

A. Our dataset G-S-H

In our experiments, we showed quantitative evaluations on multiple benchmarks, including FAUST [1], FAUST remeshed [8], MANO [10], SURREAL [12] and the SHREC20 challenge [3]. These five, as well as many more existing 3D shape datasets, can be roughly classified in two classes: (i) Synthetic datasets with dense ground-truth, near-isometries or a compatible meshing and (ii) real datasets with non-isometric pairs and sparse annotated ground-truth correspondences¹. In many cases, for (i) the objects are within the same class and therefore have a similar intrinsic geometry, but they undergo challenging, extrinsic deformations with large, non-rigid pose discrepancies. For (ii), the topological proportions of a pair of objects can be quite different, but the poses are less challenging than (i).

To address this disconnect between non-isometries and large non-rigid deformations in existing benchmarks, we create our own dataset, where the goal is to jointly address all the challenges mentioned above: Our benchmark has non-isometric pairs of objects from different classes, large-scale non-rigid poses and dense annotated ground-truth correspondences for evaluation.

A.1. The dataset

We created objects of 3 different classes for our dataset with the tool ZBrush: A dog (Galgo), a cat (Sphynx) and a human. In modeling these shapes, we took great care to obtain generic but anatomically correct instances of these distinct species, see Figure 1 for example shapes from all three classes. We furthermore endowed all objects with a UV-map parameterization, as well as a wireframe acting as a deformation cage. Moreover, the range of motions of one object is specified by a hierarchical set of joints that is consistent for all objects in the dataset. We then animate the different objects by specifying different configurations in terms of deformation handles and applying the deformation to the full shapes with a skinning technique. The UV parameterizations were defined in a way that they are consistent across all considered classes, as a patchwork of smaller components/regions of all objects.

A.2. Experiments

We performed a number of experiments on our new benchmark. For evaluation, we select a number of ~ 120 uniformly sampled keyframes for training and define 32 different poses as our test set. In the main paper, the matching accuracy for our method, as well as other unsupervised matching approaches are compared for this setup. Specifically, we followed the same evaluation protocol that we

¹Of course not all existing benchmarks fall under one of these two categories. Some notable exceptions are datasets that specify on a certain class of objects (like humans) [6, 1] or a specific type of input noise (like partiality or topological changes) [9, 5]

mentioned earlier in Section 4.1 for the results in Fig. 4, main paper. Since we have dense ground-truth correspondences that are consistent across all surfaces, we can also display the mean geodesic error at each individual point of the objects. Specifically, we take the UV-map parameterization on one pose of the ‘Galgo’ shape from our dataset and display the mean matching error of all pairs in the test set. Furthermore, we show qualitative examples of interpolations obtained with our method in Figure 2.

B. Ablation study

We now provide an ablation study where we examine how certain parts of our method contribute to our empirical results. Specifically, we perform the following ablations:

- (i) Remove the auxiliary correspondence loss ℓ_{geo} .
- (ii) Train for correspondences directly without the interpolator module from our architecture (see Fig. 2, main paper). This means that we only use ℓ_{geo} and ignore the other two loss components.
- (iii) Remove the max-pooling layers in Eq. (4) from our architecture.
- (iv) Replace the EdgeConv layer in Eq. (3) with a standard PointNet [7] layer.
- (v) Replace our feature extractor with KPConv² [11].

We then report how these changes affect the geodesic error and the mean conformal distortion (interpolation error) on FAUST remeshed, corresponding to the results in Table 1 and Figure 6, main paper. Specifically, we compare the results without post-processing:

		Geo. err.	Conf. dist
Ours		2.3	0.10
(i)	No ℓ_{geo}	13.0	0.13
(ii)	No interp.	4.7	–
(iii)	No maxpool	2.5	0.14
(iv)	EdgeConv	10.6	0.25
(v)	Use KPConv	4.2	0.28

Table 1: Ablations.

These experiments indicate that both the interpolator and the feature extractor are crucial for obtaining high quality results: Modifying technical details of our feature extractor leads to suboptimal results (iii)-(v). The difference is particularly large when EdgeConv is replaced by PointNet

²KPConv is a state-of-the-art architecture for point cloud learning, but its main emphasis is on tasks like object classification and segmentation. It was, however, used in a matching pipeline before in prior work [2].



