Creative Birds: Self-supervised Single-view 3D Style Transfer (Supplementary Material)

Abstract

This supplementary material consists of the following four parts: additional info regarding the loss function (section 1), a description of our evaluation interface for the user study (section 2), an experiment on DRGNet with different numbers of layers (section 3) and more comparison results (section 4).

1. Additional Info on 3D Reconstruction Loss

In shape transformation, single-view 3D reconstruction is performed. Due to space constraints, only mask loss and perceptual loss are presented in the main paper. However, as in UMR [6], we also use some other losses to improve the visual effect of the reconstructed results.

Mask loss alone is insufficient for shape reconstruction because it only provides information about a single viewpoint. Therefore, we use deformation loss [6] to enable the model properly incorporate the 3D prior from the mesh template. In addition, graph laplacian constraint [7, 4, 7] and edge regularization [10] are also employed to help smooth the reconstruct shape. In order to improve the visual effect of texture, we employ distance transform loss [4] to encourage texture flow select pixel inside the instance mask, and use texture cycle loss [6] to further optimize the position of the selected pixel. Moreover, we ensure the semantic consistency [6] by constraining the chamfer distance and the l2 distance of each semantic part and its center between the input image and the rendered image. Lastly, multiple camera hypothesis [9] is employed to avoid local minima (we used eight camera hypothesis here).

2. User Study

As stated in the "Experiments" section of our paper, we conducted a user study to compare our model to existing models. It consisted of three main components: a comparison of shape transformation (five questions about NC, DSN, KPD, NT and ours), a comparison of texture transformation (five questions about AdaIN and ours, five questions about LST and ours, five questions about EFDM and ours), and a judgement of realism (five T/F questions). Part of the evaluation interface is shown in Fig. 1.

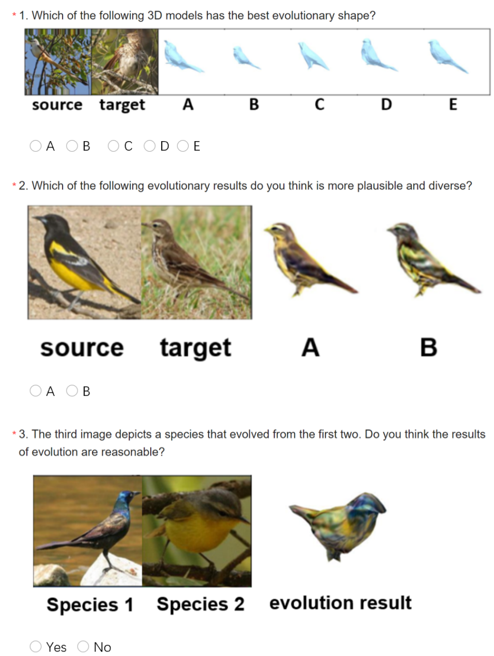

Figure 1. Evaluation Interface.

3. Additional Experiment on DRGNet

After proving the efficacy of DRGNet for feature coordination, it remains unclear how many layers of DRGNet are needed for the transformation task, so here we conducted further experiments on the number of layers of DRGNet. The results are shown in Fig. 2.

4. More Results

Here, we provide more results to help the reader evaluate the performance of our model. We show our comparison results from two aspects: 1. shape transformation (Fig. 4), 2. texture transformation (Fig. 3).

