Exemplar-Based 3D Portrait Stylization: Supplemental Material

In this supplemental document, we validate the effectiveness of some designs in our method through ablation studies, present an additional application with realistic style reference, and provide more results of our stylization method.

1 Two Stages v.s. Single stage

To show that the two-stage framework is effective in learning both geometry and texture styles, we design an ablation study to compare it with the single-stage framework, in which we do not have a separate geometry style transfer stage and let the geometry style be learned jointly with the texture styles. That means during the multi-view optimization, the style transfer objectives will be back-propagated to variables v_z and t_z via the differentiable renderer and update them. However, the freedom in updating two variables will also cause some ambiguities: whether the vertex should be moved or the texture map's color should be changed to simulate the reference style. Thus, it leads to obvious artifacts on both geometry and texture, as shown in Fig. 1.

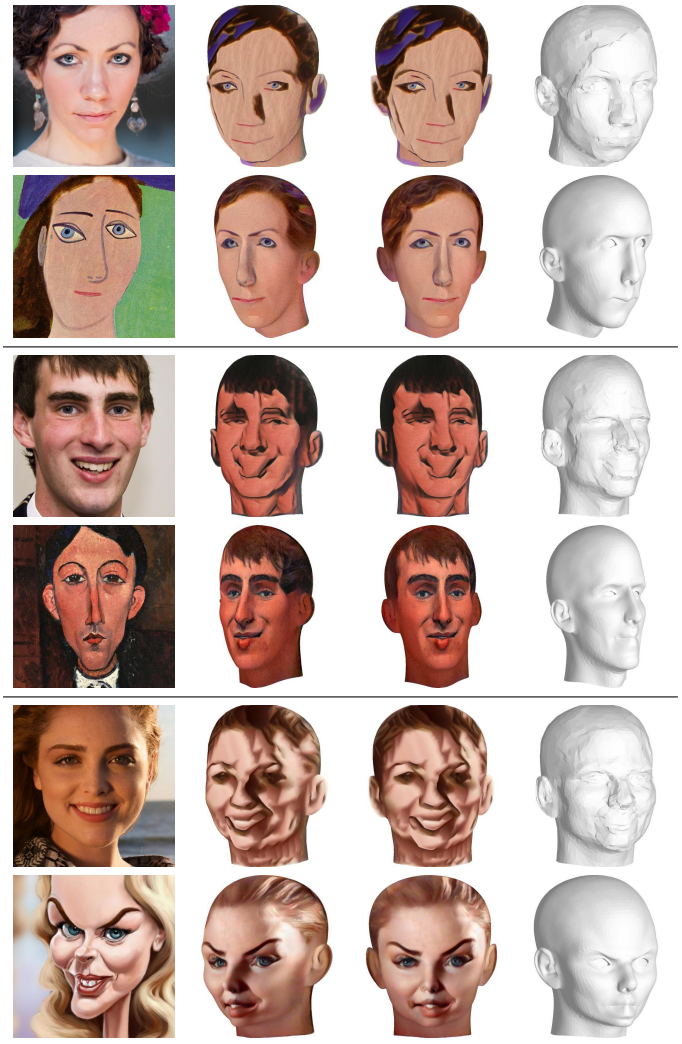


Fig. 1: **Comparison between single-stage (upper row) and two-stage (lower row) frameworks.**

2 Ablation on Loss Terms in Landmark Translation

In this section, we ablate the loss terms (besides the classification loss already given in the paper) in Eq. 7 for training our multimodal landmark-to-landmark translation network. The qualitative and quantitative comparisons are respectively given in Fig. 2 and Table. 1.

As shown in Fig. 2, without the reconstruction loss on landmarks L^Y_{recon}, there is no guarantee that it is a valid disentanglement of content and style, which makes the translation unstable. Some results are subtly deformed, while some have strange distortions and even fail to preserve the face structures. Without the content reconstruction loss L^C_{recon}, almost no content information, such as the expression and the pose, is retained because of the degeneration of the content encoder. Similarly, the style encoder degenerates without the style reconstruction loss L^S_{recon}, resulting in the loss of style information. Thus, the results fall into the normal face domain. Without the KL divergence loss L_{KL}, the styles stay the same for different style inputs, as expected, as stated in the main text, due to the degeneration of the style encoder and the invalid disentanglement. Without the adversarial loss L_{adv}, the translation to the artistic domain fails, and the results also fall into the normal face domain.