Figure 1: **The G-S-H dataset.** We show 3 examples each for the 3 classes in our G-S-H dataset. Note, that all 3 classes share the same parameterization, despite the varying body proportions. In particular this means that we can obtain dense ground truth correspondences between all pairs of shapes, which we indicate here with a texture map.

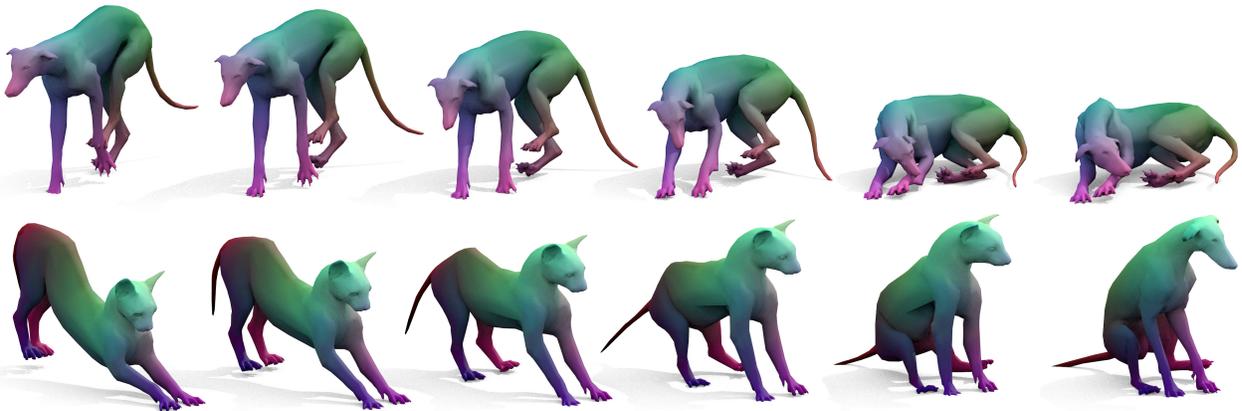


Figure 2: **Interpolation on G-S-H.** Two interpolation sequences on our own benchmark G-S-H obtained with NeuroMorph. This shows clearly that, while our method contains interesting non-isometric pairs, the non-rigid pose variety is still significant.

(iv). Similarly, without the interpolator module, the correspondence estimation is less accurate (ii), since they are not based on an explicit notion of extrinsic deformation. Finally, without the geodesic loss ℓ_{geo} , the matching accuracy deteriorates significantly (i). This can be attributed to the fact that, without a notion of intrinsic geometry, our method is prone to run into non-meaningful local minima.

C. Additional qualitative examples

Finally, we show a few more qualitative results from the SHREC20 benchmark. Specifically, we display examples of non-isometric interpolations in Figure 3 and a qualitative comparison of correspondences obtained with our method and smooth shells [4] in Figure 4.

D. Digital puppeteering

One interesting property of our method is that it is able to learn geometrically plausible pose priors for any shape \mathcal{X} . Given any target pose \mathcal{Y} , we generally obtain a meaningful new pose of the input object \mathcal{X} as the last pose of the interpolation sequence $t = 1$. Consequently, by considering a distribution of target poses \mathcal{Y} , we automatically obtain a shape space of admissible poses with the object identity \mathcal{X} . This allows for digital puppeteering as an application of our method. To that end, we jointly train NeuroMorph for a set of poses from the TOSCA dataset of animals and humans, as well as the SURREAL dataset which consists of a large collection of SMPL shapes. As a proof of concept,

