1. Supplementary

1.1. Additional Algorithm Details

Algorithm 1 Calculate α in Equ.3

Require: Hair classifier C_{hair}, generator G, latent code w^+, and hair separation boundary normal vector n_h
  $s \leftarrow 1$
  $n \leftarrow 0$
  while $s \neq 1$ **do**
    $n \leftarrow n + 1$
    $s = C_{hair}(G(w^+ - n \times n_h))$
  end while

For Alg. 2, we first train a gender classifier C_{gender} as well as training C_{hair} in Subsection 3.2. We use the attribute *Male* in CelebAHQ-mask to train C_{gender}. Male portraits will be scored by C_{gender} as $s_{gender} = 1$, female portraits will be scored as $s_{gender} = 0$, the gender classifier is also used to distinguish male latent codes and female latent codes in the training stage.

Algorithm 2 Calculate α in Equ.13

Require: Gender classifier C_{gender}, generator G, female latent code w^+, and gender separation boundary normal vector n_g
  $s \leftarrow 0$
  $n \leftarrow 0$
  while $s \neq 0$ **do**
    $n \leftarrow n + 1$
    $s = C_{gender}(G(w^+ + n \times n_g))$
  end while

1.2. Additional Details about Run Time

In the training stage, it takes about 7 hours to train M_m and 3 hours to train M. Performing the proposed two pipelines to get H_m and H takes about 2 weeks. In the testing stage, it takes 0.0959 seconds for the $e4e$ encoder to encode a real image, and 0.0036 seconds for $HairMapper$ to process an input latent code. Feeding latent code to StyleGAN generator to get the portrait without hair takes 0.1534 seconds. Generating a hair mask takes 0.0952 seconds. Poisson editing takes 0.3495 seconds. Other steps take negligible time. In sum, our method takes 0.7181 seconds to process a real image.

1.3. Additional Details about Real Image Processing

Since the e4e encoder cannot precisely encode real images, there will be artifacts in the resulting image. Therefore, we apply an additional diffusion step to eliminate artifacts. Similar to the diffusion step mentioned in Subsection 3.4, we use a synthetic image as the prior and use the edited latent code to initialize optimized latent code. After diffusion, the face region in the image generated from the optimized latent code will be similar to the one in the input real image. We apply the additional diffusion step in our non-hair-FFHQ to guarantee the quality of our dataset. If we use an additional diffusion step, our method takes 0.7181 seconds to process a real image, and it takes 41.5736 seconds to perform the diffusion.

1.4. Additional Ablation Study

Figure 1. From left to right: the original image, image edited by n_h^{fine}, and image edited by M.

Mapper vs. Separation Boundary. Given an input image (Fig. 1, left), Fig. 1 shows a typical comparison between the result by using "fine separation boundary" based on SVMs (Fig. 1, middle) and the result by training a mapper M (Fig. 1, right). We can find that a naive method not only cannot fully remove hair but also changes facial identity significantly, indicating that hair removal is not a simple linear problem.

1.5. Additional Comparisons with 3D Head Reconstruction Work

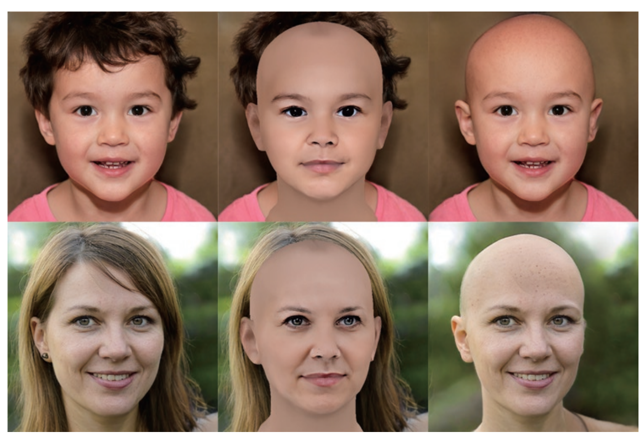

Figure 2. Comparisons with the 3D head reconstruction work.

Our method better preserves facial identity without 3D reconstruction which is likely to cause feature loss, while 3D head reconstruction would benefit relighting, novel view synthesis.

