Ranking CGANs: Subjective Control over Semantic Image Attributes- Supplementary Material

Yassir Saquil
Y.Saquil@bath.ac.uk
Kwang In Kim
K.Kim@bath.ac.uk
Peter Hall
P.M.Hall@bath.ac.uk

Visual Computing Group
University of Bath
Bath, UK

In the main paper, we reported results only on few attributes due to space limitations. We provide more results in each application which we described in the main paper to highlight the characteristics of RankCGAN. Additionally, we trained an unconditioned Stage-II GAN from StackGAN method [5] for 128×128 image generation, where the resolution is chosen due to the limited real images size.

1 Image Generation

In this section, we show more results on shoe [4], face [1], scene [2] datasets using additional attributes.

1.1 One attribute interpolation

To perform the interpolation using RankCGAN, we generate the images by varying the latent variable r in the interval $[-1,1]$, while the unconditional noise vector z is fixed.

Shoes dataset [4] comprises 4 relative attributes (sporty, pointy, open, comfortable). We defined an additional "black" attribute by comparing the color histogram of images. Figures 1, 2, 3, 4 show one attribute interpolation with respect to the attributes "black", "pointy", "open" and "sporty" respectively.

Face dataset [1] comprises 29 semantic attributes. We demonstrated one example of the "masculine" attribute in the paper. Here, we extend those results by showing two additional attributes results. Figures 5, 6, 7 demonstrate one attribute interpolation using "smiling", "young" and "masculine" attributes.

Similarly to face dataset, we highlight two additional attributes for the scene dataset. Figure 8, 9, 10 demonstrate the one attribute interpolation using "open", "perspective" and "natural" attributes. Generally, forest and tall building scenes are considered less open than coast and highway scenes, while the coast and open-country scenes are considered less perspective than street and tall building scenes.

Finally, Figure 11 show one attribute interpolation on 128×128 images using "sporty" and "masculine" attributes.

1.2 Two attributes interpolation

To perform two attributes interpolation using RankCGAN, we vary the latent variable r_1, r_2 in the interval $[-1,1]$, corresponding to the first and second attributes respectively, while fixing the

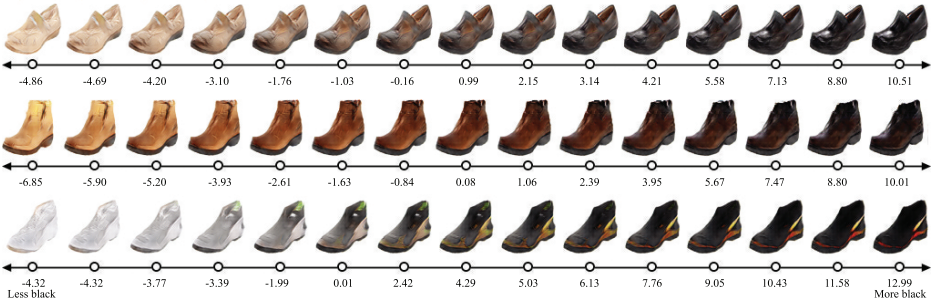

Figure 1: Examples of generated shoe images with respect to “black" attribute. The abscissa values represent the ranking score of each image.

Figure 2: Examples of generated shoe images with respect to “pointy" attribute. The abscissa values represent the ranking score of each image.

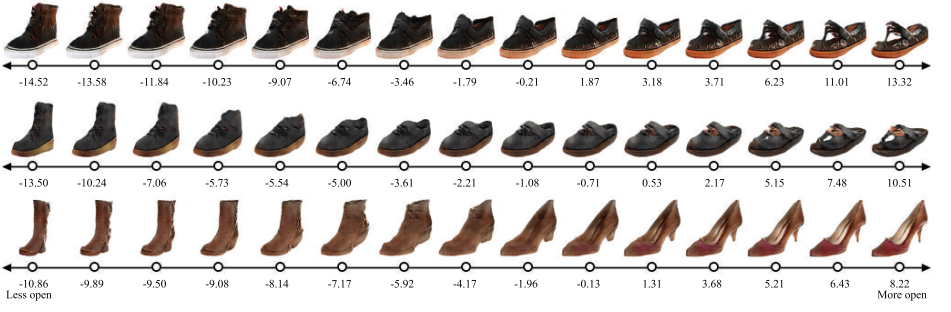

Figure 3: Examples of generated shoe images with respect to “open" attribute. The abscissa values represent the ranking score of each image.

unconditional noise vector z. The additional results support the evidence that RankCGAN decorrelates multiple semantic attributes.

On shoe dataset, Figure 12 shows the two attributes interpolation using respectively the couple of attributes (“black", “sporty”) and (“black”, “open”). Also, on face dataset, Figure 13 shows the two attributes interpolation using the couple of attributes (“masculine”, “smiling”) and (“masculine”, “young”). Moreover, Figure 14 illustrates the same interpolation on scenes dataset using the couple attributes (“open”, “natural”). Lastly, Figures 15 and 16 demonstrate similar interpolation using the couple of attributes (“black”, “sporty”) and (“masculine”, “smiling”) respectively on 128×128 images.