The results in Table. 1 also reflect that removing any of these terms will significantly degrade the result quality. To conclude, all the loss terms in Eq. 7 in the main text are necessary for this network to achieve the full performance.

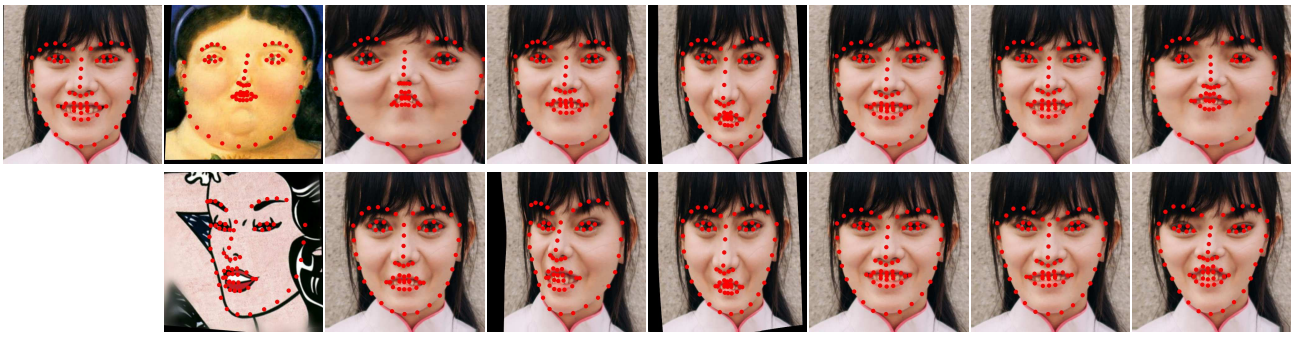


Fig. 2: Ablations on loss terms in the landmark-to-landmark translation. We show the landmark translation results by training versions without each terms in the Eq. 7 of the main text.