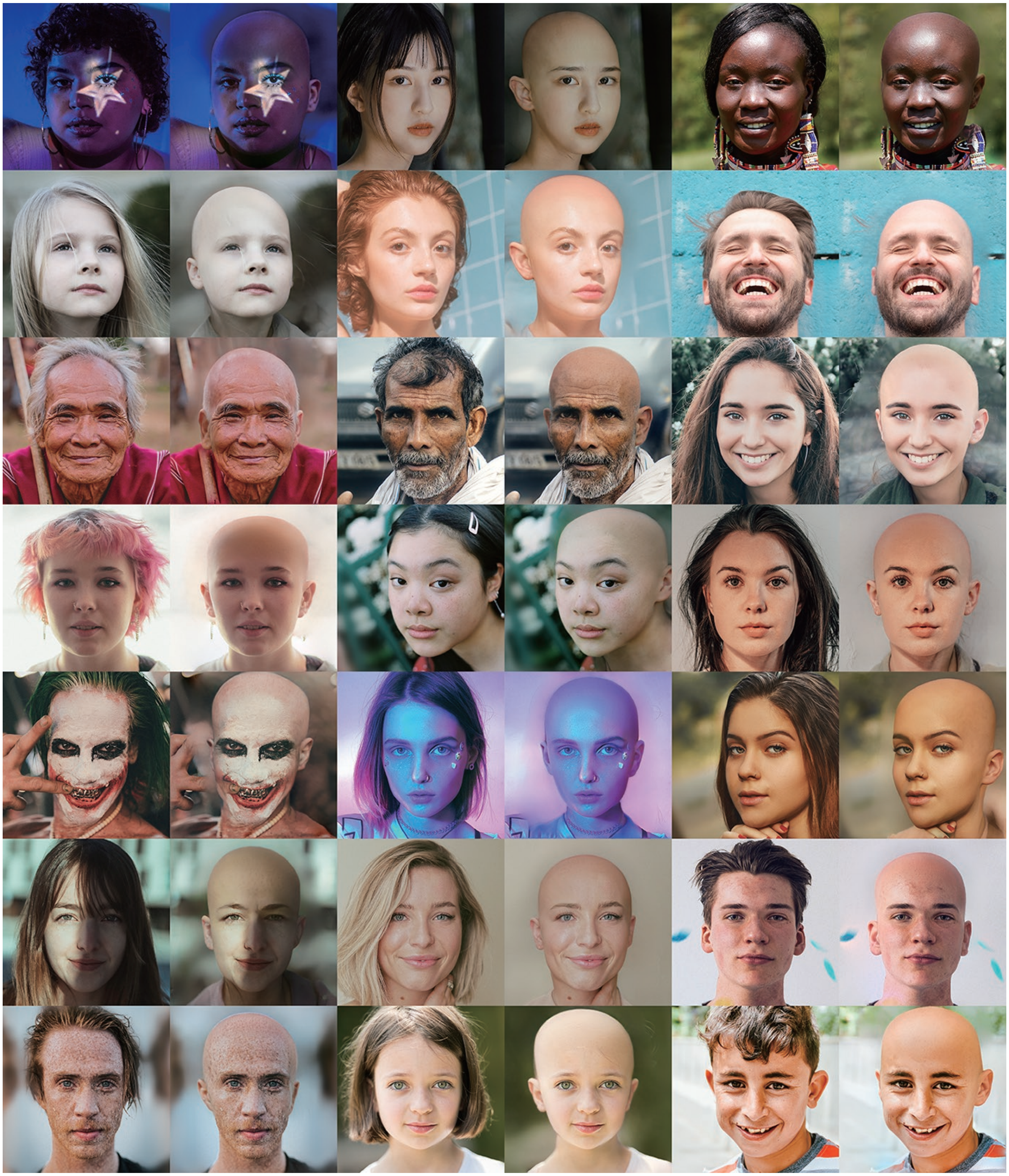

Figure 3. Qualitative results of $HairMapper$. Experimental results show that our method can deal with portrait images with different gender, age, race, hairstyle, expression, pose, hair occlusion, shadow, and lighting successfully.

