3DCaricShop: A Dataset and A Baseline Method for Single-view 3D Caricature Face Reconstruction
– *Supplemental Material* –

Yuda Qiu[1] Xiaojie Xu[1] Lingteng Qiu[1] Yan Pan[1]
Yushuang Wu[1] Weikai Chen[2] Xiaoguang Han[1,*]
[1]SRIBD, The Chinese University of Hong Kong, Shenzhen[†] [2]Tencent Game AI Research Center

1. Structure of G2L Network

In this section, we illustrate the structure of the Global to Local (G2L) moduleof VC-GCN. As shown in Fig. 1, the outputs of local-view GCN $\mathbf{X}_l$ and those of global-view GCN $\mathbf{X}_g$ are fed into G2L network. First, we change the channels of global and local features ($\hat{\mathbf{X}}_g$ and $\hat{\mathbf{X}}_l$ respectively) by local-GCN and global-GCN, of which the structures are the same as that employed in the whole pipeline. Then, trainable G2L weights $\mathbf{W}$ are obtained by matrix multiplication between $\hat{\mathbf{X}}_g$ and $\hat{\mathbf{X}}_l$, followed by a softmax operation. Finally, we get the updated local-view features $\hat{\mathbf{Z}}_l$ processed by the following formula:

$$\hat{\mathbf{Z}}_l = \mathbf{W} \bigotimes \mathbf{X}_l + \mathbf{X}_l, \tag{1}$$

where $\bigotimes$ means matrix multiplication. In Fig. 1, $\mathbf{B}$ is the batch size of inputs. N_l and N_g represent the number of nodes in the local and global graph, with C and C_1 defined as the number of feature channels. Empirically, C_1 is set to 32, considering the trade-off between efficiency and accuracy.

2. More Qualitative Results

Fig. 2 shows that the reconstruction results using our proposed 3D landmark localization approach could capture the large exaggerations more accurately than other settings. For example, the long chin of the second sample is not distorted. Fig. 3 shows the necessity of landmark-guided registration. Without 3D landmarks, the outputs of NICP [1] fail to fit the accurate shape of the face, and PCA [2] projection helps to further reduce artifacts in the final results.

We show a failure case in Fig. 4 where the estimated normal map is blurry, especially at the mouth region, that leads to inaccurate result.

We present more visual results in Fig. 5 to show the effectiveness of our framework. In addition, We show more qualitative results for ablation studies on the framework.

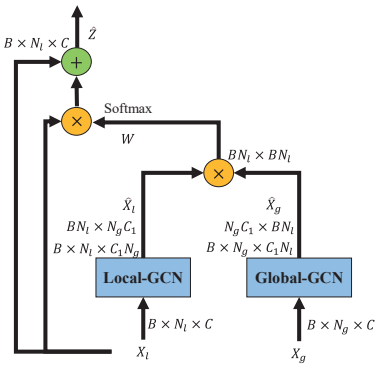

Figure 1: The pipeline of Global to Local(G2L) module in VC-GCN. $\mathbf{X}_l$ means local-view features and $\mathbf{X}_g$ represents global-view features. $\bigotimes$ is matrix multiplication and $\bigoplus$ stands for the pixel-wise addition. The trainable global-to-local weights are defined as W while $\hat{\mathbf{X}}_l$ and $\hat{\mathbf{X}}_g$ represents the outputs of local-GCN and Global-GCN respectively. $\hat{Z}$ is the updated local features fused with global ones. B is the batch size of inputs. N_l and N_g represent the number of nodes of local and global graph, with C defined as the number of feature channels.

3. Applications

The proposed framework generates caricature meshes with uniform topology. With the well-defined topology, various applications could be developed. In Fig. 6 we present the mesh generation via interpolating among the predict caricature meshes.

In Fig. 7, we compare the rigging results with AliveCaric-DL (ADL). Both results are animated using the same skeleton and skinning weights for fair comparison. We show that our method supports faithful rigging of our results and preserves better geometric details than ADL.