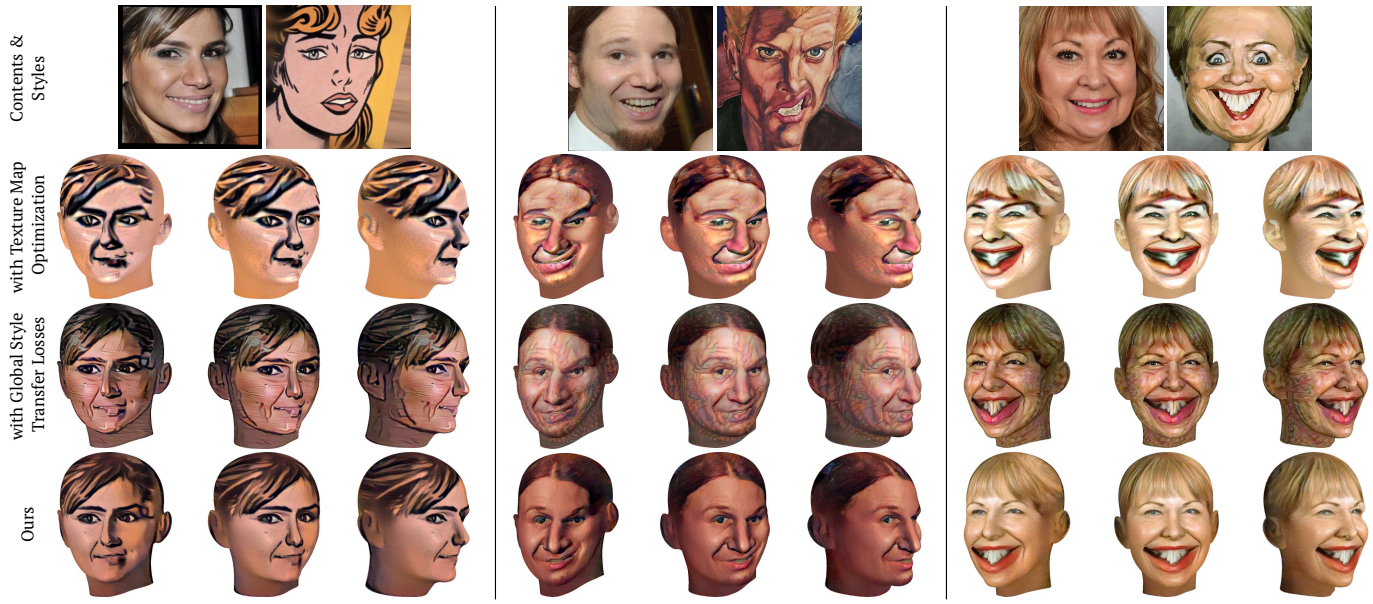


Fig. 3: **Comparisons on texture stylization.** We compare our results using multi-view optimization and local style transfer losses with the results using single texture map optimization and the results using global style transfer losses.

methods	w/o L^Y_{recon}	w/o L^C_{recon}	w/o L_{KL}	w/o L^S_{recon}	w/o L_{adv}	Ours
FIDs	25.771	28.331	34.920	47.693	46.096	23.587

TABLE 1: FID comparison for ablations on the landmark-to-landmark translation without different loss terms.

3 Comparison on Deformation Methods

In this section, we compare the deformation method used in the paper, the Laplacian deformation, with two other deformation methods, an adapted version [1] of As-Rigid-As-Possible (ARAP) [2] and the Biharmonic deformation [3], both implemented by a public library, libigl [4]. Some results are shown in Fig. 4. From the results, it can be seen that ARAP cannot handle this specific task well. This is because ARAP aims to perform the deformation with rigid transformations as much as possible, while given the landmark constraints from translation, the deformation is required to be non-rigid to maintain the smoothness of the face, such as the inflation of the face in the first example in Fig. 4.

Nonetheless, the deformation method can be replaced with other methods, and they can obtain comparable results with the landmark constraints. In Fig. 4, Biharmonic deformation is given as an example, and its results show similar geometric variation while keeping the surface smooth. Meanwhile, we agree that using more powerful deformation methods would potentially improve the final results, which can be studied in the future.

4 Application: 3D Realistic Style Transfer

Although the style transfer between photo-realistic images is beyond our focus, our method is able to obtain reasonable results with geometry maintaining unchanged from the input content and texture style transferred from the