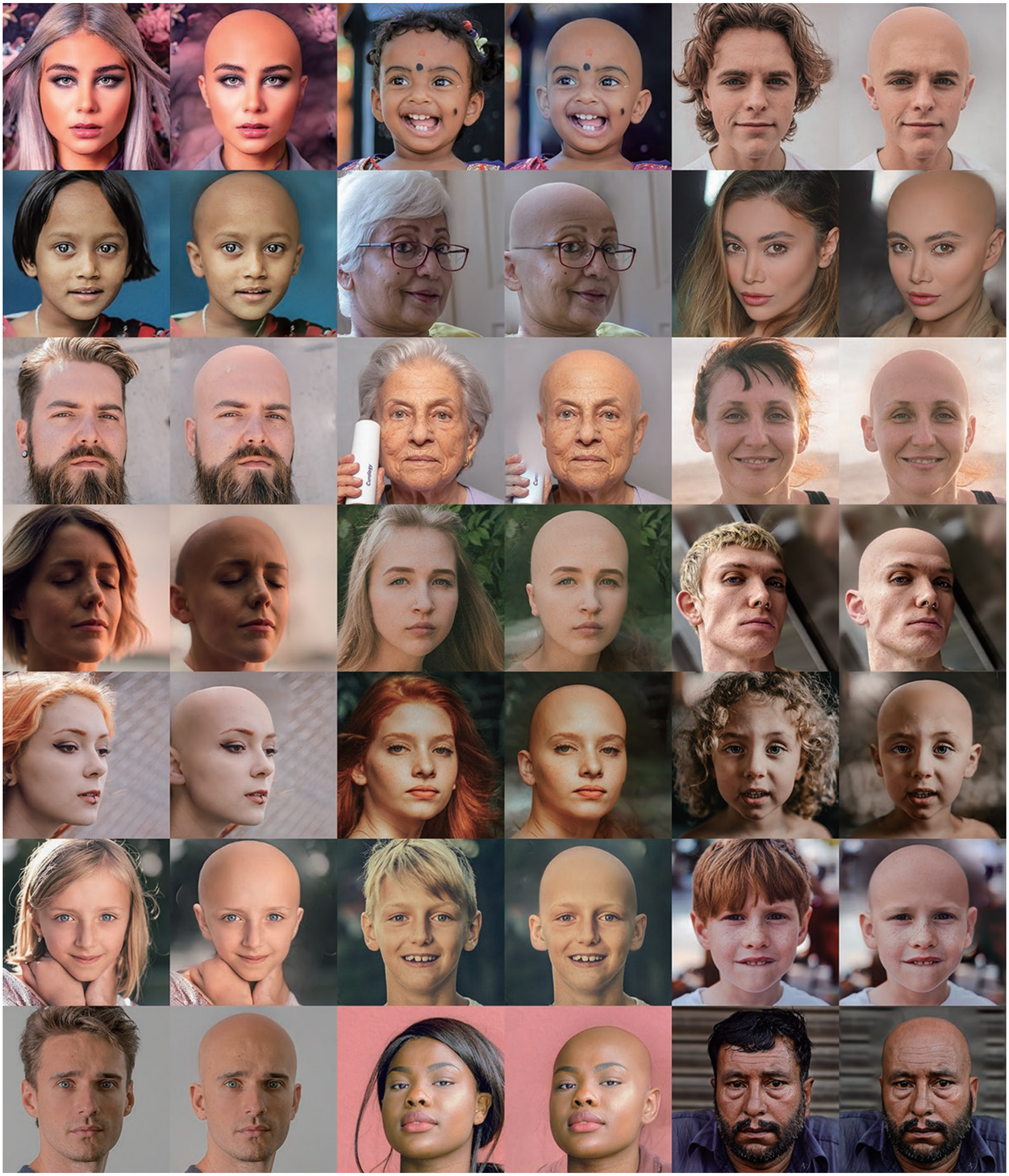

Figure 4. Qualitative results of $HairMapper$ on diverse portrait images.