Figure 4: Examples of generated shoe images with respect to “sporty" attribute. The abscissa values represent the ranking score of each image.

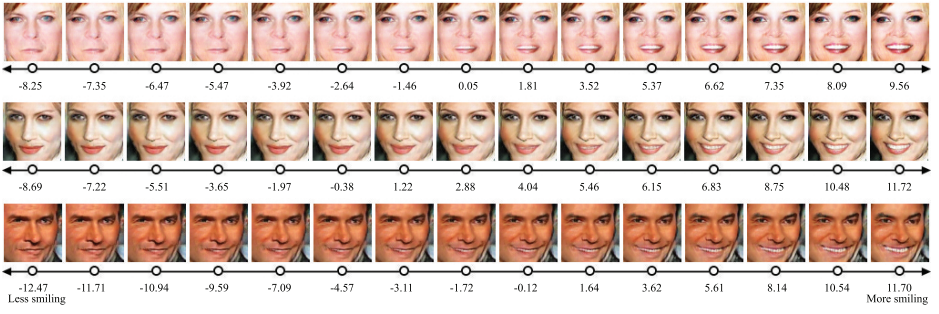

Figure 5: Examples of generated face images with respect to “smiling" attribute. The abscissa values represent the ranking score of each image.

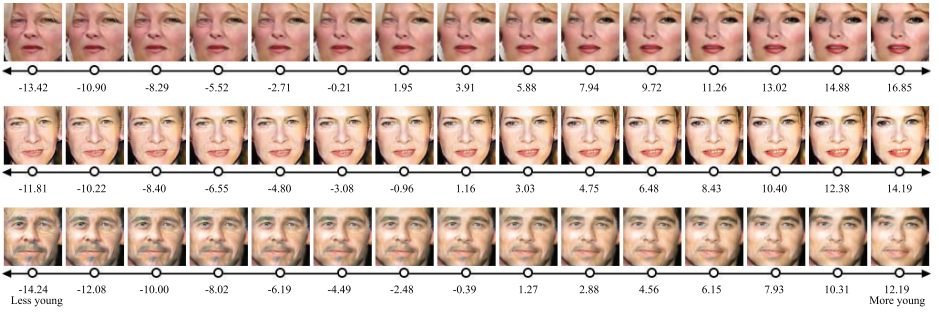

Figure 6: Examples of generated face images with respect to “young" attribute. The abscissa values represent the ranking score of each image.

2 Image Editing

In this section, we demonstrate more results on image editing tasks on all three datasets with different attributes. In order to achieve image editing, we load a real image from the dataset and estimates the latent variables r,z of RankCGAN that generate a similar image to the real one, called the reconstructed image. Then, we interpolate the reconstructed image by varying the latent variable r in the interval $[-1,1]$ and fixing the noise vector z. The estimation of the latent variables r, z is performed by encoding the real image using the encoders E_r, E_z as detailed in [3] and using the

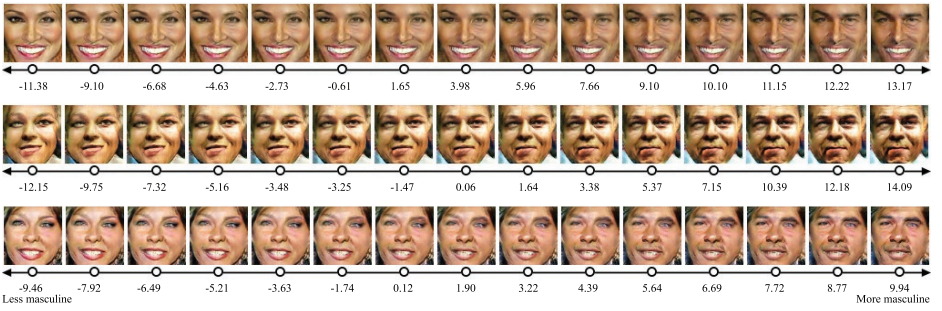

Figure 7: Examples of generated face images with respect to “masculine" attribute. The abscissa values represent the ranking score of each image.

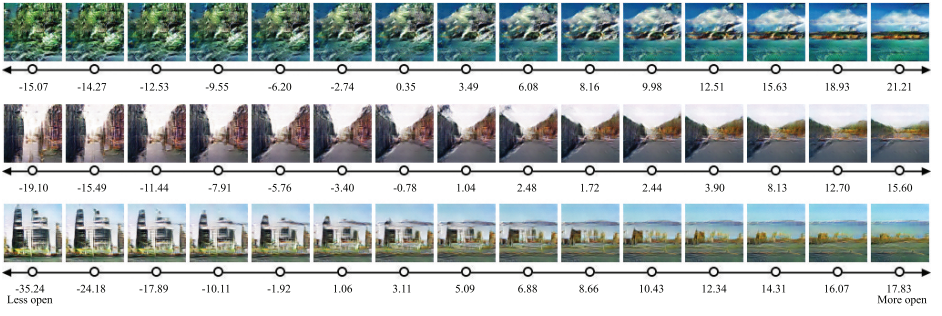

Figure 8: Examples of generated scene images with respect to “open" attribute. The abscissa values represent the ranking score of each image.

