Unsupervised 3D Structure Inference from Category-Specific Image Collections

Supplementary Material

Weikang Wang Dongliang Cao Florian Bernard

University of Bonn, Germany

9. Implementation Details

We set $\lambda_1 = 1e^{-3}$, $\lambda_2 = 1e^{-4}$ in Eq. 10, $n = 10$, $\sigma^2 = 5e^{-5}$ in Eq. 2 and $h = 0.15$ in Eq. 9 for all experiments. Our experiments show that more training iterations lead to better results, which is a common phenomenon in many unsupervised learning settings. In this work, we set the maximum number of training iterations to 40,000 (one iteration means forward and back-propagation computation of one input image), which balances result quality and training time/energy cost.

10. Number of Bases of the Shape Space Model

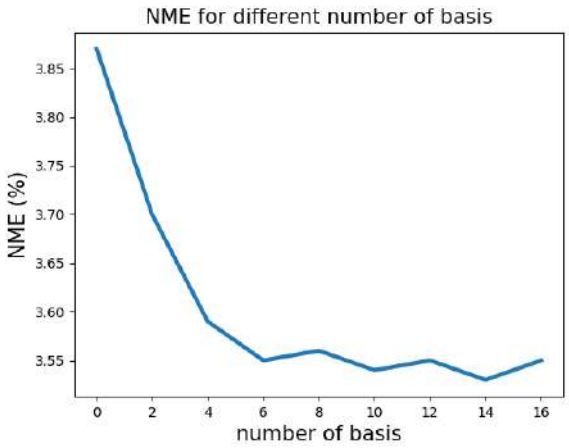

Figure 10. NME on CELEBA WILD for different number of basis of space shape model.

We conduct an ablation study on the number of bases of the shape model on CELEBA WILD, see Fig. 10. We observe that once the number of bases is sufficiently large (above 5-6 in this case), the error (NME) remains constant, indicating that the variability of this object category is sufficiently explained.

11. Results on HUMAN3.6M Dataset

As explained in the main paper, our approach is less effective for the HUMAN3.6M dataset due to large pose variations and articulation. Nevertheless, for the sake of completeness, we include additional quantitative experiments using the HUMAN3.6M dataset, which can serve as baseline for future works. The normalized mean error (NME) is used as evaluation for a varying number of keypoints, which we show in Table 5.

K=8	K=16	K=24	K=32
0.837	0.487	0.446	0.392

Table 5. NME on HUMAN3.6M dataset for different number of keypoints.

12. Shape Synthesis

By varying the coefficients of a specific basis vector of the learned space shape model, we can synthesize new shapes. Fig. 11 illustrates this for one instances of the HORSE dataset.

Figure 11. Animation of a horse shape by varying the coefficient of a basis vector of the learned space shape model.

13. Analysis of Space Shape Model

As specified in Sec. 9, we set the number of basis $n = 10$ for all experiments. In order to visualize the effectiveness of the space space model, we show the 3D shape with an increasing number of basis functions, i.e. $M + \alpha_1 \times B_1$, $M + \alpha_1 \times B_1 + \alpha_2 \times B_2$, ..., where α_i and B_i are the i-th coefficient and basis vector, respectively. Fig. 12 and Fig. 13 visualize this process.

From Fig. 12 and Fig. 13 we observe that, for human faces the mean shape already captures most shape information, with small deformations caused by bases. Instead, for horses the bases take the main role to represent the shape. The reason may be that human faces are relatively rigid, while horses have stronger deformations stemming from articulations, which can, to some extent, be explained by the bases of our space shape model.

Fig. 14 shows results on the CELEBA WILD dataset trained with the space shape model only consisting of a mean shape. We can see that these 3D shapes are still reasonable and captures the shape of the human faces, which confirms our statements in the last paragraph.

14. Influence of h in the Repulsion Loss L_{rep}

The repulsion loss L_{rep} defined in Eq. (9) is used to penalize keypoints within 2D space that are too close to each other. The temperature parameter h in L_{rep} controls the definition of "closeness". More specifically, the smaller the h, the less penalization for close keypoints. Thus small h may result in a degenerate solution where all keypoints converge in a small region, while large h may result in all keypoint diverging from each other and cannot capture the detailed structure of the object. Thus the choice of h depends on the category of object, as well as the number of keypoints.

Fig. 15 gives results for various choices of h on the CELEBA WILD dataset. From the results we can conclude that too small h result in clustered keypoints and cannot cover the full object (the first row of Fig. 15); while too large values of h result in points distributing in the whole image space, thus lacking details in specific areas (the fourth and fifth rows in Fig 15).

