FAUST [1]. The mean of the variances for the different components are: c_i : 0.54, c'_i : 0.50. The figure shows that the network learns similar distribution for the coefficients c_i and c'_i.

1.2 Symmetry Plane Parametrization.

As mentioned in Sec. 5.3 in the main paper, we observe that handling misaligned data with unsupervised methods can lead to some rotation ambiguities. More specifically, we observe that different combination of basis shapes can result in different alignments.

As we show in Fig. 5 in the text, predicting the symmetry plane of the object category allows to have more control over the predicted instance poses. We came up with the idea of learning an additional common parameter, $\mathsf{R}_{\mathcal{C}}$, which is directly related to the symmetry plane. By adding this category-specific parameter, the network learns a common rotation for all the objects in the category. As a consequence, the instance-wise rotation, R_i, can be thought like an offset from the reference basis alignment. Several evaluations confirmed that this strategy helps the learning process, reducing the rotation ambiguities.

2 Additional Experiments

2.1 Keypoints correspondence

We provide a complete overview for all the object categories evaluated regarding the keypoints correspondences across instances in Fig. 2. This demonstrates the

ability of our model to capture and model the inter-subject shape variations and intra-subject deformations in a category.

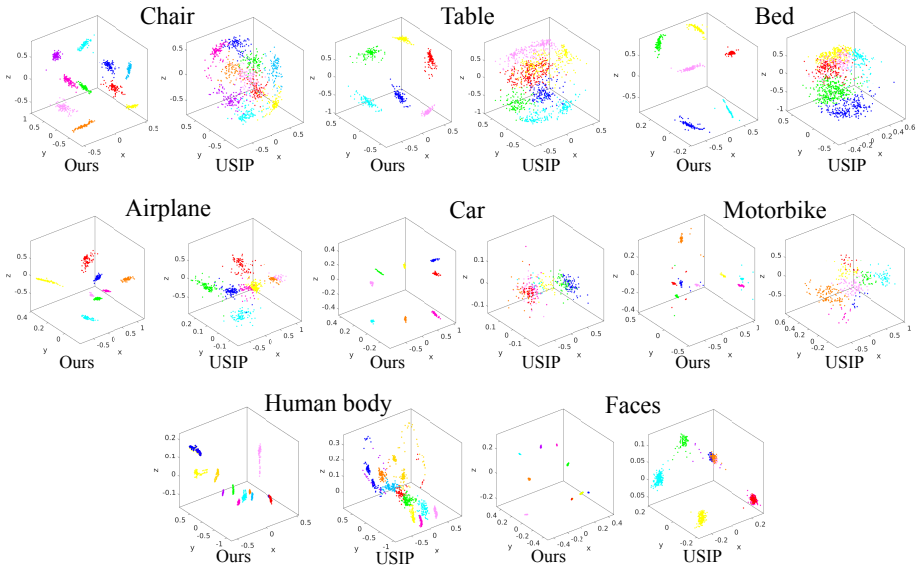

Fig. 2: **Keypoints correspondence across instances**. We cluster the keypoints predicted for all the instances of a category to show their geometric consistency. Note how our keypoints get neatly clustered creating a general 3D shape template.

2.2 Segmentation Label Transfer

As demonstrated in Sec. 5.2 in the main paper, our predicted keypoints correspond to semantically meaningful locations. Therefore, here we explore the utility of the proposed category-specific keypoints for the segmentation label transfer task. In this experiment, for every point in the original shape $\mathsf{s}_{ij} \in \mathsf{S}_i$, we find its closest category-specific keypoint $\mathsf{p}_{ik} \in \mathsf{P}_i$, and transfer the corresponding semantic label to it. We assume the keypoints labels are known and correspond to those in Fig. 4 in the paper.

Some qualitative results are shown in Fig. 3. Our method achieves full correspondence between instances, therefore avoiding placing keypoints in less representative parts. An example is the engine, in grey, in the case of airplanes. This is reflected in the label transfer since there is no distinction of these parts. Besides that, only with eight keypoints in the example, we achieve reasonable results, close to the ground truth data.