[1]**Corresponding email:** hanxiaoguang@cuhk.edu.cn
[2]Shenzhen Research Institute of Big Data

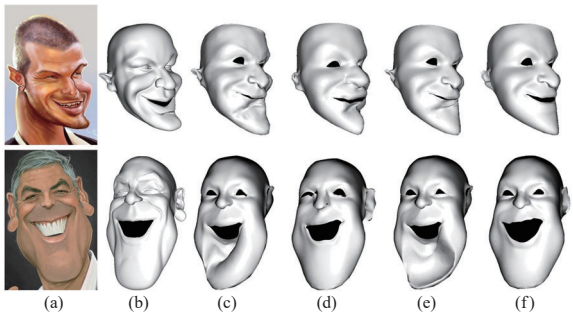

Figure 2: Visualized reconstruction results using different setting of 3D landmark detection, with (a) input; (b) GT; (c) pure projection; (d) global only; (e) w/o G2L; (f) ours. It demonstrates that our method could capture the face geometry more accurately.

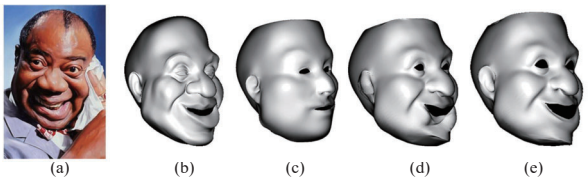

Figure 3: Visualized reconstruction results on different setting of registration: (a) input; (b) GT; (c) NICP w/o landmark; (d) NICP w/o PCA projection; (e) ours. It demonstrates that a better template is obtained with finer details (e.g., the nose) by using PCA projection.

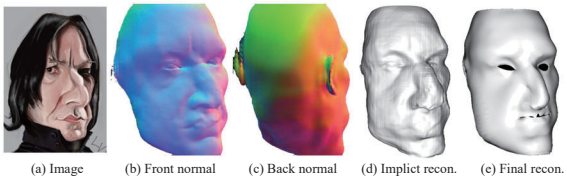

Figure 4: A Failure case. The normal is broken on the ear and blur on the mouth, generating a low quality reconstruction.

References

[1] Brian Amberg, Sami Romdhani, and Thomas Vetter. Optimal step nonrigid icp algorithms for surface registration. In *2007 IEEE Conference on Computer Vision and Pattern Recognition*, pages 1–8. IEEE, 2007.

[2] Volker Blanz and Thomas Vetter. A morphable model for the synthesis of 3d faces. In *Proceedings of the 26th annual conference on Computer graphics and interactive techniques*, pages 187–194, 1999.

Figure 5: Results gallery of the proposed framework on 3DCaricShop. The framework has the capability to reconstruct 3D shapes from caricature images with diverse texture and geometry shapes.