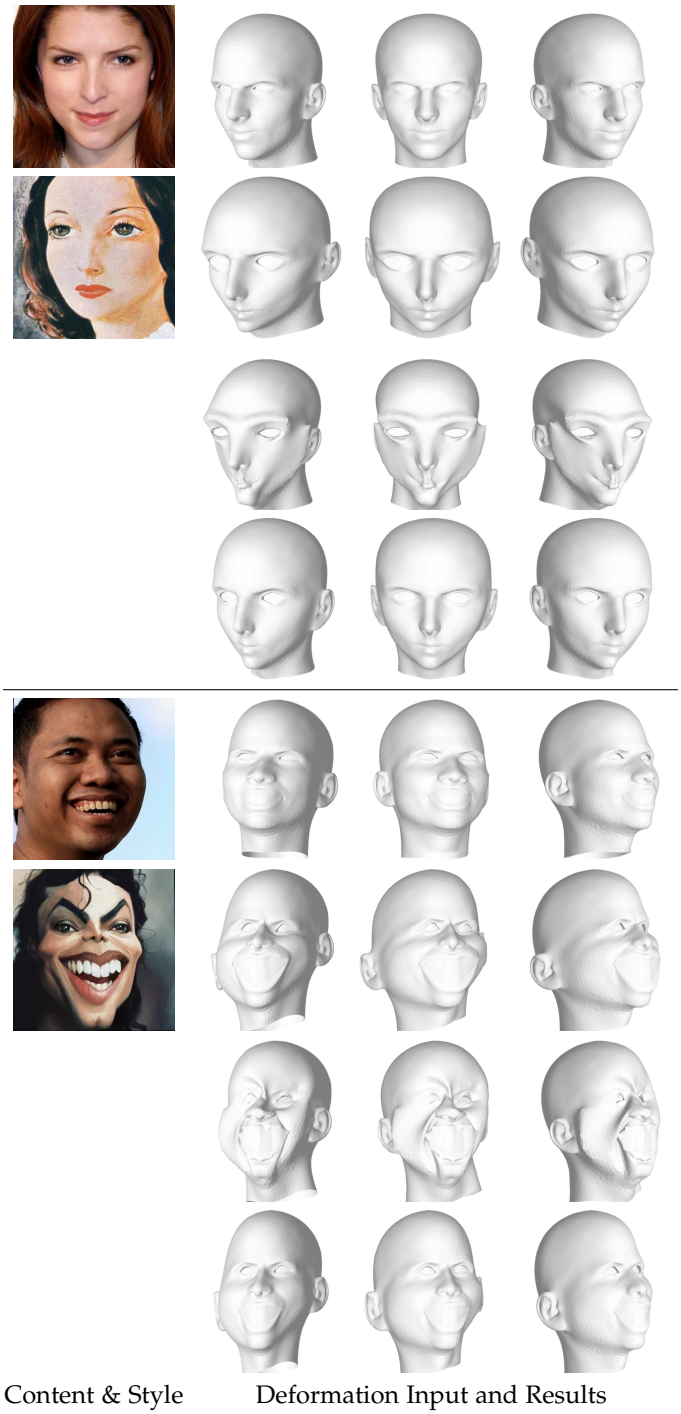


Fig. 4: **Comparison among different deformation methods.** First row: deformation input, second row: Laplacian deformation (ours), third row: As-Rigid-As-Possible, fourth row: Biharmonic deformation.

realistic style reference. We have tested our method on several typical examples, and the results are given in Fig. 5. As discussed in Sec. 8 of the paper, the geometric deformations in the results are not obvious since the landmark-to-landmark translation network is trained to learn the artistic exaggeration from the style. If the style landmarks are also in the normal face domain, the content landmarks will remain unchanged. For the texture style transfer, the facial

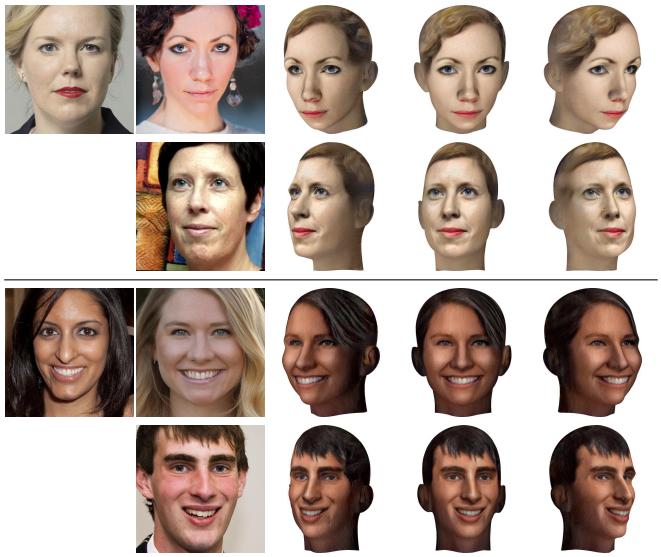


Fig. 5: **Results with real images as style reference.**

attributes such as makeup, skin tone, and hair color can be successfully transferred, and the result images maintain to be realistic. However, it is necessary to point out that the artistic style transfer does not naturally preserve the local structures of the content, and sometimes local texture distortions may occur. A further solution to prevent such a problem is to replace texture transfer losses with color transfer losses, which constrain the transfer operation to happen only in color space and would be more suitable for the photo-realistic cases.

5 Ablation on Texture Style Transfer

Once the geometry style is transferred in the first stage, we have two options to transfer the texture style. The first one is to apply image style transfer on the texture map, and the second one is to optimize the texture map via the differential render in the multi-view framework. We compare the results of these two options in Fig. 3. With the first option, the texture has lots of artifacts because the texture map, which is often a severely distorted face, will cause misalignment and texture pattern distortions in rendered results. The second option, our choice, can avoid these problems and achieve much better texture quality by rendering the 3D model into normal face images for style transfer. And the multi-view optimization framework can seamlessly combine the multi-view transfer results into a full texture map.

Next, we ablate the choice for texture style transfer loss. As claimed, the texture transfer loss is inherited from neural image style transfer methods, which have two major categories. One is to match the global statistics of deep features, while the other is to build pixel-wise feature correspondences and transfer the local statistics. The first category's representative method is Gatys [5] and the STROTSS [6] represents the state of the art of the second category. We apply the loss functions from these two methods, respectively, in our multi-view optimization and show the comparison results in Fig. 3. Since our style transfer focuses on faces

with strong semantic correspondences, the STROTSS loss, our choice, leads to more visually pleasing results, where the style of a local part can be precisely transferred to its corresponding part.