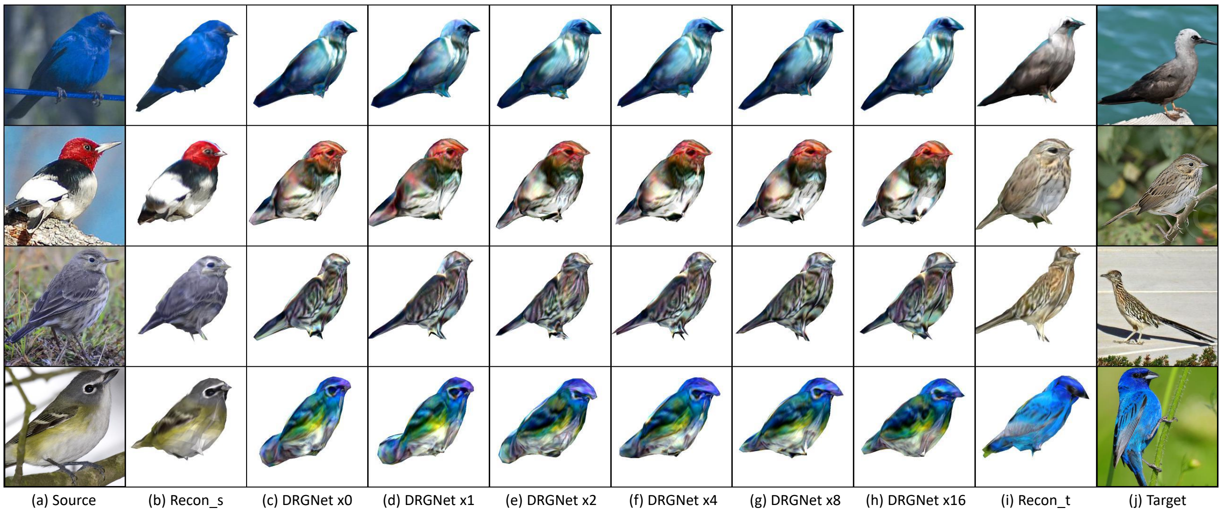

Figure 2. Shape ablation study of DRGNet with different number of layers using our SLST+SG as texture transformation method.