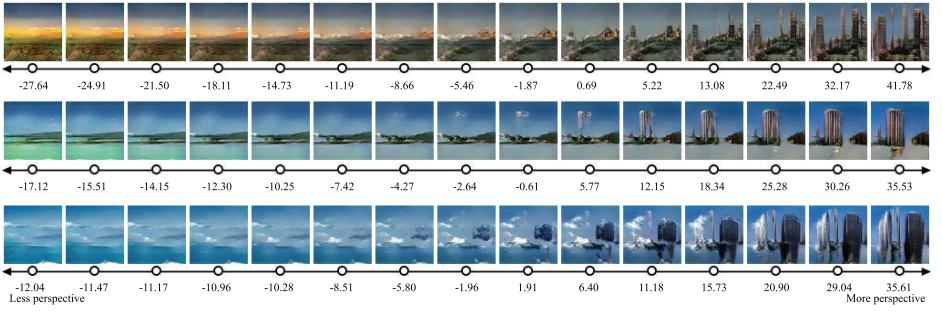

Figure 9: Examples of generated scene images with respect to “perspective" attribute. The abscissa values represent the ranking score of each image.

manifold projection method [6] as a refining step.

The Figures 17, 18, 19 illustrate results on image editing tasks using shoe, face, scenes datasets respectively, where the reconstructed images are framed in red color, and the images generated with subjectively less of the chosen attribute than the reconstructed image are shown on the top line, while the images generated with subjectively more are on the bottom line. Additionally, the figure 20 shows 128×128 image editing on shoes and faces, using the “sporty” and “masculine” attributes respectively.

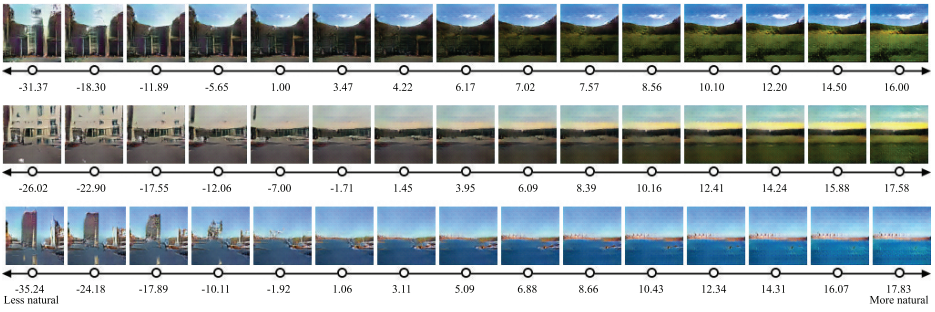

Figure 10: Examples of generated scene images with respect to “natural" attribute. The abscissa values represent the ranking score of each image.

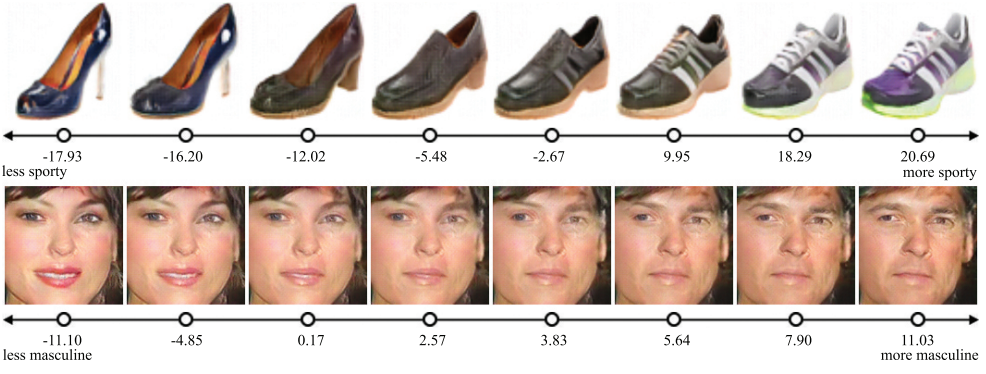

Figure 11: Examples of 128×128 generated face and shoe images with respect to “masculine” and “sporty” attributes.

3 Semantic Attribute Transfer

In this section, we provide extended results of attribute transfer tasks. Figure 21 shows the results of transferring the semantic strength of “smiling” attribute from twelve references images to ten target images. The column shows the edited target images after transferring the semantic attribute value from the reference image. It indicates that the “smiling” attribute can be transferred in images with different pose and gender. The row shows the edited target image after transferring different semantic attribute values from the reference images. It indicates that the identity of the edited image is retained regardless of the semantic strength values. In addition, figure 22 demonstrates the same attribute transfer task on 128×128 generated images.