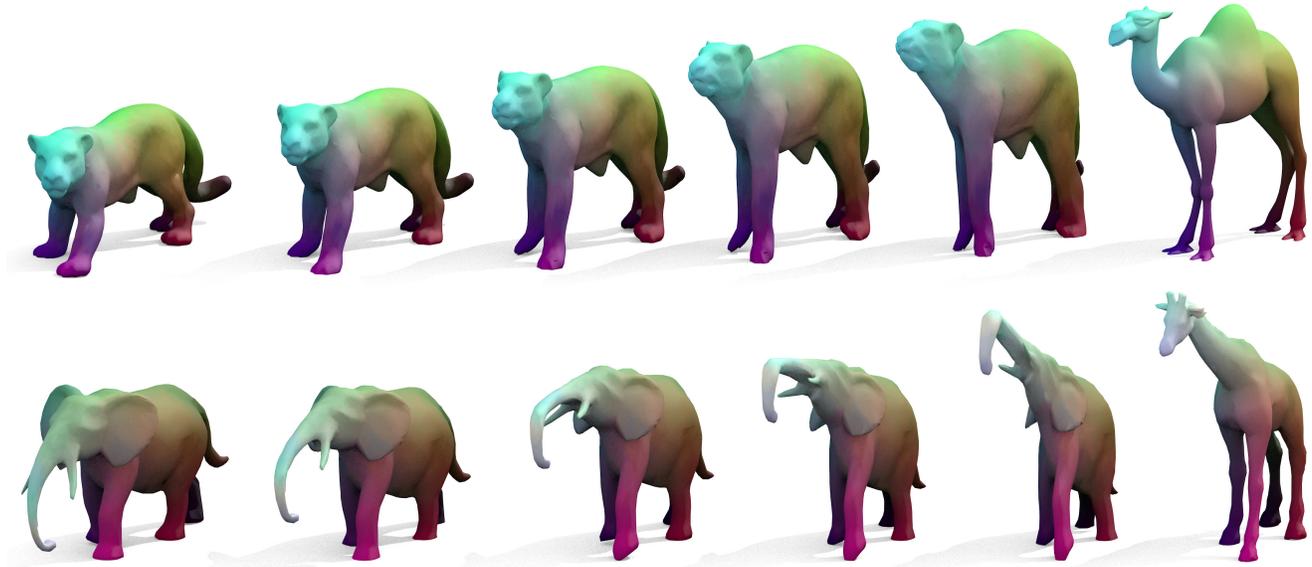


Figure 3: **Interpolation on SHREC20.** We show two additional examples of interpolation sequences obtained with our method for pairs of shapes from the SHREC20 [3] dataset. For each input pair, our method acts on the pose of the first input objects (left) while mostly preserving its identity. The elephant uses its trunk to imitate the shape of the giraffe’s head. While this can be considered meaningful from a geometric perspective, it also reveals a limitation of our approach. The fully unsupervised setup occasionally fails to find semantically exact correspondences, if the geometric features have a very different appearance.