15. Video Results

We provide additional animation videos of 3D shapes from several datasets, which visualize the 3D shapes and rotations more clearly.

16. More Qualitative Results

Fig. 16 to Fig. 23 give more qualitative results on the CELEBA WILD, CUB-200-2011, HORSE and AFHQ (several subsets) datasets.

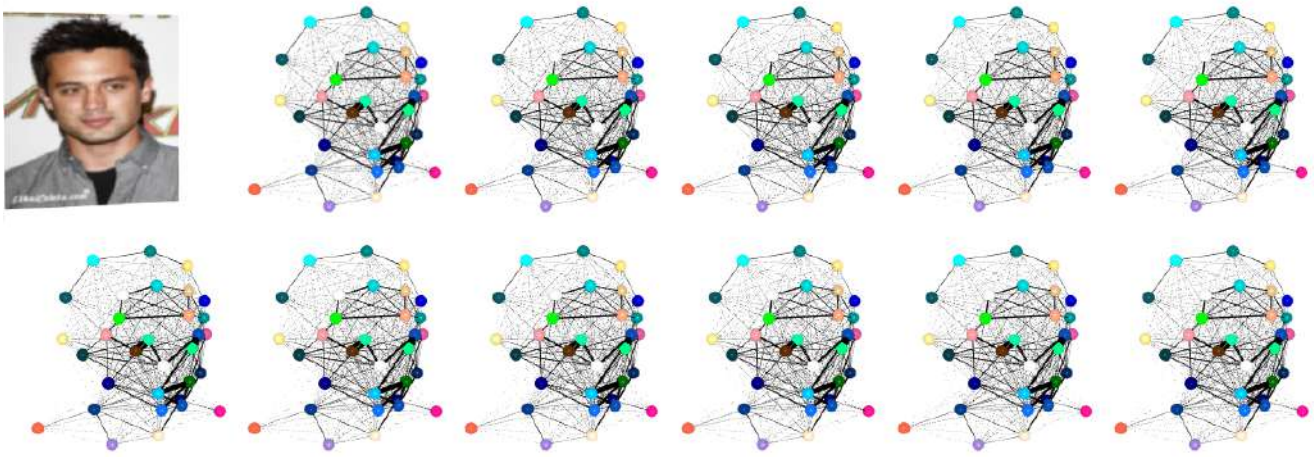

Figure 12. We visualize the effectiveness of our 3D shape space model on the CELEBA WILD dataset. From top to bottom, from left to right, we visualize the 3D shape of M, $M + \alpha_1 \times B_1$, $M + \alpha_1 \times B_1 + \alpha_2 \times B_2$, ..., until we get the full 3D shape $M + \sum_{i=1}^{n} \alpha_i \times B_i$. (Shapes are also transformed and scaled using rotation R, translation T and scaling factor s)

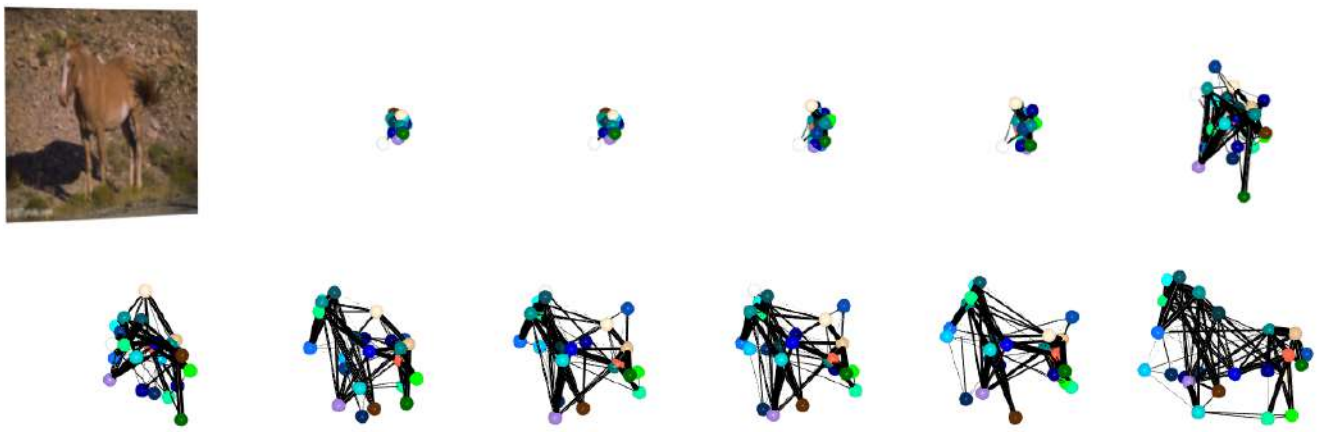

Figure 13. We visualize the effectiveness of of our 3D shape space model on the HORSE dataset. From top to bottom, from left to right, we visualize the 3D shape of M, $M + \alpha_1 \times B_1$, $M + \alpha_1 \times B_1 + \alpha_2 \times B_2$, ..., until we get the full 3D shape $M + \sum_{i=1}^{n} \alpha_i \times B_i$. (Shapes are also transformed and scaled using rotation R, translation T and scaling factor s)

Figure 14. Visualization of 3D shapes of CELEBA WILD dataset trained with 3D space shape model only consisting of mean shape.