4 Generation Quality

Our model RankCGAN is built on top of GAN architecture and therefore, it inherits many characteristics of GAN in term of diversity, generation capability and quality. Many failure examples are due to the inability of GAN to generate good quality images on the studied dataset. We notice that the generated images on shoe dataset [4] have globally a good quality because the images have the same white background and orientation. Also, on face dataset [1], we observe the appearance of some artifacts on

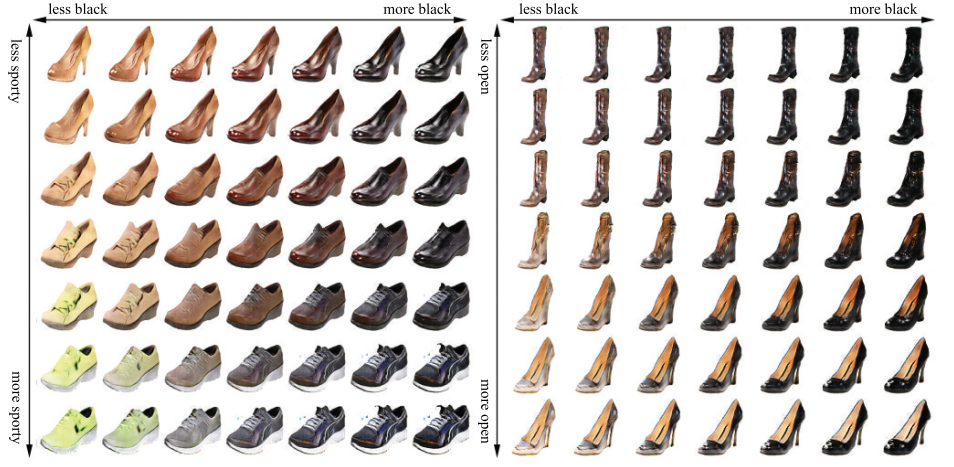

Figure 12: Example of two-attributes interpolation on shoe images using (“sporty",“black") and (“open",“black") couple attributes.

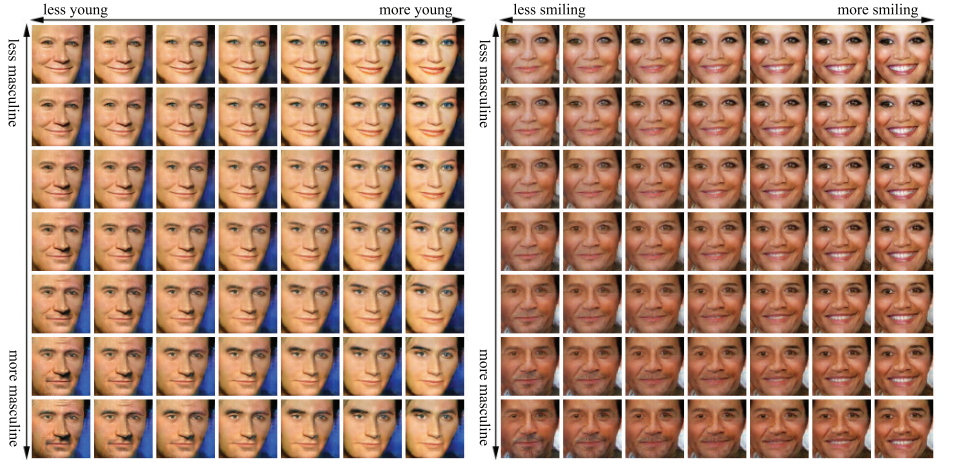

Figure 13: Example of two-attributes interpolation on face images using (“masculine",“young") and (“masculine",“smiling") couple attributes.

generated faces e.g. in Figure 21. Besides, on scene dataset [2], GAN struggles to generate some categories in the scene dataset due to its high variability such as forest in Figure 8, inside-city in Figure 9.

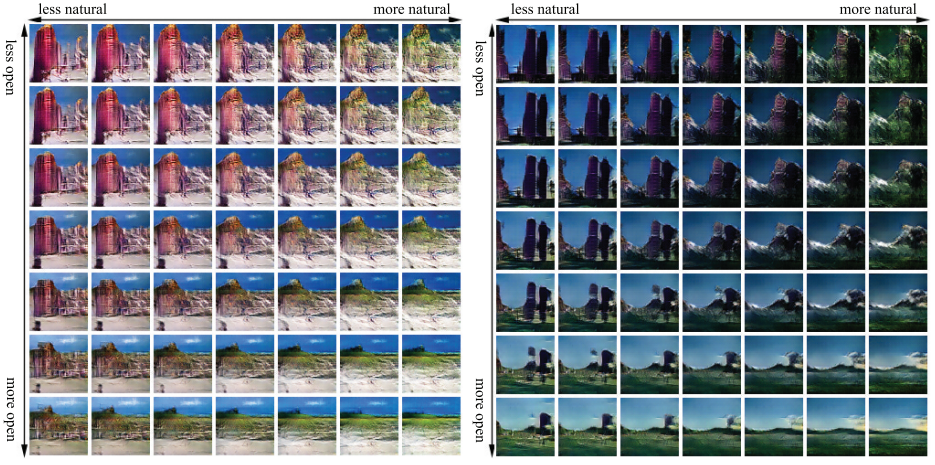

Figure 14: Examples of two-attributes interpolation on scene images using (“open",“natural") couple attributes.