6 Perceptual Study Questionnaire

We show the 16 groups of questions used in the perceptual study along with the content recognition rate (CR) and the style recognition rate (SR) of each of the groups in Fig. 6.

References

[1] I. Chao, U. Pinkall, P. Sanan, and P. Schröder, "A simple geometric model for elastic deformations," *ACM transactions on graphics (TOG)*, vol. 29, no. 4, pp. 1–6, 2010.

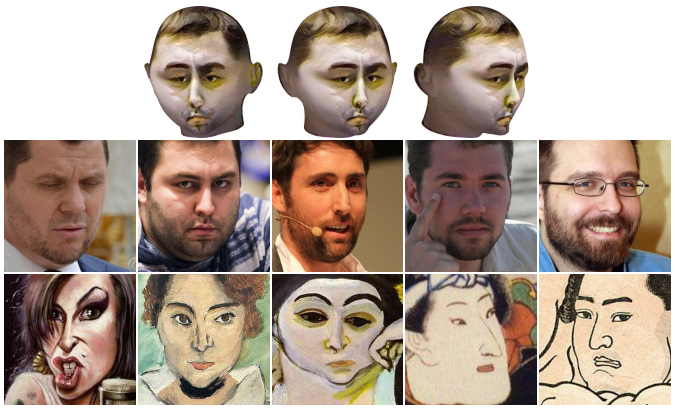
[2] O. Sorkine and M. Alexa, "As-rigid-as-possible surface modeling," in *Symposium on Geometry processing*, vol. 4, 2007, pp. 109–116.

[3] B. T. Helenbrook, "Mesh deformation using the biharmonic operator," *International journal for numerical methods in engineering*, vol. 56, no. 7, pp. 1007–1021, 2003.

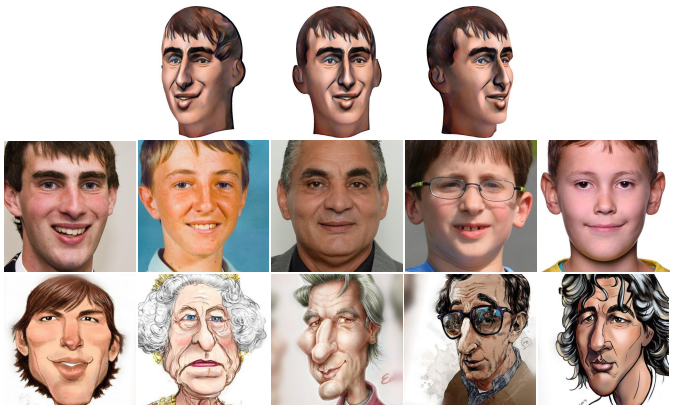
[4] A. Jacobson, D. Panozzo *et al.*, "libigl: A simple C++ geometry processing library," 2018, https://libigl.github.io/.

[5] L. A. Gatys, A. S. Ecker, and M. Bethge, "A neural algorithm of artistic style," *CoRR*, vol. abs/1508.06576, 2015.

[6] N. I. Kolkin, J. Salavon, and G. Shakhnarovich, "Style transfer by relaxed optimal transport and self-similarity," in *CVPR 2019*, 2019, pp. 10 051–10 060.