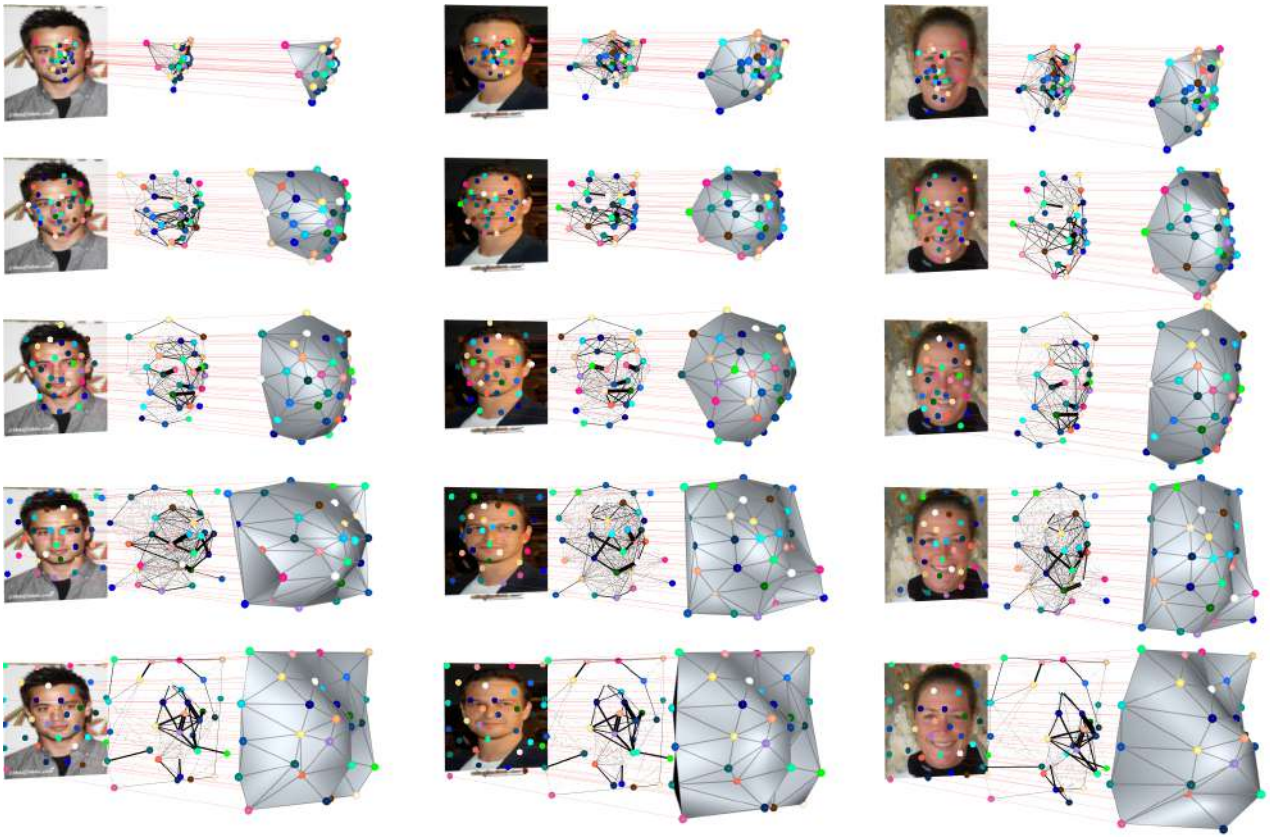

Figure 15. Visualization of 3D structure on CELEBA WILD dataset for various choice of $h = 0.05, 0.1, 0.25, 0.5, 0.75$ (from top rows to bottom rows).

