Unsupervised Learning of Dense Shape Correspondence –Supplementary Material–

Oshri Halimi
Technion, Israel
oshri.halimi@gmail.com

Or Litany
Facebook AI Research
orlitany@fb.com

Emanuele Rodolà
Sapienza University of Rome
rodola@di.uniroma1.it

Alex Bronstein
Technion, Israel
bron@cs.technion.ac.il

Ron Kimmel
Technion, Israel
ron@cs.technion.ac.il

1. Implementation

The code to our paper is available at https://github.com/OshriHalimi/unsupervised_learning_of_dense_shape_correspondence and will be accessible after the conference.

2. Training with scarce data

As discussed in Section 5.1 of the original manuscript, having an unsupervised learning method bridges the gap between axiomatic solvers and supervised learning methods. The latter can excel on particular data but often suffer from limited generalization capabilities, while the former offer a general purpose tool for solving matches between unseen pairs but suffer from computational inefficiency. Our method can do both, while demonstrating improved capabilities in both regimes. The TOSCA experiment in Section 5.4 (Figure 6) shows that our network trained on human scans generalizes well to non-human shapes. The single-pair experiment in Section 5.1 (Figure 1) shows that our method can efficiently solve for a single unseen pair of shapes. In addition, we have emphasized its usefulness in fast inference given a scarce unlabeled train set. In what follows we present additional evidence that did not fit into the main paper due to page limitation.

Fast inference. As discussed above, when fast inference is required on newly encountered unlabeled data, axiomatic methods are no longer an option. Also, one cannot afford full retraining and therefore has two options: either using a pre-trained network on a labelled similar data using supervised learning, or use the unsupervised network to train quickly on few examples. We demonstrate this using an artistic model of *Deadpool*, a super-hero comics character, provided in a variety of poses sampled from animations. To convert the artistic mesh to a manifold we used [3]. The models were remeshed to a 7K resolution, using edge contraction [2]. We wish to stress that the artistic models do *not* have any ground-truth labeling, emphasizing the usefulness of an unsupervised approach.

In Figure 1 we compare the performance of the unsupervised network, trained with only 3 shapes for a total of 15 minutes (100 iterations); and the supervised network trained on FAUST synthetic human dataset (80 shapes) for 8 hours (3K iterations). Visualized are the test examples (*i.e.*, pairs of shapes unavailable to the network at training time). In both methods we show the network predictions, without any further post processing. While both methods demonstrate equivalent inference time of less than one second, the performance gap significantly shows a clear advantage for our method.

In addition, we show the results of an axiomatic method, using SHOT descriptors and Functional Maps framework and refined using product manifold filter (PMF). Comparing the processing time between the methods, for the unsupervised network the training process duration was 15 minutes, while for the supervised network it was 8 hours, and for both networks, given a test pair of shapes, the inference time was less than a second. On the other hand, the axiomatic method had to solve the problem from scratch for any new pair of test shapes and the PMF refinement phase consumed one hour for each pair of shapes. We can observe that even a very short training on a very small fraction of a newly encountered dataset, could lead to fairly good results on the remaining unseen pairs of shapes. A PMF refinement procedure we can be applied also in the unsupervised case, consuming more time and resulting in perfect matching.

3. Synthetic FAUST additional visualization

In section 5.2 of the main paper we showed in Figure 5 the results of the unsupervised network trained on syn-