References

[1] Neeraj Kumar, Alexander C. Berg, Peter N. Belhumeur, and Shree K. Nayar. Attribute and simile classifiers for face verification. In *ICCV*, 2009.

[2] Devi Parikh and Kristen Grauman. Relative attributes. In *ICCV*, 2011.

[3] Guim Perarnau, Joost van de Weijer, Bogdan Raducanu, and Jose M. Álvarez. Invertible conditional gans for image editing. In *NIPS Workshop on Adversarial Training*, 2016.

[4] A. Yu and K. Grauman. Fine-Grained Visual Comparisons with Local Learning. In *CVPR*, 2014.

[5] Han Zhang, Tao Xu, Hongsheng Li, Shaoting Zhang, Xiaogang Wang, Xiaolei Huang, and Dimitris Metaxas. Stackgan: Text to photo-realistic image synthesis with stacked generative adversarial networks. In *ICCV*, 2017.

[6] Jun-Yan Zhu, Philipp Krähenbühl, Eli Shechtman, and Alexei A. Efros. Generative visual manipulation on the natural image manifold. In *ECCV*, 2016.

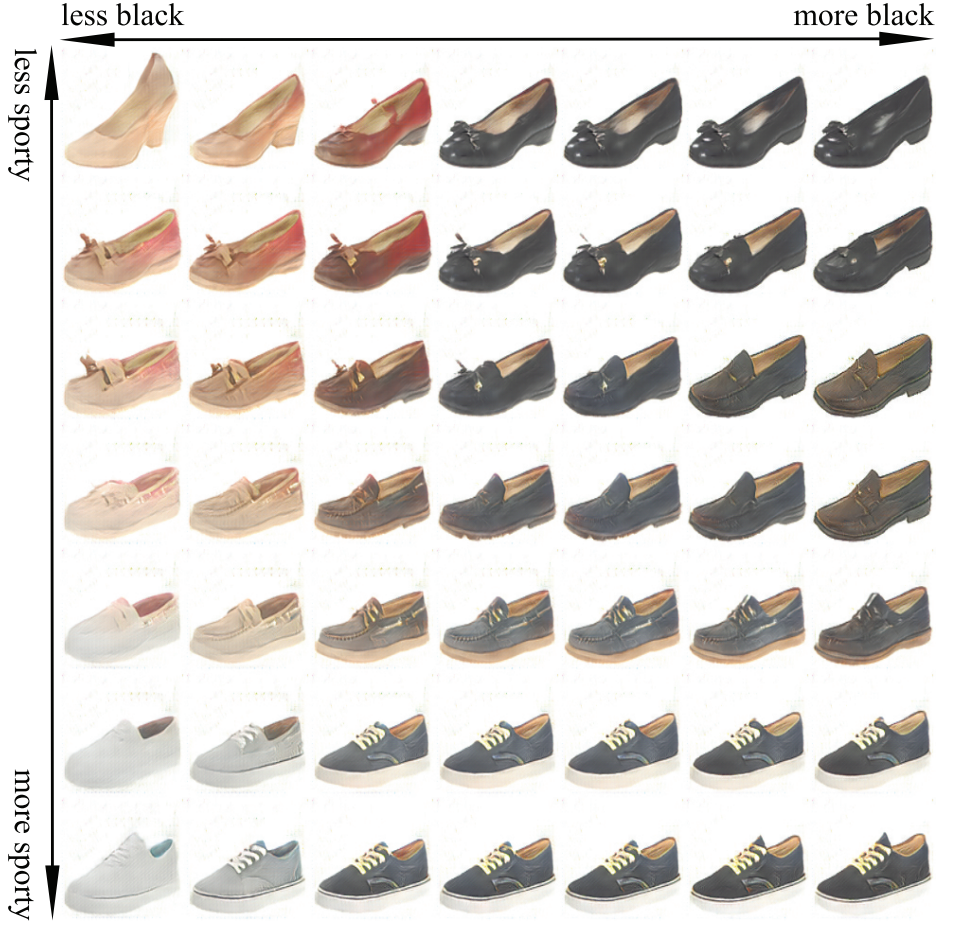

Figure 15: Example of two-attributes interpolation on 128×128 shoe images using (“sporty",“black") couple attributes.