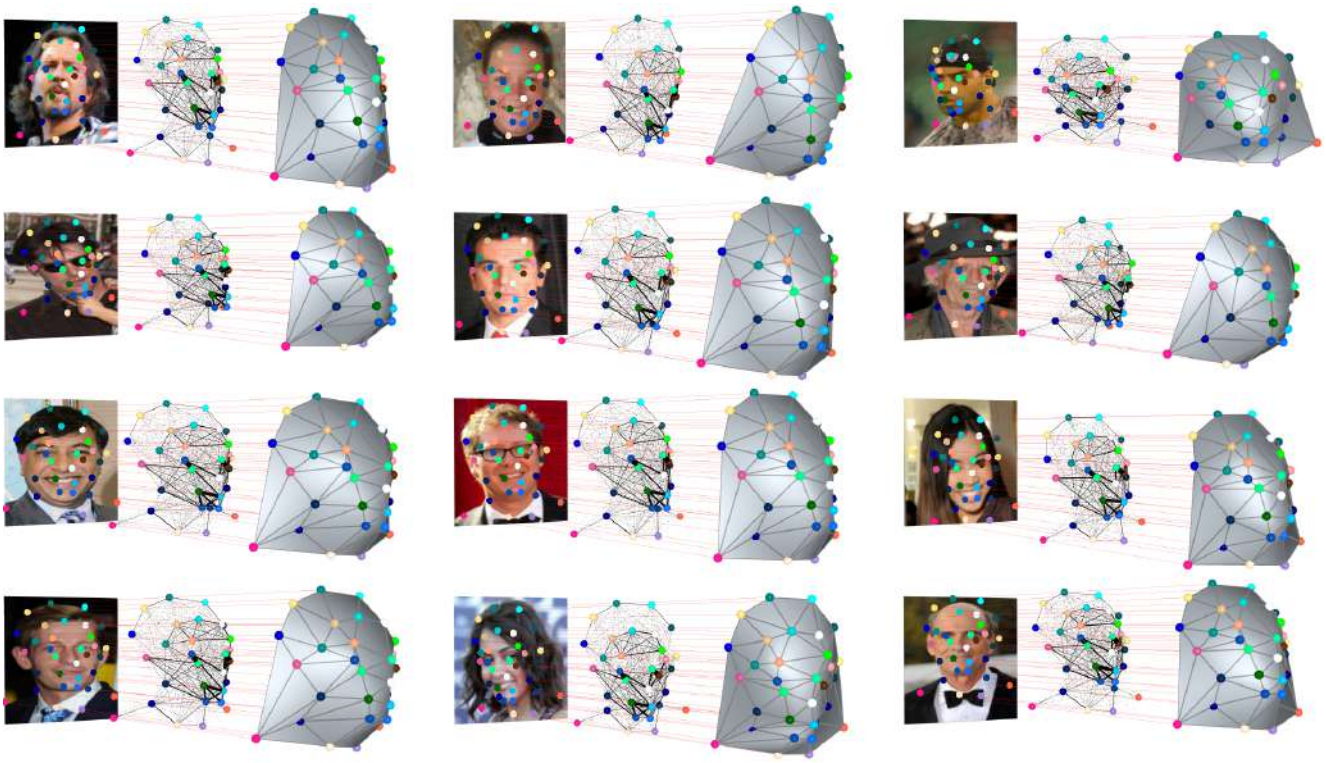

Figure 16. Qualitative results on CELEBA WILD (keypoints number $K = 32$).