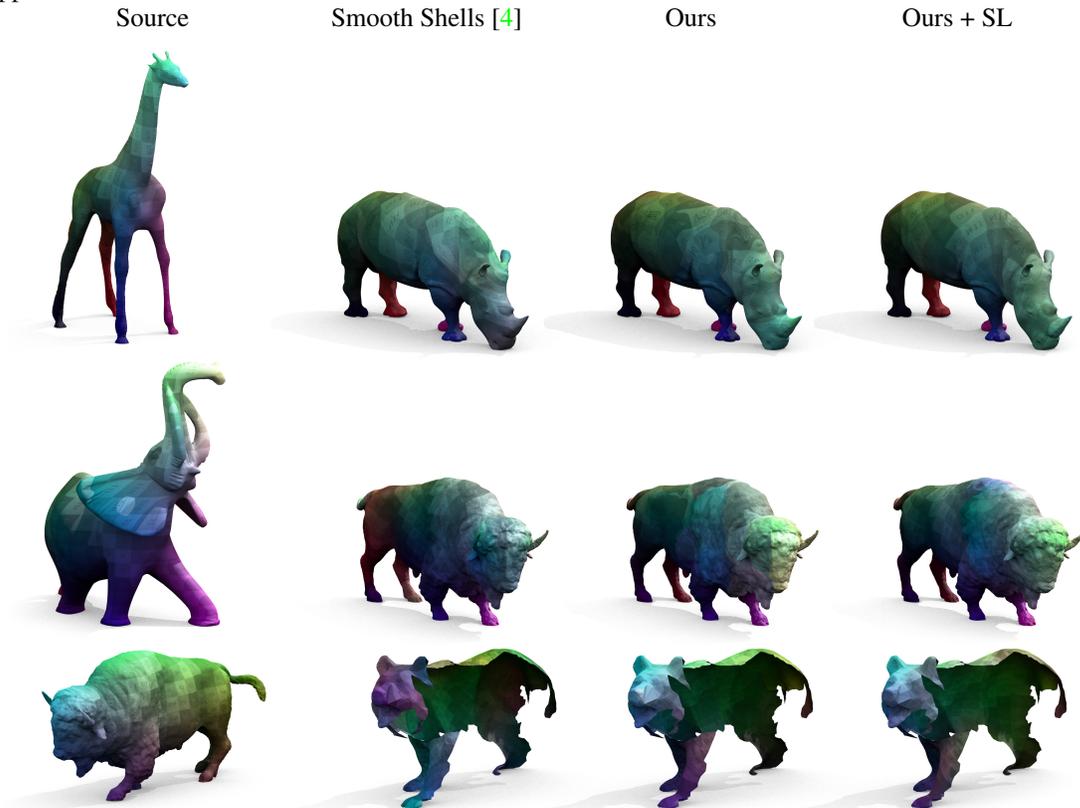


Figure 4: **Unsupervised correspondences on SHREC20.** We show two more qualitative comparisons of correspondences obtained with different methods on the SHREC20 benchmark.

