Supplementary Materials
AnyBald: Toward Realistic Diffusion-Based Hair Removal In-The-Wild

Yongjun Choi[1*] Seungoh Han[1*] Soomin Kim[2] Sumin Son[2]
Mohsen Rohani[3] Edgar Maucourant[3] Dongbo Min[2] Kyungdon Joo[1†]

[1]Ulsan National Institute of Science and Technology (UNIST), [2]Ewha Womans University, [3]ModiFace
{ccyjun123, sohan, kyungdon}@unist.ac.kr, {kim.soomin, sumin.son, dbmin}@ewha.ac.kr
{mohsen.rohani, edgar.maucourant}@loreal.com

Overview

In the supplementary materials, we provide further details on the following topics:

- Section 1 presents an extensive discussion of evaluation metrics for in-the-wild datasets, along with detailed descriptions of our Bald Head Augmentation pipeline, evaluation set construction, and implementation settings.
- Section 2 offers additional qualitative comparisons on the CelebA and DeepFashion2 datasets against baselines.
- Section 3 describes the detailed setup of the user study, along with examples of the survey composition, and includes the full set of images used in the evaluation.
- Section 4 showcases three representative applications that demonstrate the utility of bald images generated by our `AnyBald` model.
- Section 5 discusses the limitations of our work and outlines potential directions for future research.

1. Experiment Details

1.1. Analysis on Metrics of Unpaired Set

We decide to measure alternatives for quantitative comparisons of the in-the-wild images, such as CelebA [7] and DeepFashion2 [3] without GT bald images. We clarify our empirical observation that several metrics fail to properly evaluate generated images when the reference ($i.e.$, GT bald) does not exactly match the output. In Fig. 1, we present CLIP-I scores along with the corresponding images for different methods. CLIP-I computes similarity between the generated and input images in the CLIP embedding space. Notably, although PowerPaint (PPT) retains visible hair artifacts, it still achieves the highest CLIP-I score.

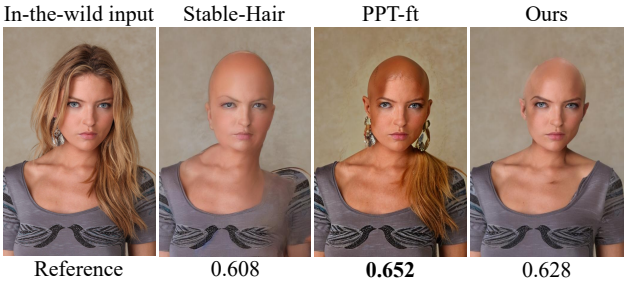

Figure 1. CLIP-I results on unpaired images. Despite artifacts such as residual hair and color inconsistency around the mask boundary, PPT achieves a higher score than the clean bald image produced by our method.

1.2. Details of Bald Head Augmentation

Preprocess Stage. To remove remaining artifacts from non-hair-FFHQ [8], we adopt the face parser [9]. Given a bald image I_b, we extract a face segmentation map with 19 semantic classes. By filtering out non-face classes such as *background*, *necklace*, *neck*, and *clothing* from the segmentation map, we obtain a cleaned bald reference $\tilde{I}_b$ that contains only face-related pixels.

Effect of Synthesizing Proxy I_{proxy}. We introduce an additional step to generate a proxy image I_{proxy} which conditioned on head pose, to obtain an aligned hair mask. Although this may seem indirect, the design is driven by the need for reliable and pose-aligned hair masks, which are difficult to obtain directly from the original input. As shown in Fig. 2, we qualitatively compare the resulting images between our two-stage hairstyle generation and a single-stage generation from the augmented bald image $\hat{I}_b$ with FLUX.1-Kontext [6].

Synthesized Data. We visualize examples of paired data from `AnyBald` dataset as shown in Fig. 3.

*Equal contribution.
†Corresponding author.