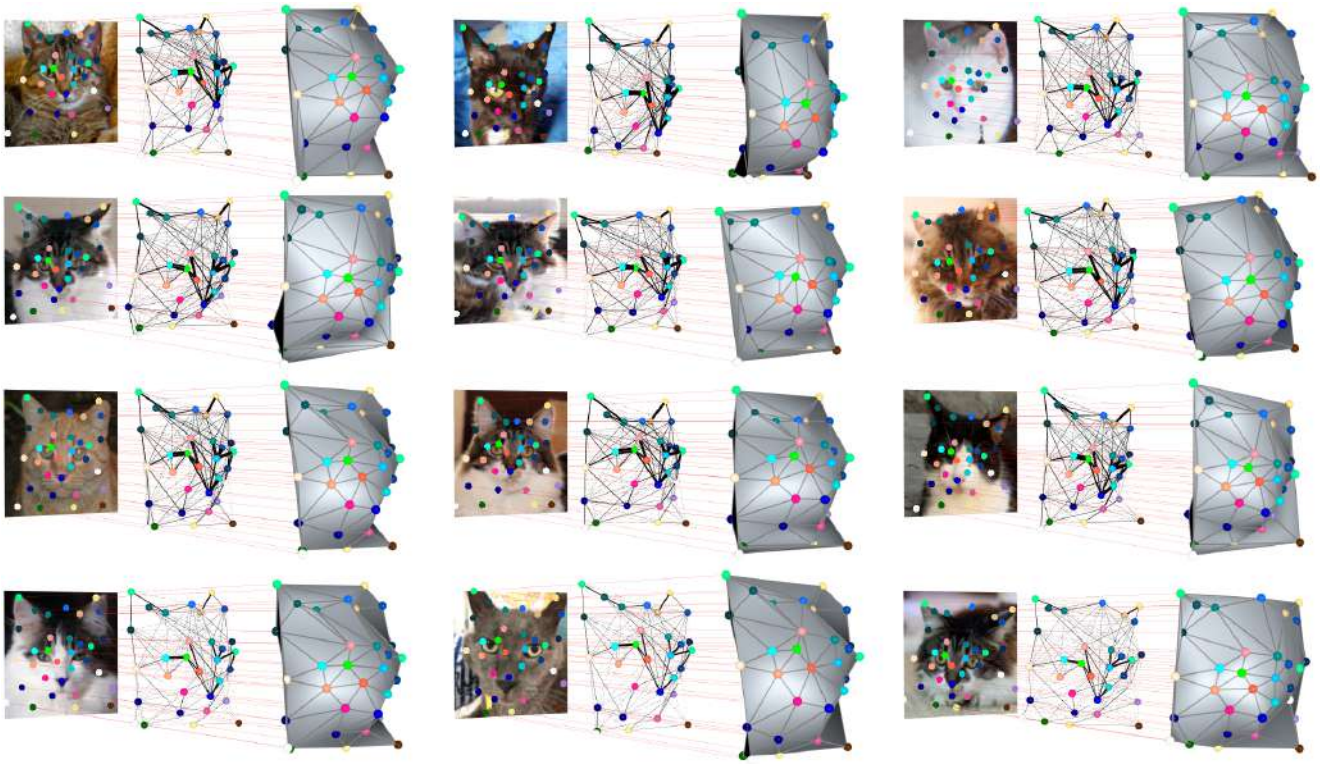

Figure 17. Qualitative results on CAT from AFHQ (keypoints number $K = 32$).

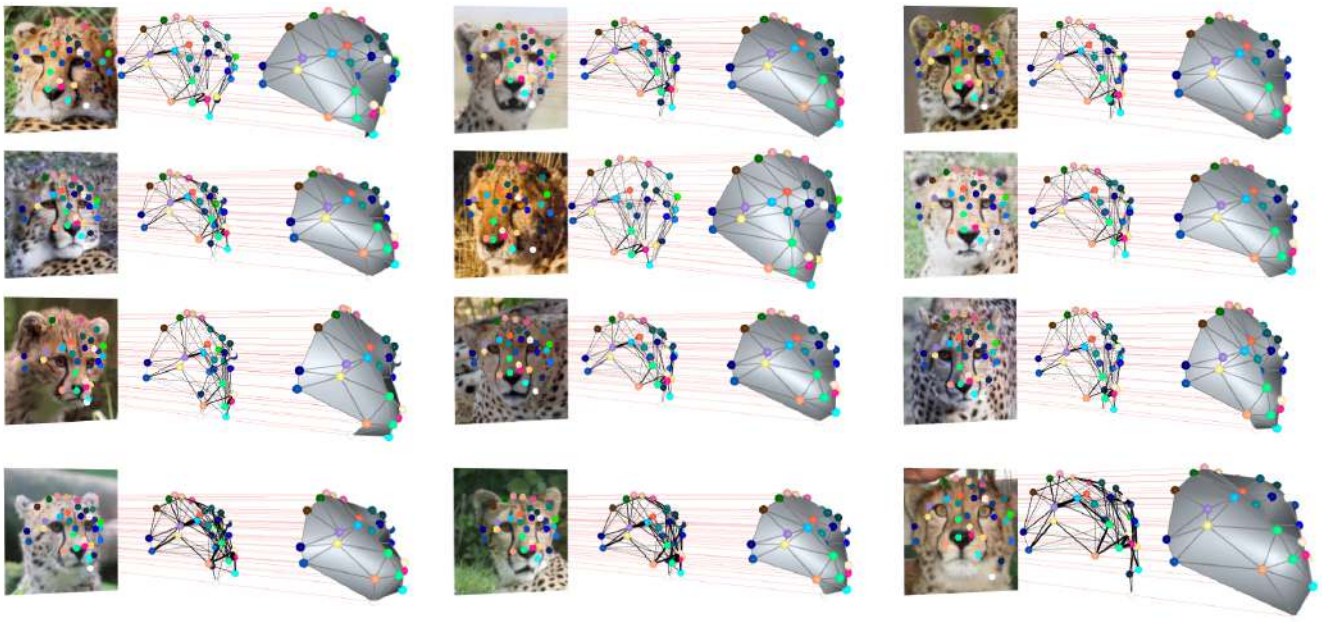

Figure 18. Qualitative results on CHEETAH from AFHQ (keypoints number $K = 32$).