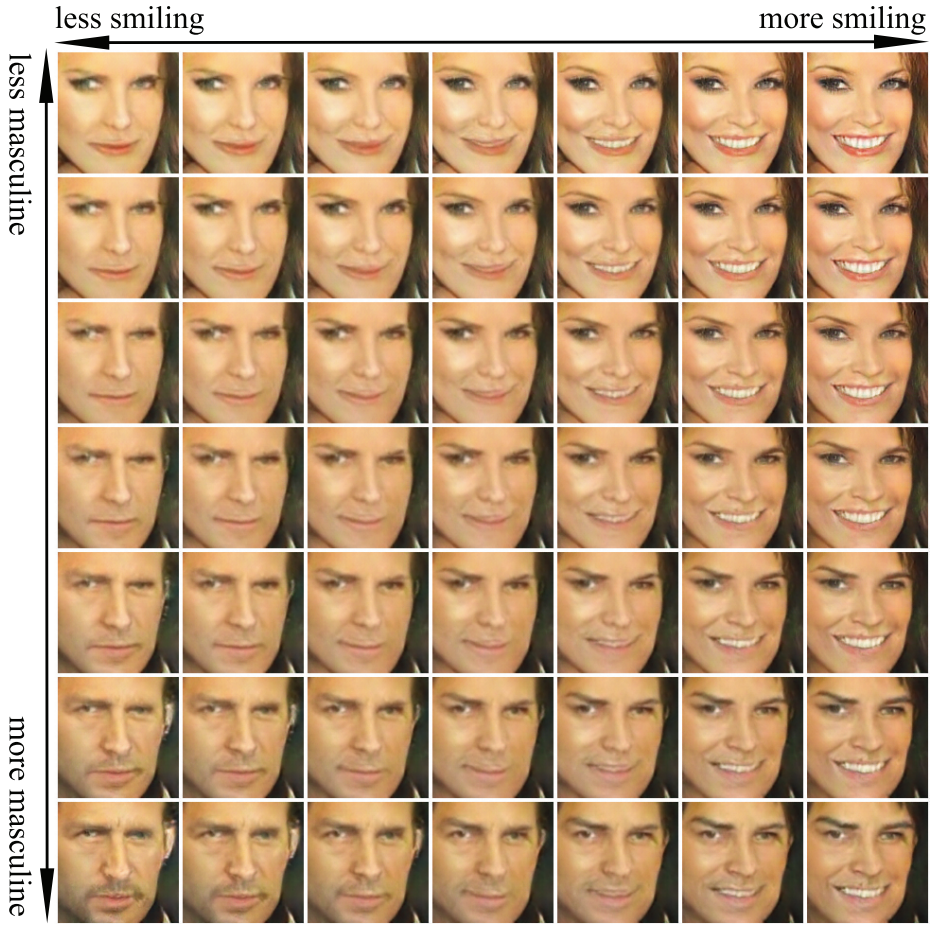

Figure 16: Example of two-attributes interpolation on 128 × 128 face images using (“masculine",“smiling") couple attributes.

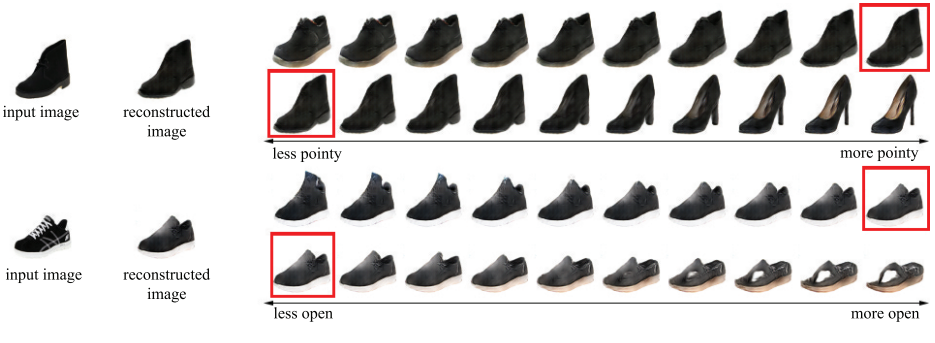

Figure 17: Examples of image editing task, where the latent variables are estimated for an input image to perform the editing with respect to the “pointy" attribute (top) and “open" attribute (bottom).

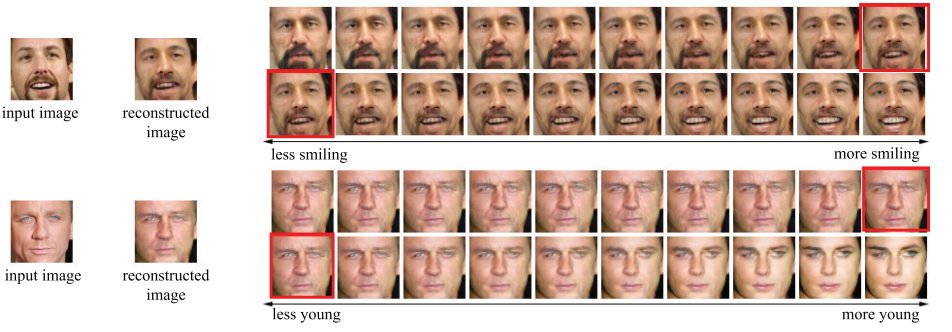

Figure 18: Examples of image editing task, where the latent variables are estimated for an input image to perform the editing with respect to the “smiling" attribute (top) and “young" attribute (bottom).