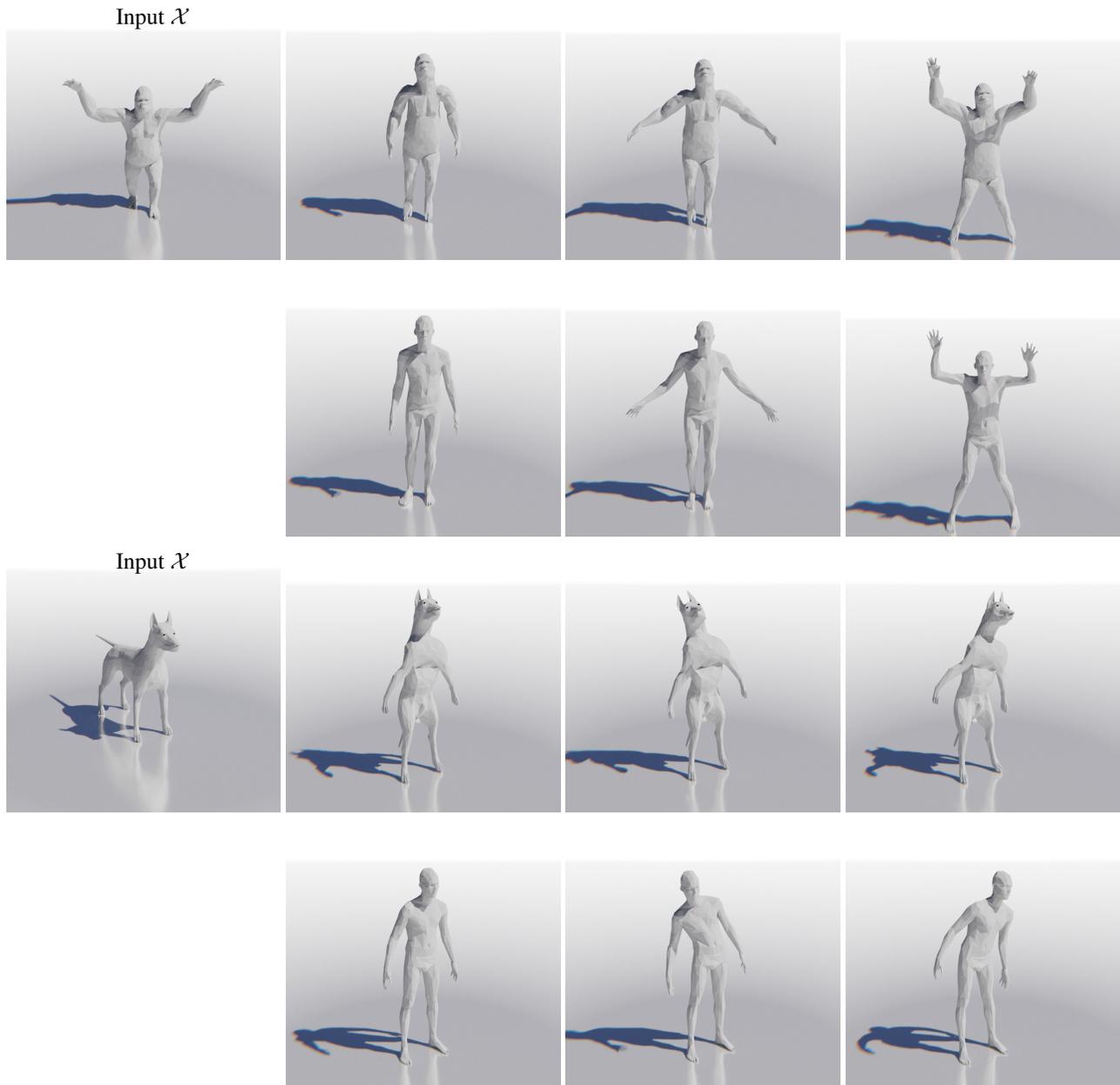


Figure 5: **Digital puppeteering.** We train NeuroMorph jointly for a collection of animal and human (SMPL) shapes. In that manner, we effectively learn a pose prior for the animal shapes which allows us to animate them according to a reference sequence of SMPL shapes from DFAUST. See also our attached videos for the full, animated versions of the two sequences shown here.

we then query our network for a time-continuous sequence of SMPL shapes from the DFAUST dataset and animate the sequence by replacing the human shape with different animals, see Figure 5 and also see our attached videos in the supplementary material.

E. Additional quantitative comparisons

For the sake of completeness, we also provide quantitative comparisons on the SHREC19 [6] benchmark, see Figure 6. Note that, like for FAUST, we again use the more recent remeshed version of the dataset, first introduced in [2].

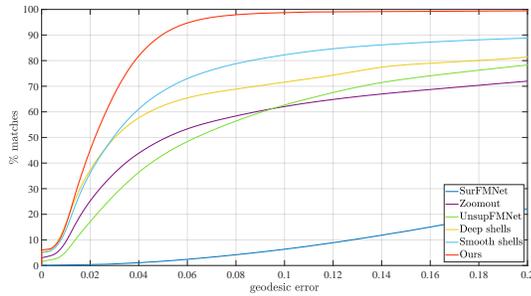


Figure 6: **Unsupervised correspondences on SHREC19 remeshed.** A comparison of unsupervised approaches, showing the cumulative geodesic error curves on the 430 challenge test pairs.

References

- [1] 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, June 2014. IEEE. 1
- [2] Nicolas Donati, Abhishek Sharma, and Maks Ovsjanikov. Deep geometric functional maps: Robust feature learning for shape correspondence. *arXiv preprint arXiv:2003.14286*, 2020. 1, 4
- [3] 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. 1, 3
- [4] Marvin Eisenberger, Zorah Lahner, and Daniel Cremers. Smooth shells: Multi-scale shape registration with functional maps. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 12265–12274, 2020. 2, 3
- [5] 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. *Proceedings of Eurographics Workshop on 3D Object Retrieval (3DOR)*, 2:11, 2016. 1
- [6] Simone Melzi, Riccardo Marin, Emanuele Rodolà, and Umberto Castellani. Matching humans with different connectivity. *Proceedings of Eurographics Workshop on 3D Object Retrieval (3DOR)*, 2019. 1, 4
- [7] Charles R Qi, Hao Su, Kaichun Mo, and Leonidas J Guibas. Pointnet: Deep learning on point sets for 3d classification and segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 652–660, 2017. 1
- [8] Jing Ren, Adrien Poulénard, Peter Wonka, and Maks Ovsjanikov. Continuous and orientation-preserving correspondences via functional maps. *ACM Trans. Graph.*, 37(6):248:1–248:16, Dec. 2018. 1
- [9] Emanuele Rodolà, Luca Cosmo, Michael Bronstein, Andrea Torsello, and Daniel Cremers. Partial functional correspondence. *Computer Graphics Forum (CGF)*, 2016. 1
- [10] Javier Romero, Dimitrios Tzionas, and Michael J. Black. Embodied hands: Modeling and capturing hands and bodies together. *ACM Transactions on Graphics, (Proc. SIGGRAPH Asia)*, 36(6), Nov. 2017. 1
- [11] Hugues Thomas, Charles R Qi, Jean-Emmanuel Deschaud, Beatriz Marcotegui, François Goulette, and Leonidas J Guibas. Kpconv: Flexible and deformable convolution for point clouds. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 6411–6420, 2019. 1
- [12] Gul Varol, Javier Romero, Xavier Martin, Naureen Mahmood, Michael J Black, Ivan Laptev, and Cordelia Schmid. Learning from synthetic humans. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 109–117, 2017. 1