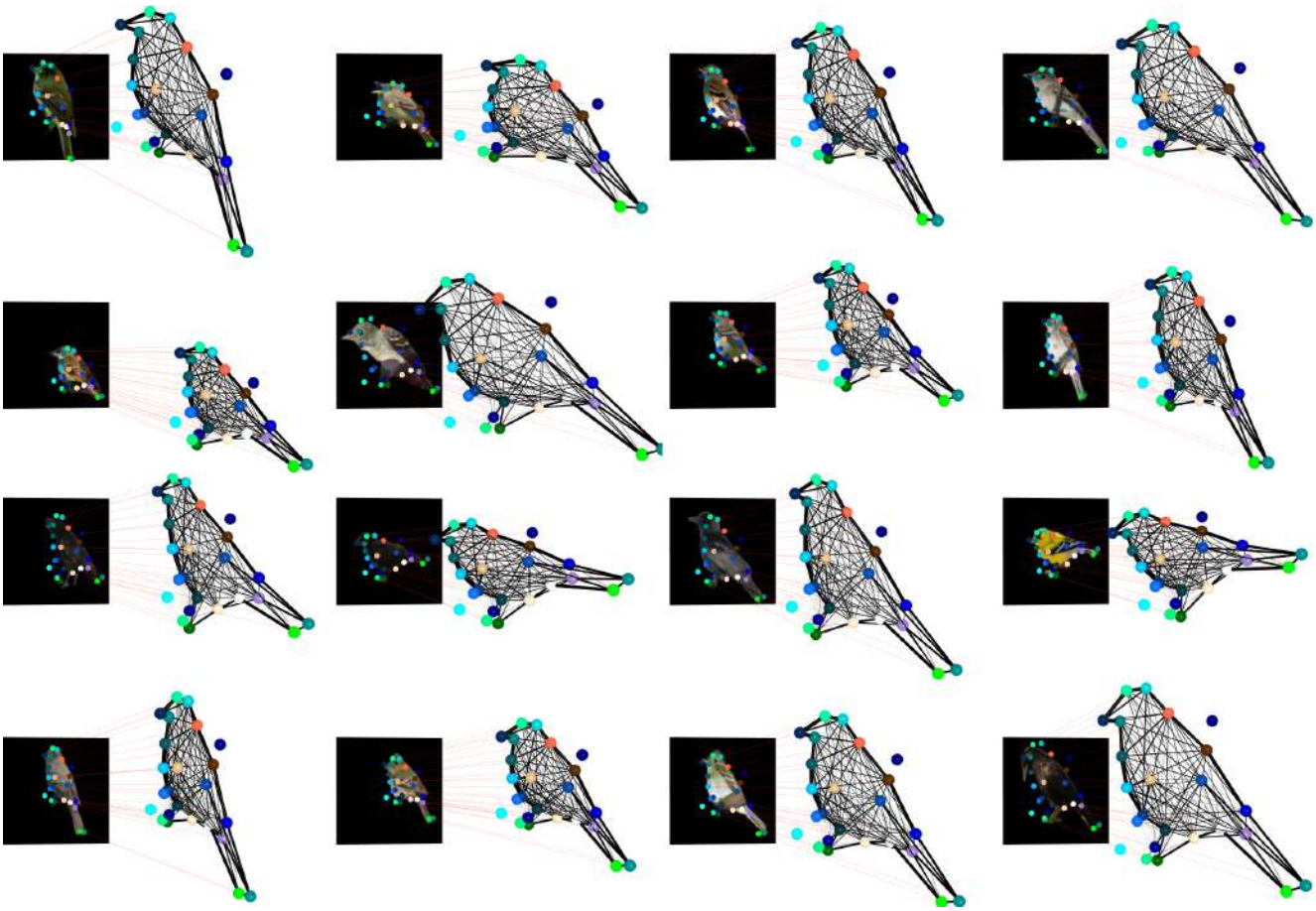

Figure 19. Qualitative results on CUB-200-2011 (keypoints number $K = 32$).