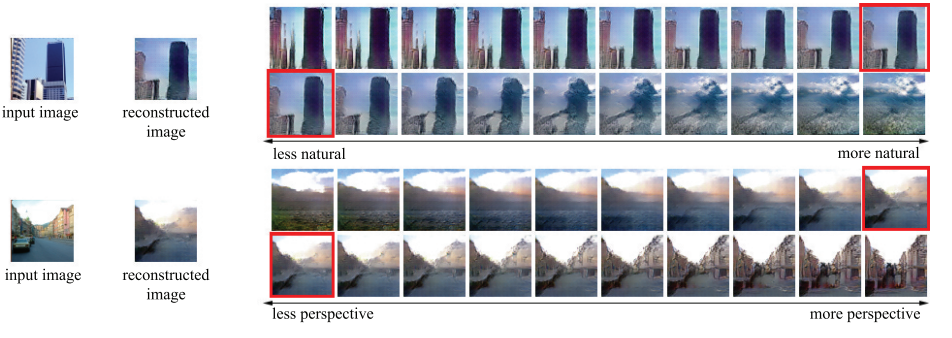

Figure 19: Examples of image editing task, where the latent variables are estimated for an input image to perform the editing with respect to the “natural" attribute (top) and “perspective" attribute (bottom).

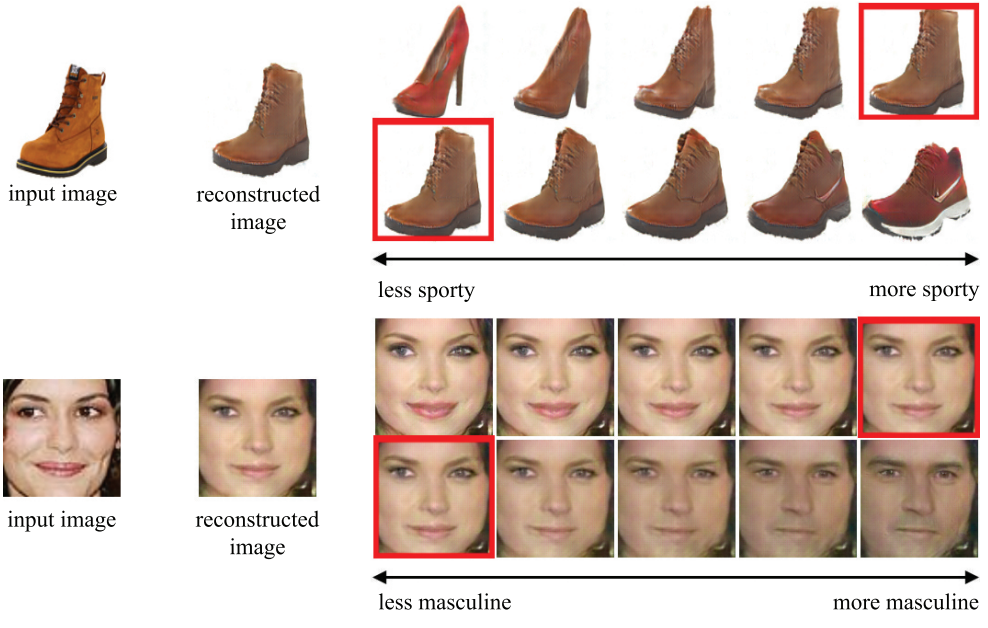

Figure 20: Examples of 128×128 image editing task, where the latent variables are estimated for an input image to perform the editing with respect to the "sporty" attribute (top) and "masculine" attribute (bottom).