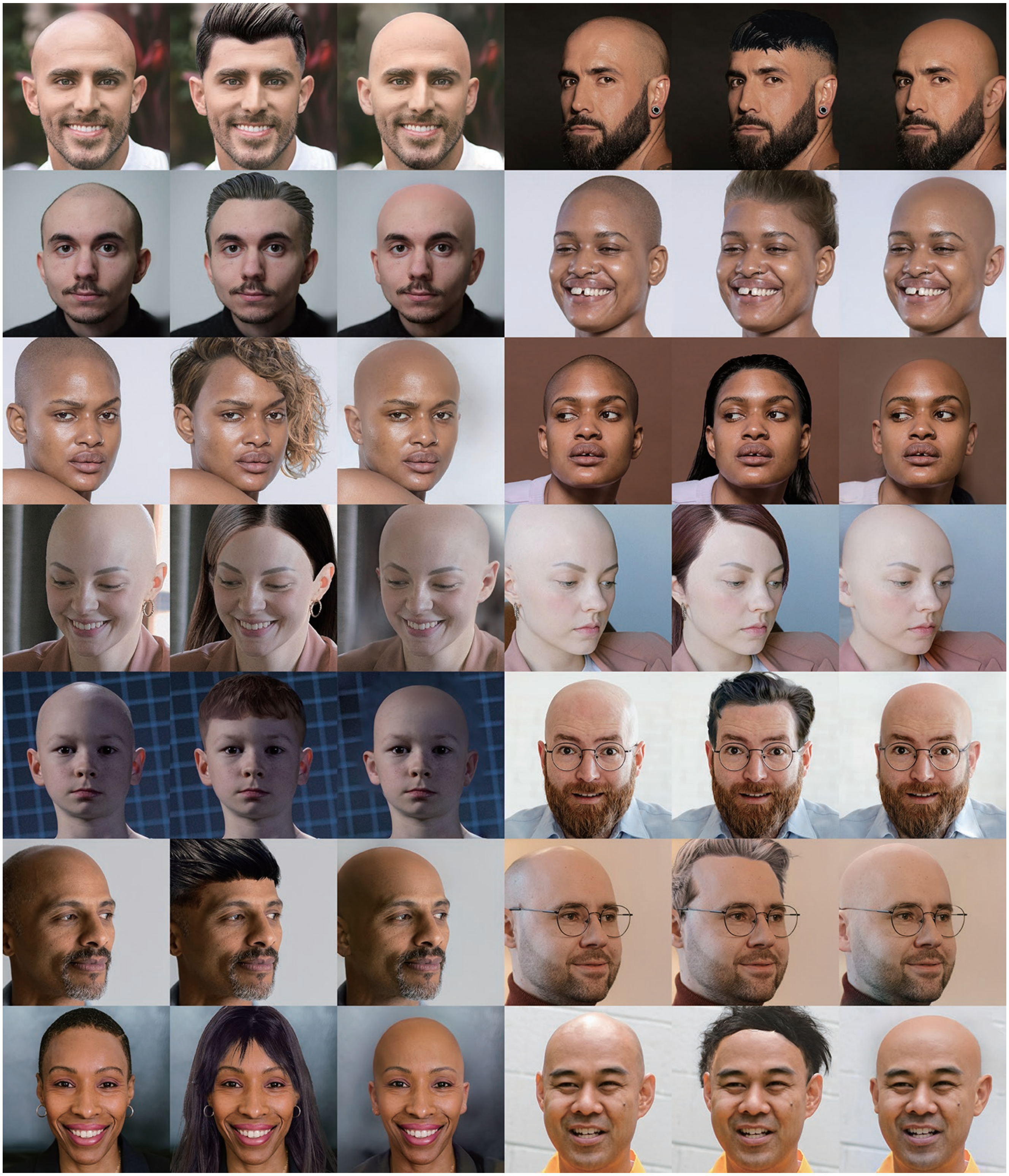

Figure 5. "pseudo ground-truth" samples. Given the real bald portraits (columns 1 and 4), we manually add hair templates to them (columns 2 and 5), then apply our method to the hair-added portraits (columns 3 and 6).

Figure 6. GANFIT reconstruction results. From left to right: input portrait image, face texture, and reconstructed face with textures.