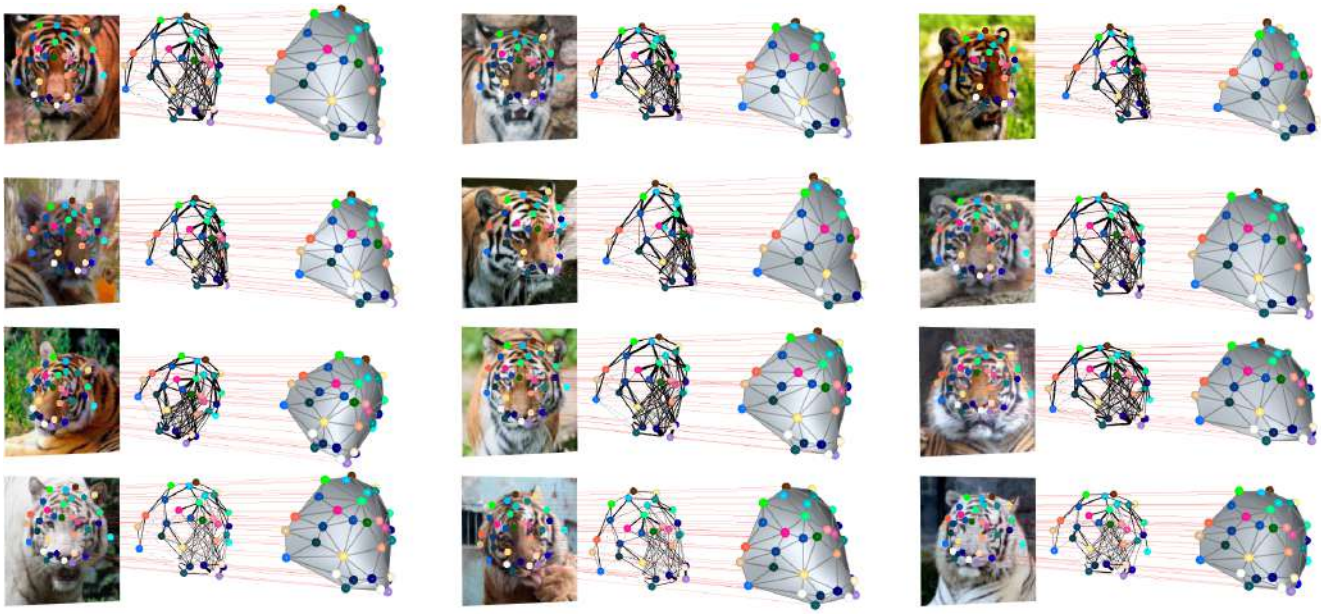

Figure 20. Qualitative results on TIGER from AFHQ (keypoints number $K = 32$).

Figure 21. Qualitative results on HORSE (keypoints number $K = 32$).