Group 1, CR: 96.55% , SR: 72.41%

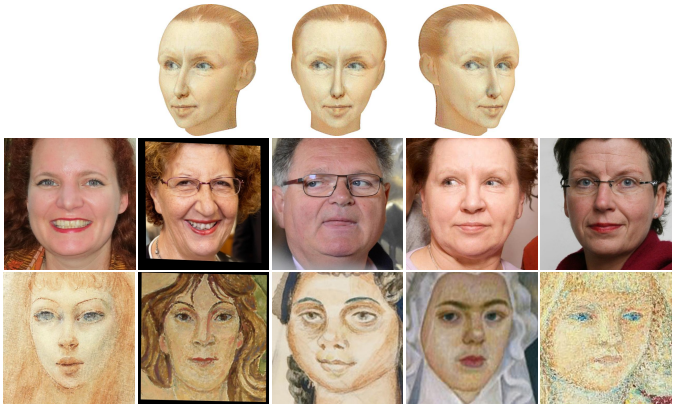


Group 2, CR: 100.00% , SR: 51.72%

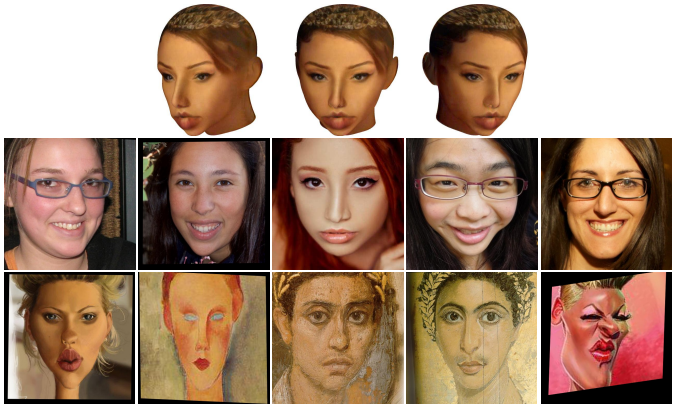
Group 3, CR: 86.21% , SR: 100.00%

Group 4, CR: 96.55% , SR: 93.10%

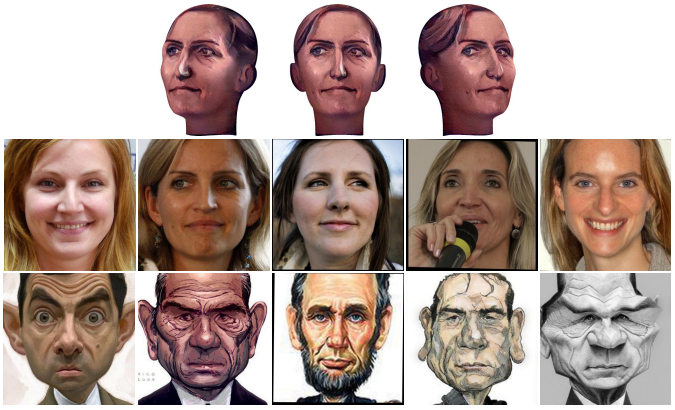
Group 5, CR: 93.10% , SR: 75.86%



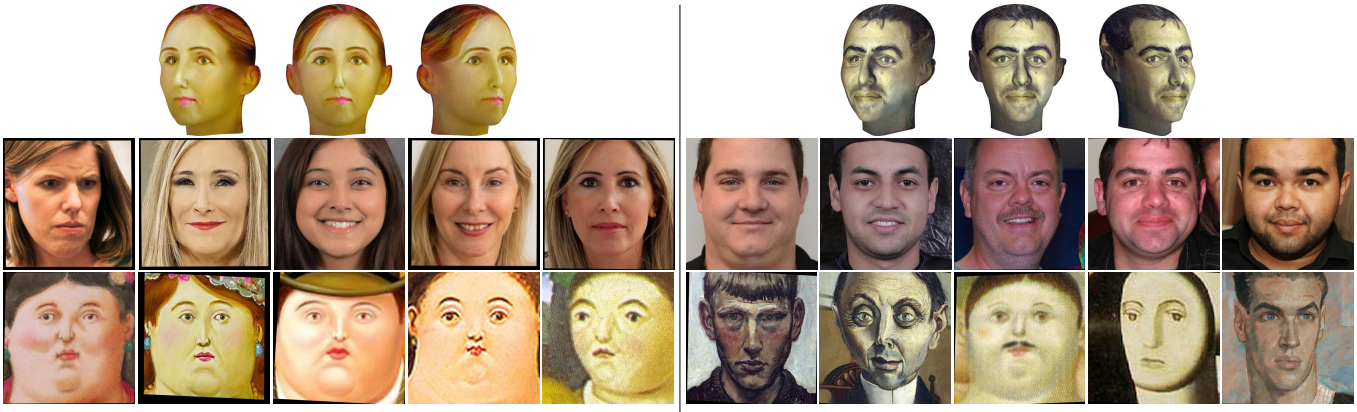
Group 6, CR: 72.41% , SR: 100.00%



Group 7, CR: 86.21% , SR: 82.76%



Group 8, CR: 96.55% , SR: 96.55%

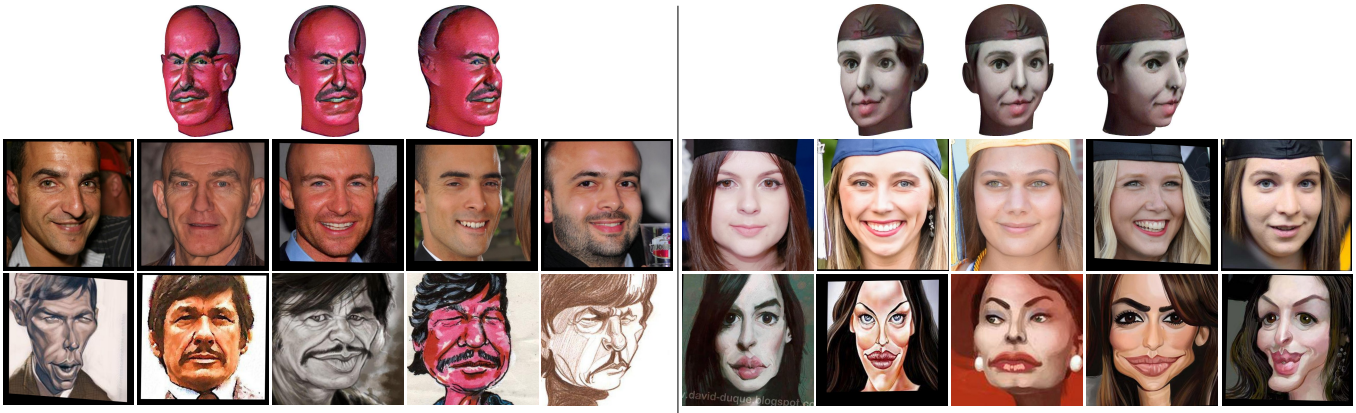