2.3 Real Data

In this section, we show the performance of our method for real data in Fig. 4. For this experiment, the network is trained on the chair category from the

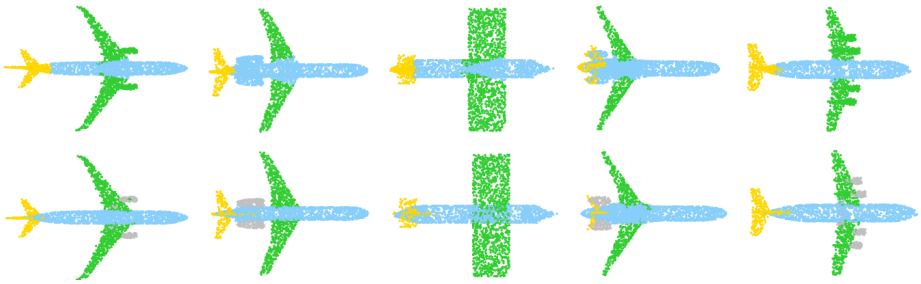

Fig. 3: **First row:** results of performing semantic label transfer with our keypoints. **Second row:** ground truth. This is evaluated in ShapeNet part dataset [2] using eight keypoints for the label transfer.

ModelNet10 dataset [3] and tested on real chairs from the SUNRGBD dataset [4]. To generate the real data dataset from [4], we crop the points inside the ground truth 3D bounding boxes provided by the authors. Real data entail additional challenges. This is not only because shapes appear incomplete and noisy, but also because other objects may cause occlusions, e.g. part of a table occluding a chair. As illustrated in Fig. 4, even though real data is fairly challenging, our network can still produce corresponding meaningful keypoints.

Being able to generalize to previously unseen real objects as demonstrated in Fig. 4 is crucial and really useful for many tasks such as guide for shape completion or shape generation.

Fig. 4: Results in real chairs from SUNRGBD dataset [4] training with CAD chairs from ModelNet10 dataset [3].

3 Qualitative results

In this section, we provide additional qualitative results on various object categories from the datasets evaluated in the paper; ModelNet10 [3] in Fig. 5, ShapeNet parts [2] in Fig. 6, Dynamic FAUST [1] in Fig. 7 and Basel Face Model 2017 [5] in Fig. 8.

Again, we note that our network predicts corresponding keypoints between instances of the same category and consistently associates the same keypoint with the same semantic part. For instance, for the chair object category, the keypoint colored in pink is always associated with the chair back, the keypoint colored in cyan is associated with the front left leg, etc.

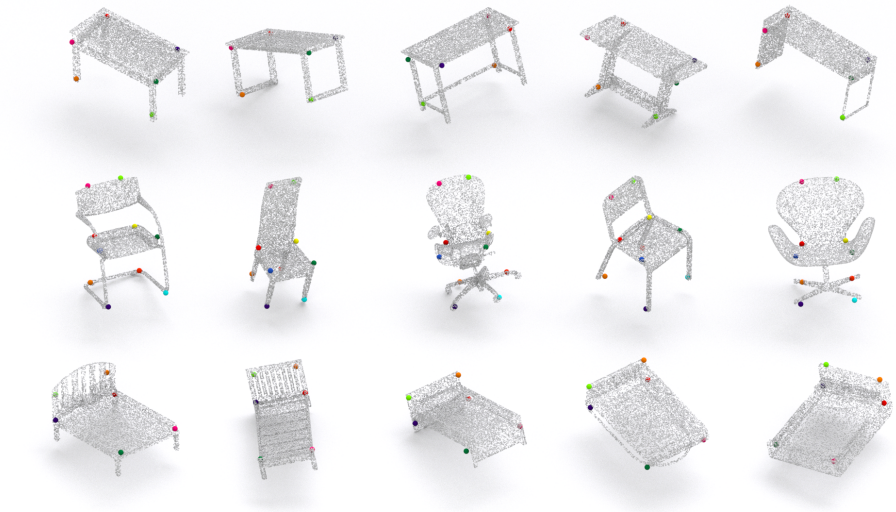

Fig. 5: Qualitative results in table, chair and bed categories from ModelNet10 dataset [3].