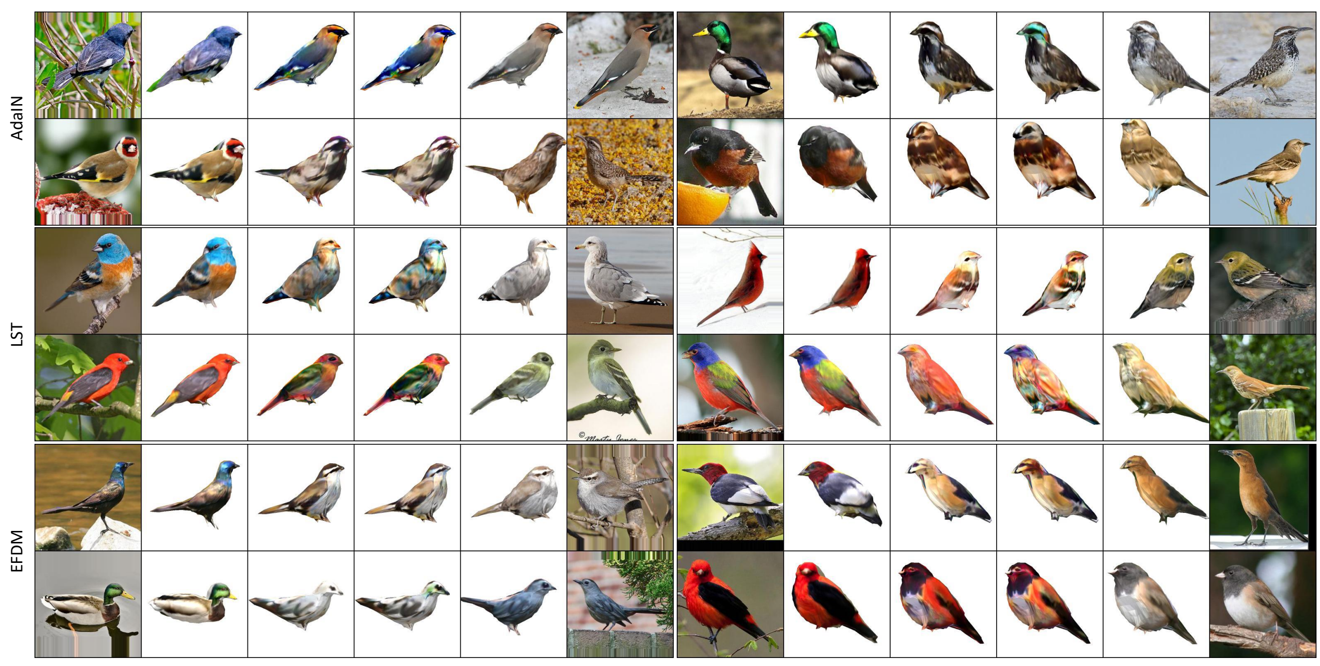

Figure 3. More visual comparisons on textures using style transfer methods (*e.g.*, AdaIN [1], LST [5]), EFDM [12] and our methods (*e.g.*, semantic UV mask (+M)).