Group 9, CR: 93.10% , SR: 93.10%

Group 10, CR: 75.86% , SR: 58.62%

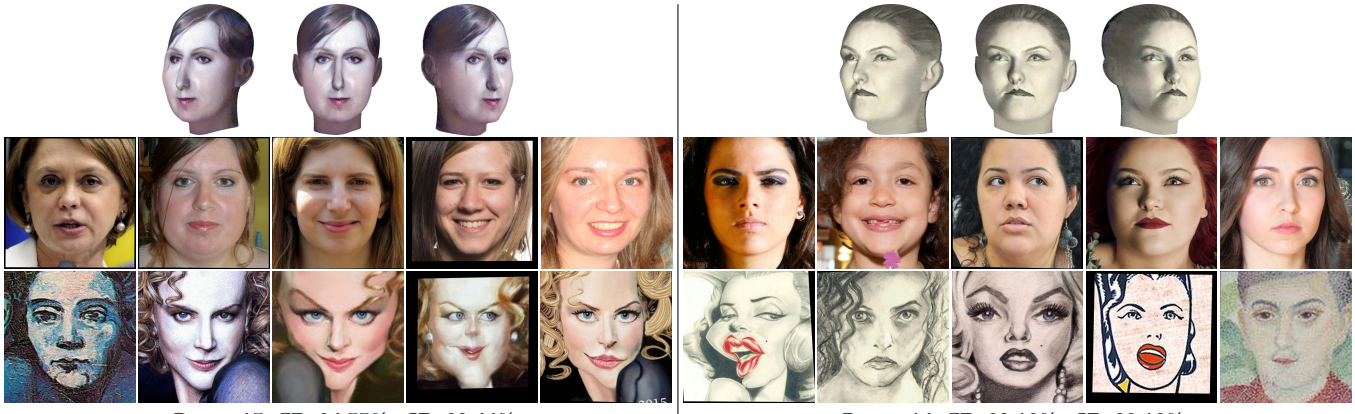
Group 11, CR: 79.31% , SR: 93.10%

Group 12, CR: 100.00% , SR: 72.41%



Group 13, CR: 72.41% , SR: 96.55%

Group 14, CR: 79.31% , SR: 75.86%



Group 15, CR: 96.55% , SR: 89.66%

Group 16, CR: 93.10% , SR: 93.10%

Fig. 6: **Perceptual study questionnaire and results.** From top to bottom of each group: three views of the stylization result, five similar content images and five similar style images.