References

1. Bogo, F., Romero, J., Pons-Moll, G., Black, M.J.: Dynamic faust: Registering human bodies in motion. In: CVPR. (2017) 6233–6242
2. Yi, L., Kim, V.G., Ceylan, D., Shen, I.C., Yan, M., Su, H., Lu, C., Huang, Q., Sheffer, A., Guibas, L.: A scalable active framework for region annotation in 3d shape collections. ACM Transactions on Graphics (TOG) **35**(6) (2016) 1–12
3. Wu, Z., Song, S., Khosla, A., Yu, F., Zhang, L., Tang, X., Xiao, J.: 3d shapenets: A deep representation for volumetric shapes. In: CVPR. (2015) 1912–1920
4. Song, S., Lichtenberg, S.P., Xiao, J.: Sun rgb-d: A rgb-d scene understanding benchmark suite. In: Proceedings of the IEEE conference on computer vision and pattern recognition. (2015) 567–576

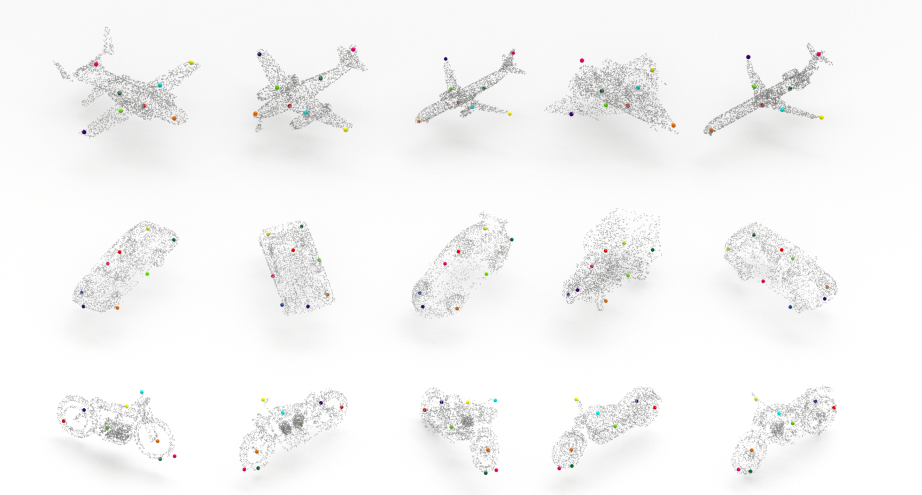

Fig. 6: Qualitative results in airplane, car and motorbike categories from ShapeNet parts dataset [2].

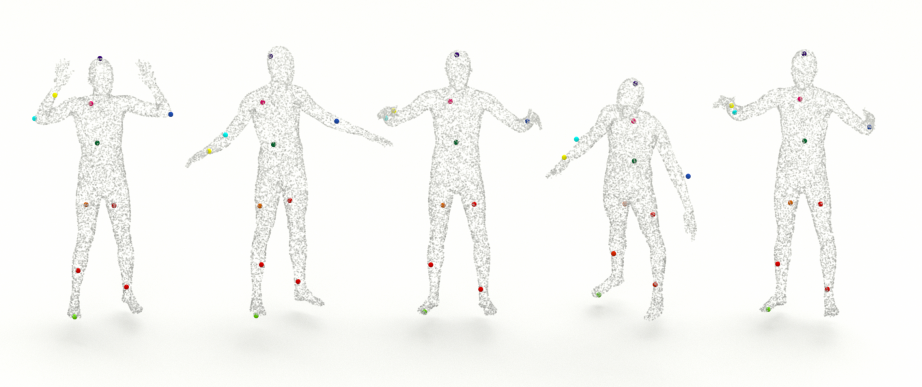

Fig. 7: Qualitative results in human bodies from Dynamic FAUST dataset [1].

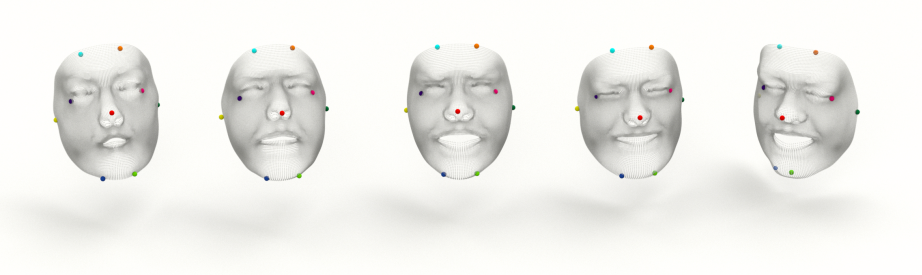

Fig. 8: Qualitative results in faces from Basel Face Model 2017 dataset [5].

5. Gerig, T., Morel-Forster, A., Blumer, C., Egger, B., Luthi, M., Schönborn, S., Vetter, T.: Morphable face models-an open framework. In: 2018 13th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2018), IEEE (2018) 75–82