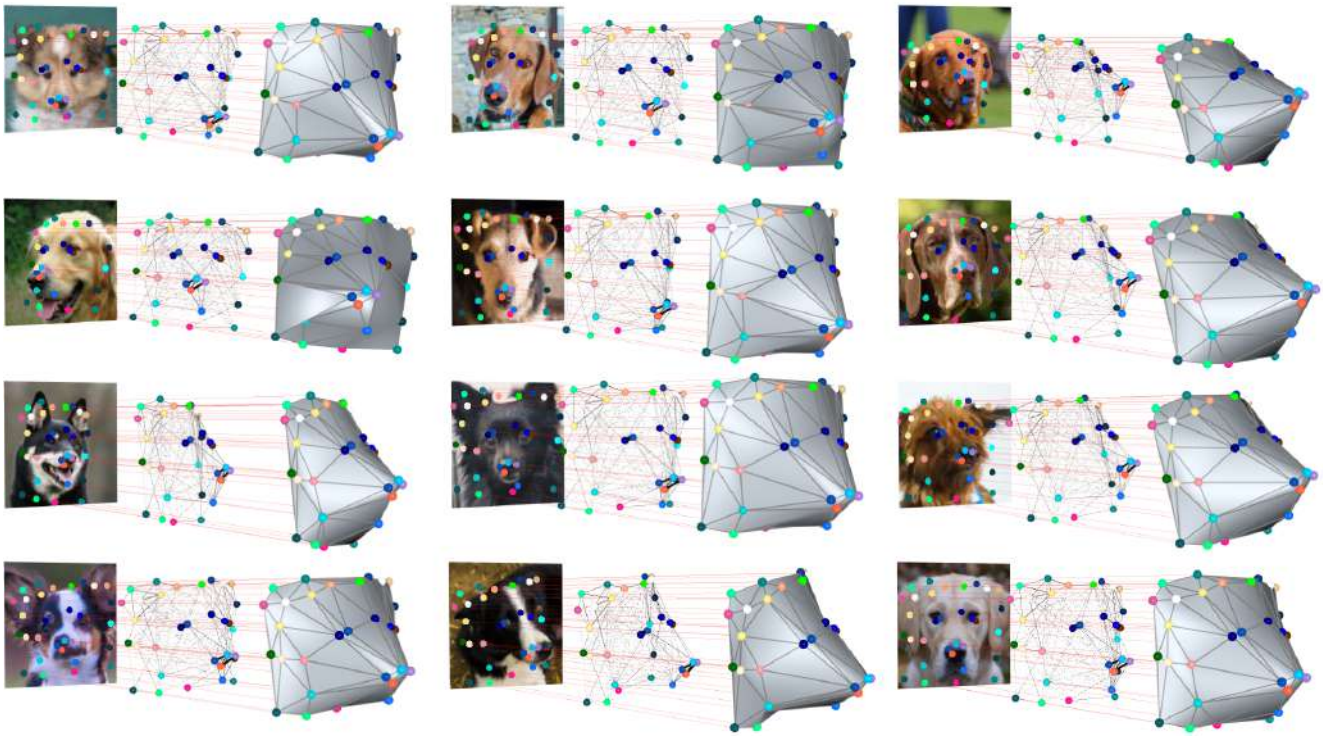

Figure 22. Qualitative results on DOG from AFHQ (keypoints number $K = 32$).

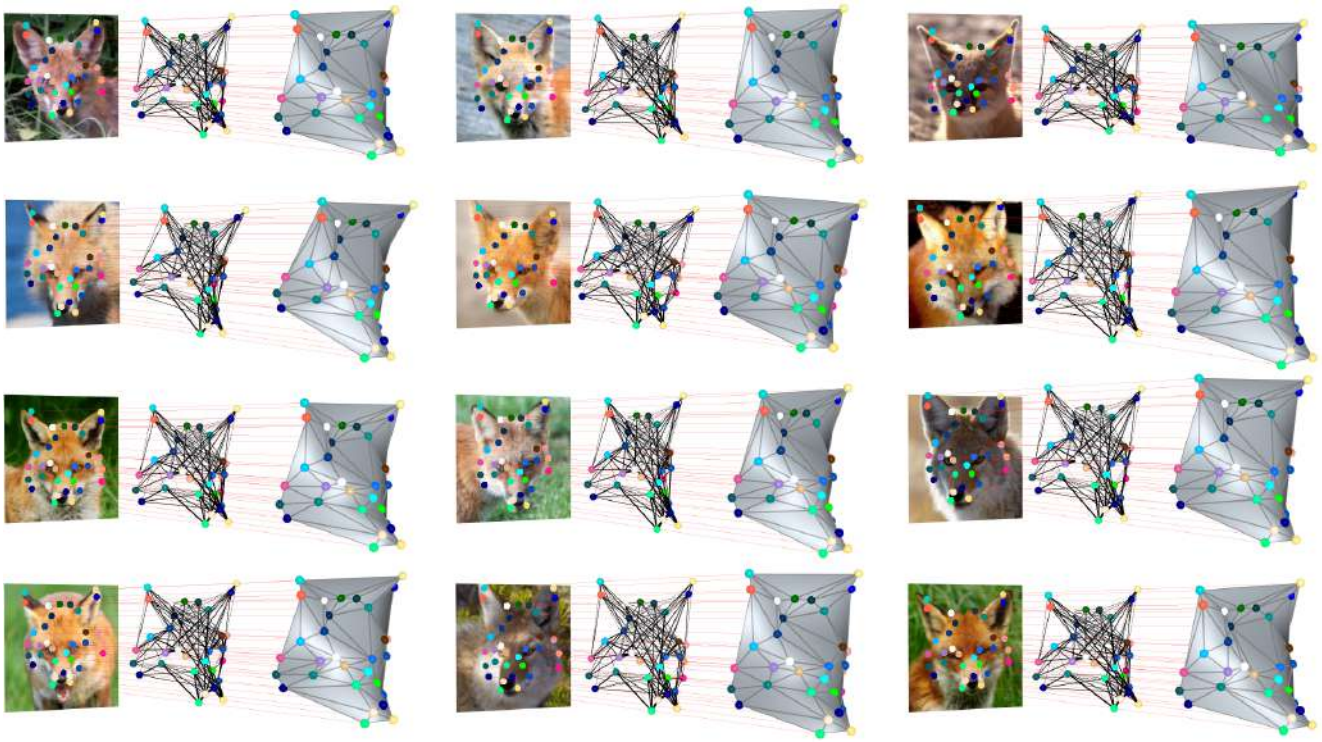

Figure 23. Qualitative results on FOX from AFHQ (keypoints number $K = 32$).