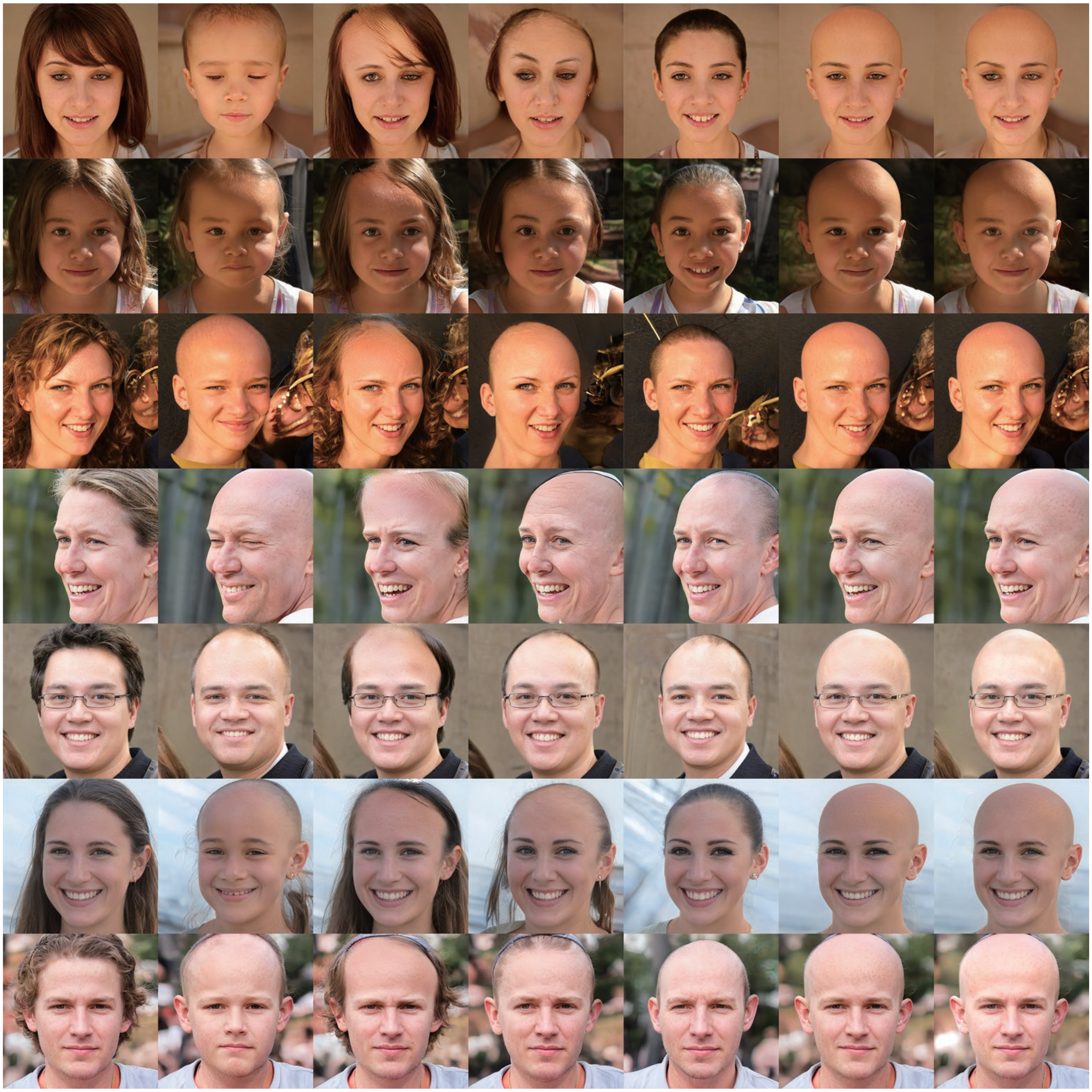

Figure 7. More qualitative comparison results for state-of-the-art methods. From left to right: the original image, InterfaceGAN, StyleSpace,StyleClip, StyleFlow, ours w/o blending, and ours w/ blending