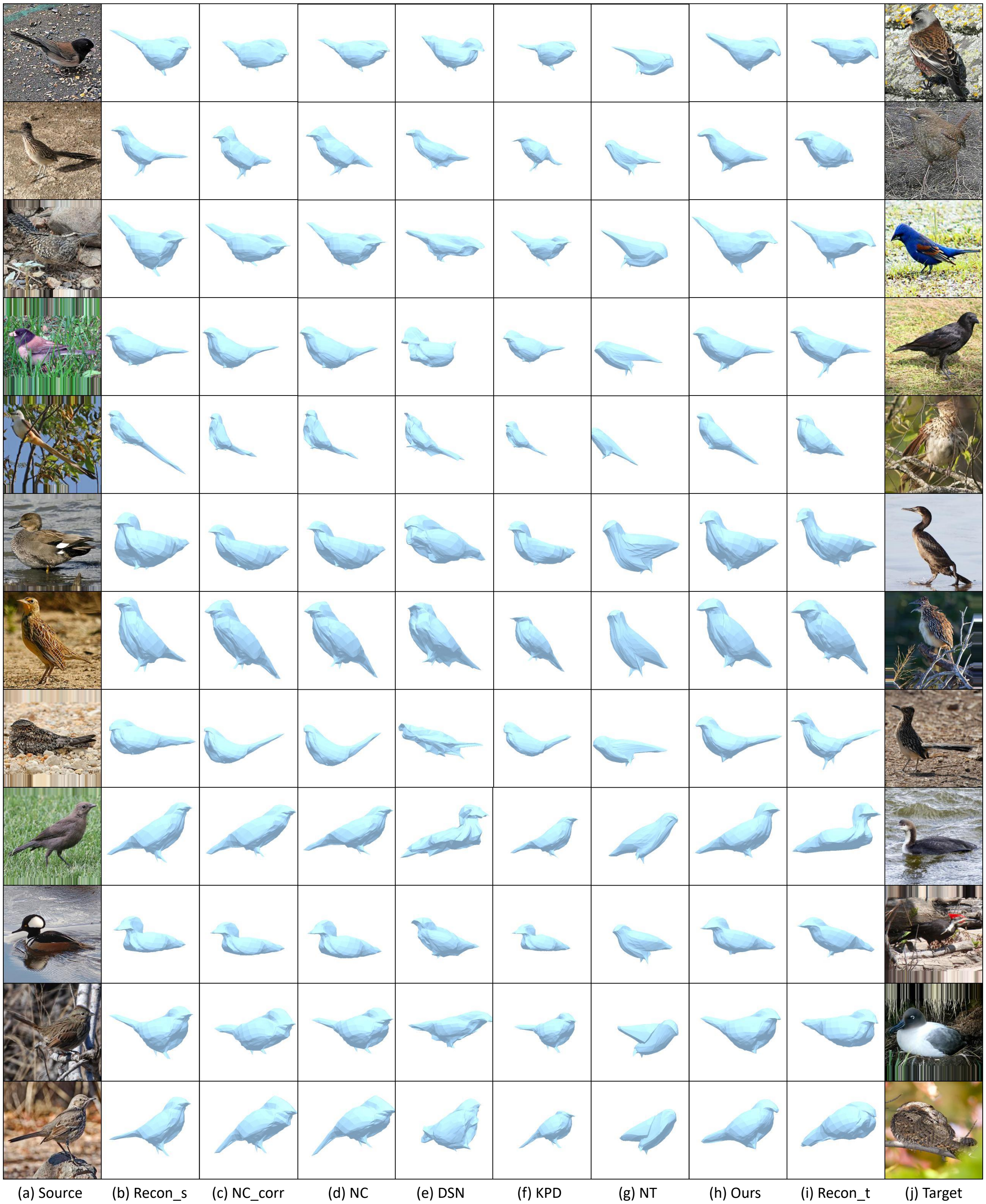

Figure 4. More visual comparisons on shape using 3D shape deformation methods (*e.g.*, NC [11], DSN [8], KPD [3], NT [2] and our method.

References

[1] Xun Huang and Serge J Belongie. Arbitrary style transfer in real-time with adaptive instance normalization. In *ICCV*, pages 1501–1510, 2017. 2

[2] Ka-Hei Hui, Ruihui Li, Jingyu Hu, and Chi-Wing Fu. Neural template: Topology-aware reconstruction and disentangled generation of 3d meshes. In *CVPR*, pages 18572–18582, 2022. 3

[3] Tomas Jakab, Richard Tucker, Ameesh Makadia, Jiajun Wu, Noah Snavely, and Angjoo Kanazawa. Keypointdeformer: Unsupervised 3d keypoint discovery for shape control. In *CVPR*, 2021. 3

[4] Angjoo Kanazawa, Shubham Tulsiani, Alexei A. Efros, and Jitendra Malik. Learning category-specific mesh reconstruction from image collections. In *ECCV*, pages 386–402, 2018. 1

[5] Xueting Li, Sifei Liu, Jan Kautz, and Ming-Hsuan Yang. Learning linear transformations for fast arbitrary style transfer. In *CVPR*, pages 3809–3817, 2019. 2

[6] Xueting Li, Sifei Liu, Kihwan Kim, Shalini De Mello, Varun Jampani, Ming-Hsuan Yang, and Jan Kautz. Self-supervised single-view 3d reconstruction via semantic consistency. In *ECCV*, pages 677–693, 2020. 1

[7] Shichen Liu, Tianye Li, Weikai Chen, and Hao Li. Soft rasterizer: A differentiable renderer for image-based 3d reasoning. In *ICCV*, pages 7708–7717, 2019. 1

[8] Minhyuk Sung, Zhenyu Jiang, Panos Achlioptas, Niloy J. Mitra, and Leonidas J. Guibas. Deformsyncnet: Deformation transfer via synchronized shape deformation spaces. *ACM Trans. Graph*, 39(6):Article 262, 2020. 3

[9] Shubham Tulsiani, Alexei A Efros, and Jitendra Malik. Multi-view consistency as supervisory signal for learning shape and pose prediction. In *CVPR*, pages 2897–2905, 2018. 1

[10] Nanyang Wang, Yinda Zhang, Zhuwen Li, Yanwei Fu, Wei Liu, and Yu-Gang Jiang. Pixel2mesh: Generating 3d mesh models from single rgb images. In *ECCV*, pages 52–67, 2018. 1

[11] Yifan Wang, Noam Aigerman, Vladimir G Kim, Siddhartha Chaudhuri, and Olga Sorkine-Hornung. Neural cages for detail-preserving 3d deformations. In *CVPR*, pages 75–83, 2020. 3

[12] Yabin Zhang, Minghan Li, Ruihuang Li, Kui Jia, and Lei Zhang. Exact feature distribution matching for arbitrary style transfer and domain generalization. In *CVPR*, pages 8035–8045, 2022. 2