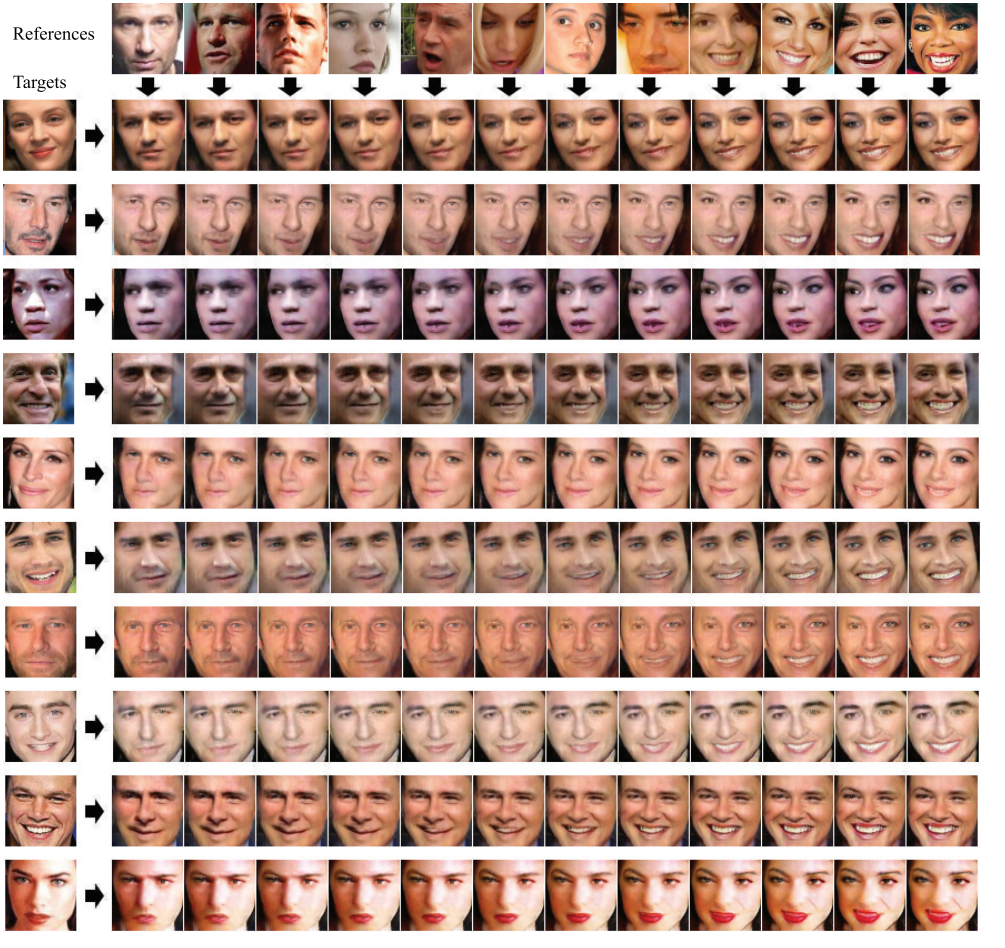

Figure 21: Examples of “smiling" attribute transfer task. The latent variable r of reference images is estimated and then transferred to the target image in order to express the same strength of attribute.

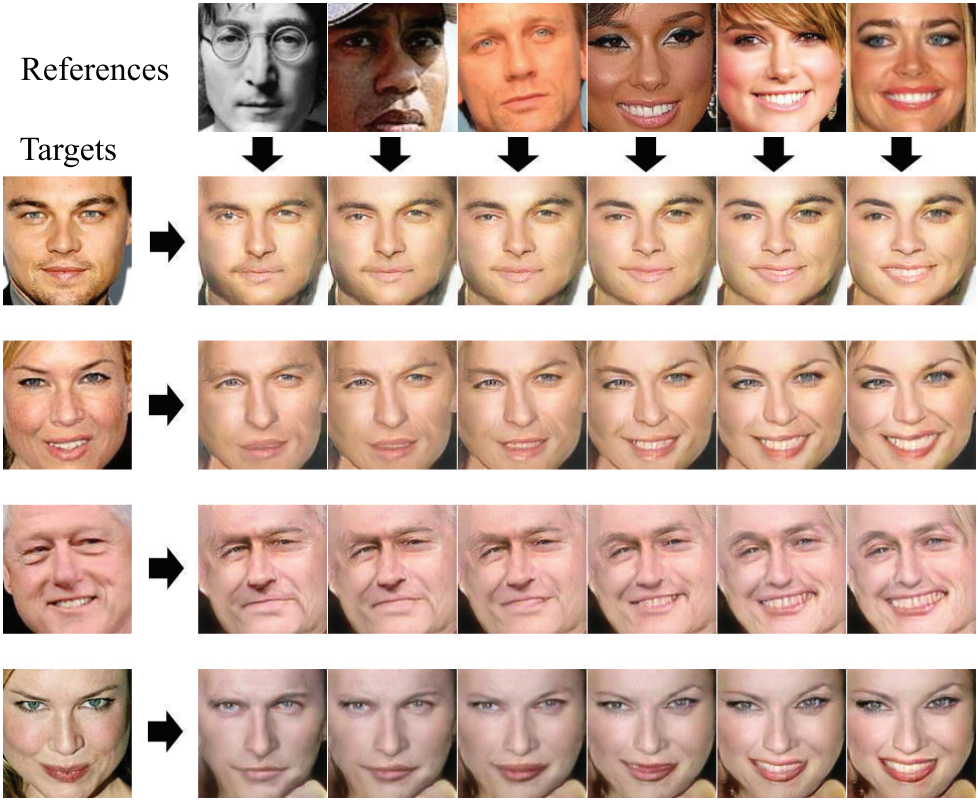

Figure 22: Examples of “smiling" attribute transfer task on 128×128 images. The latent variable r of reference images is estimated and then transferred to the target image in order to express the same strength of attribute.