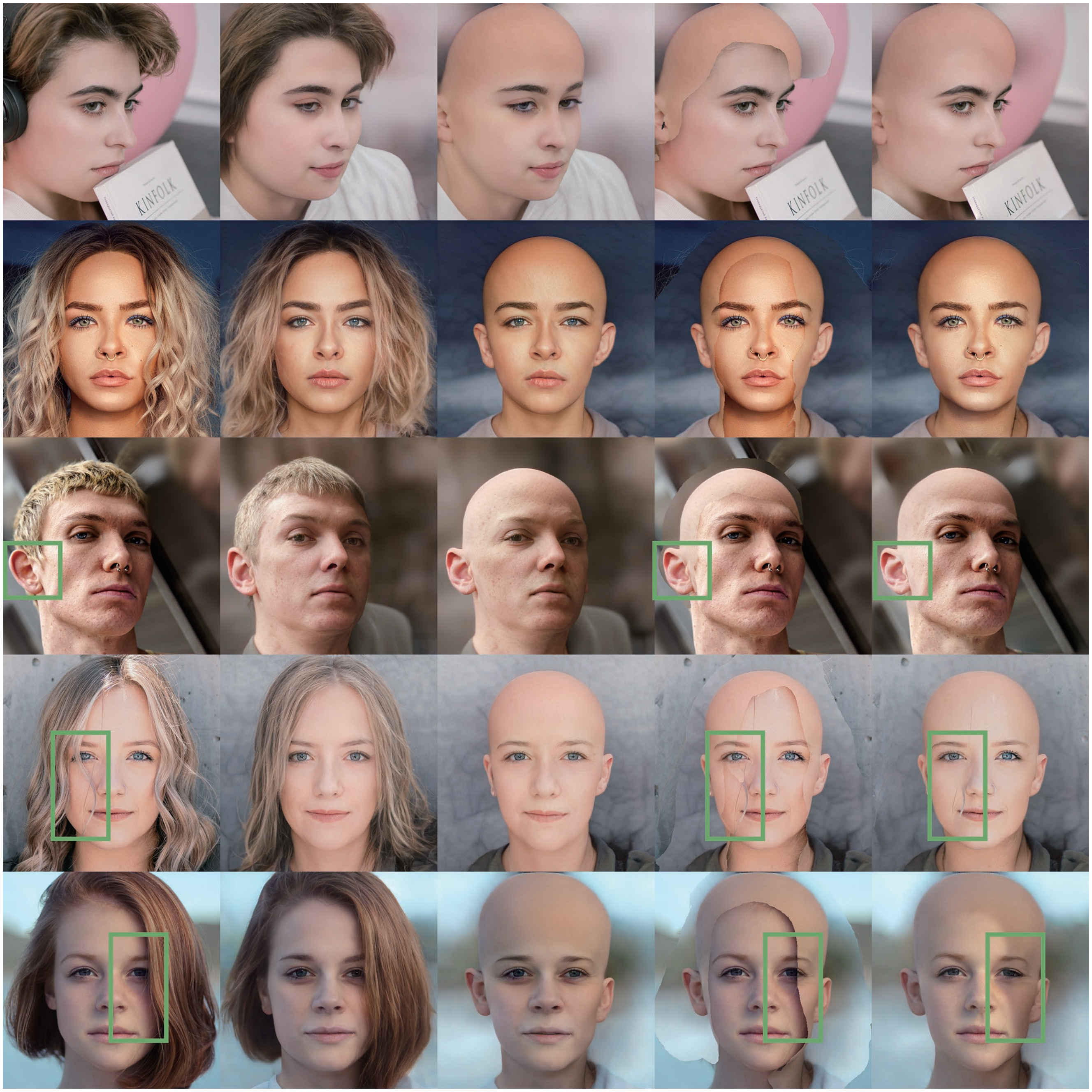

Figure 8. From left to right: the original real image, encoded image, our $HairMapper$ result, image (synthesized by hair mask, real image, and $HairMapper$ result), and Poisson editing result. In rows 1-3, the real input portraits cannot be precisely encoded by $e4e$, thus the resulting image in row 1 has artifacts: the resulting image in row 2 has the overall color changed, the resulting image in row 3 has the ear shape changed (highlighted by green box). In row 4, hair not covered by the hair mask can be noticeable in the result (highlighted by green box). In row 5, extreme lighting conditions bring up the “lightened” face.