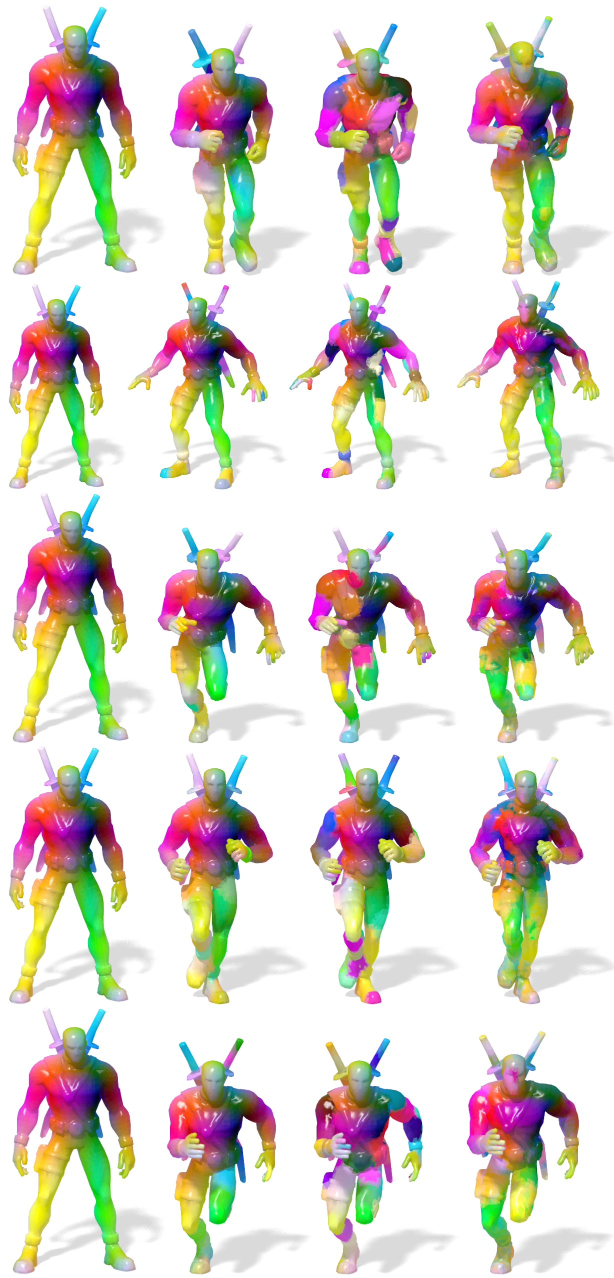

Figure 1: From left to right: Reference model; Correspondences calculated using our unsupervised network, trained on just 3 poses of Deadpool; Correspondences calculated using a supervised network, trained on FAUST synthetic human dataset (80 shapes); Correspondences calculated using the purely axiomatic method of functional maps with SHOT descriptors. Note that only the axiomatic results are refined using PMF, while for the former we show the raw network predictions. Corresponding points are assigned the same color.

thetic Faust train-set, when applied on the test pairs. We provided a visualization of the resulting correspondence for intra-subject as well as for inter-subject test pairs. Due to lack of space, we only included few visualizations. In Figure 2 we provide more visualizations.

4. Dynamic FAUST additional visualization

Section 5.4 of the main paper discusses the generalization of our network, trained on the FAUST synthetic human dataset, on the recent Dynamic FAUST dataset [1]. As demonstrated in Figure 8 of the original paper, when tested on 256 test pairs comprising of 4 different subjects at 4 different poses, our method showed extremely good generalization results. Due to lack of space, we only included few visualizations. In Figure 3 we show many more results via texture transfer.

References

[1] F. Bogo, J. Romero, G. Pons-Moll, and M. J. Black. Dynamic FAUST: Registering human bodies in motion. In *IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, July 2017. 2, 4

[2] M. Garland and P. S. Heckbert. Surface simplification using quadric error metrics. In *Proceedings of the 24th annual conference on Computer graphics and interactive techniques*, pages 209–216. ACM Press/Addison-Wesley Publishing Co., 1997. 1

[3] J. Huang, H. Su, and L. Guibas. Robust watertight manifold surface generation method for shapenet models. *arXiv preprint arXiv:1802.01698*, 2018. 1

[4] M. Vestner, R. Litman, E. Rodolà, A. M. Bronstein, and D. Cremers. Product manifold filter: Non-rigid shape correspondence via kernel density estimation in the product space. In *CVPR*, pages 6681–6690, 2017. 4

Figure 2: Visualization of the calculated correspondences for synthetic Faust test pairs, illustrated by texture transfer according to the estimated map. In each row, the first column shows the reference shape to which the remaining shapes are matched.

Figure 3: Generalization of our network trained on synthetic Faust dataset to Dynamic FAUST [1], illustrated by texture transfer according to the estimated map. In order to convert our method's raw outputs to a bijection, results are refined using PMF [4]. In each row, the first column shows the reference shape to which the remaining shapes are matched.