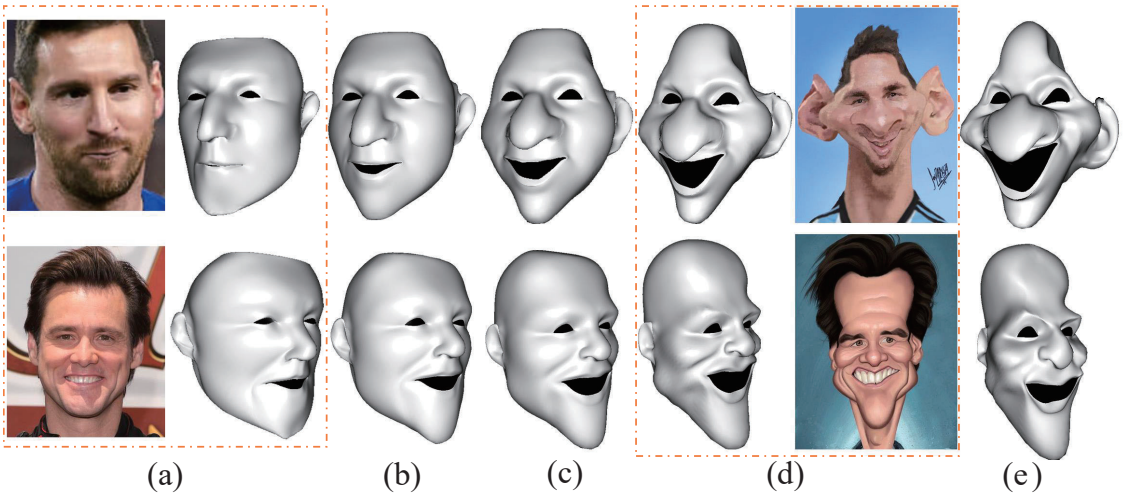

Figure 6: Samples of interpolation application of our method. Thanks to the uniform topology, it's feasible to generate novel caricature shapes by interpolating the predicted meshes among different inputs. As shown in the figure, (a) are the input real photos and the corresponding reconstructed meshes, (d) are the input caricature images and the generated meshes. (b)(c) are the interpolation results for the meshes from (a)(d). Also, we can perform extrapolating between the meshes (a)(d) to create more exaggerated results, as shown in the last column (e).

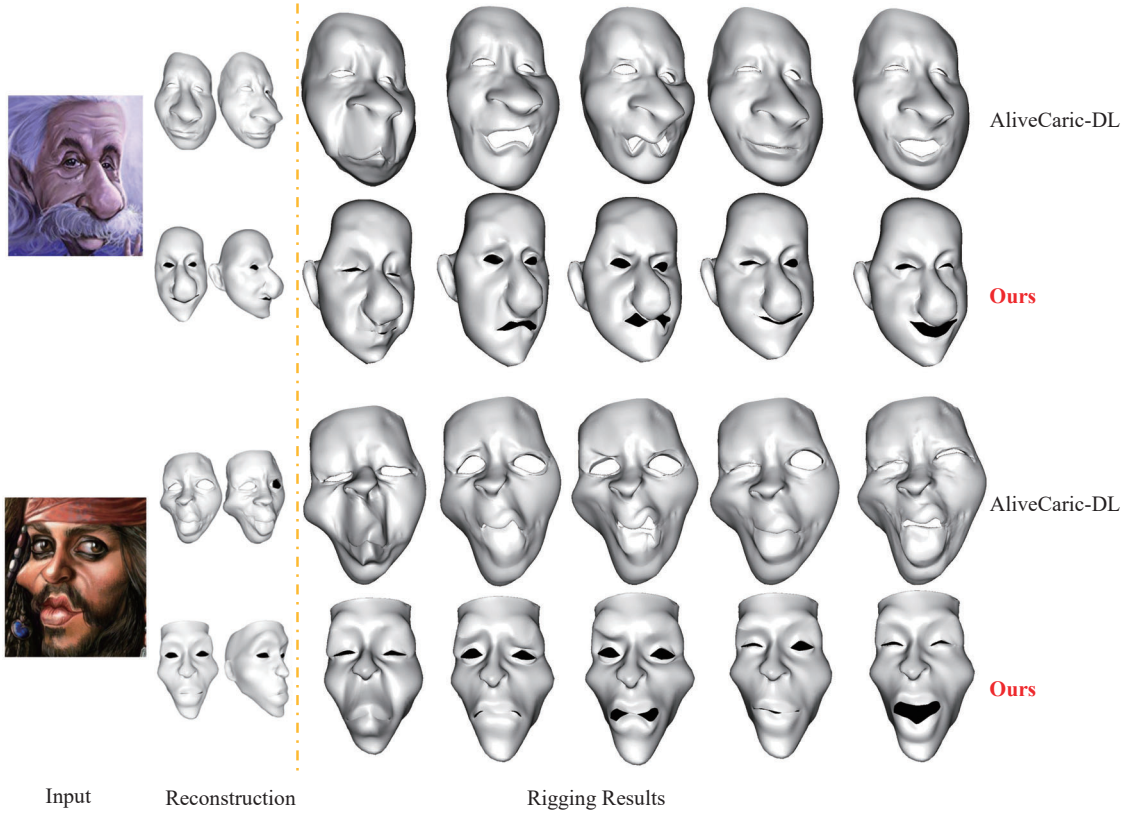

Figure 7: Rigging samples.