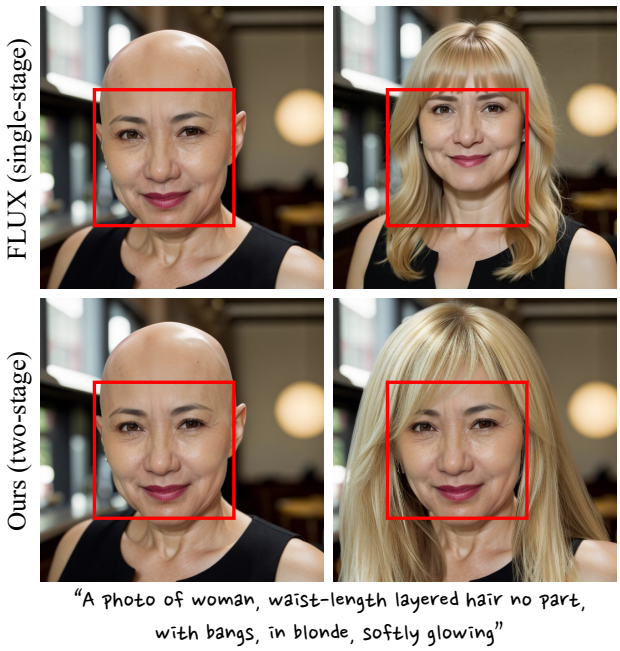

Figure 2. Effect of two-stage augmentation. Red boxes indicate the same regions within each image. We highlight that a single-stage generation suffers from misalignment.

Prompts. Table 1 and Table 2 show the sample prompts for bald augmentation and hairstyle generation, respectively. For the realistic augmentation, we categorize three important components: *Head Pose*, *Clothing*, and *Background*. And we combine each component as follows: "A photo of a woman from the chest up, {Head Pose}, {Clothing}, {Background}". For the hair inpainting, we combine the base prompt with a selected style prompt as "A photo of a woman, {style}".

1.3. Evaluation Dataset

CelebA. We use "in-the-wild" images as raw images before preprocessing of CelebA [7] (hereafter referred to as CelebA In-the-Wild). We remove unrelated attributes from CelebA: *Wearing Hat*, *Blurry*, *Bald*, and *Receding Hairline*. Then, we also skip low-quality images whose shorter side is less than 512 pixels. Finally, a face detection model[1] is used to remove images without detected faces, as well as those containing huge faces where the hairstyle is not clearly visible.

DeepFashion2. This dataset [3] is originally proposed for virtual try-on tasks, but it also contains meaningful in-the-wild images suitable for hair removal. To leverage this, we utilize the preprocessing pipeline from StreetVTON [1], which filters and selects images with clearly visible human bodies based on pre-computed human body annotations. Then, we employ the same strategy as CelebA, removing low-quality images and adopting the face detection model.

[1] https://huggingface.co/AdamCodd/YOLOv11n-face-detection

1.4. Implementation Details

Our training pipeline is built upon the PowerPaint v2 architecture [11], initialized from its official checkpoint. As PowerPaint expects a 9-channel input (noisy latent, masked image latent, and mask), we load only the first 8 channels from the pretrained weights to match the input format of our model. For our learnable prompt P_{learn}, we initialize it with the context-aware prompt P_{ctxt} from original PowerPaint, which is one of its three learnable prompts. Additionally, we reinitialize the pretrained zero convolution layers to adapt the model from a mask-based inpainting setup to our mask-free setting.

2. Additional Qualitative Results

Fig. 5 and Fig. 6 present additional qualitative results on the CelebA In-the-Wild and DeepFashion2 datasets, respectively. For comparison, we evaluate the same baseline models used in the main paper: HairMapper (HM) [8], Stable-Hair (SH) [10], ControlNet-Inpainting (CNI), PowerPaint v2 (PPT) [11], and BrushNet (Brush) [4]. As shown in the figures, HairMapper often generates images with incorrect identities or distorted body structures, particularly for in-the-wild inputs. This is likely due to its reliance on a StyleGAN-based encoder [5] trained on limited data, which struggles to accurately interpret facial regions under diverse poses. While Stable-Hair demonstrates better generalization in hair removal compared to HairMapper, it still fails to maintain body structure in challenging cases, such as when the upper body is visible (see the last row in Fig. 5). Mask-based inpainting models maintain global consistency by restricting editing to masked regions. However, they often produce noticeable boundaries due to color mismatches between the masked and unmasked areas, and may also introduce undesired background artifacts in some cases. In contrast, our method achieves more natural and consistent hair removal across diverse inputs and better preserves non-hair regions than the baselines.

3. User Study Details

To validate our survey, we attach user study format in Fig. 7. All participants are anonymized in the survey form and provided consent for the experimental study. The evaluation included two distinct survey sets, each comprising 15 questions based on 15 different original input images. Given a reference input, participants were shown 6 output images, each generated by a different model, including ours. These images are randomly ordered to prevent potential bias. Participants are asked to evaluate the images based on the following three criteria: (1) Accuracy of hair removal,

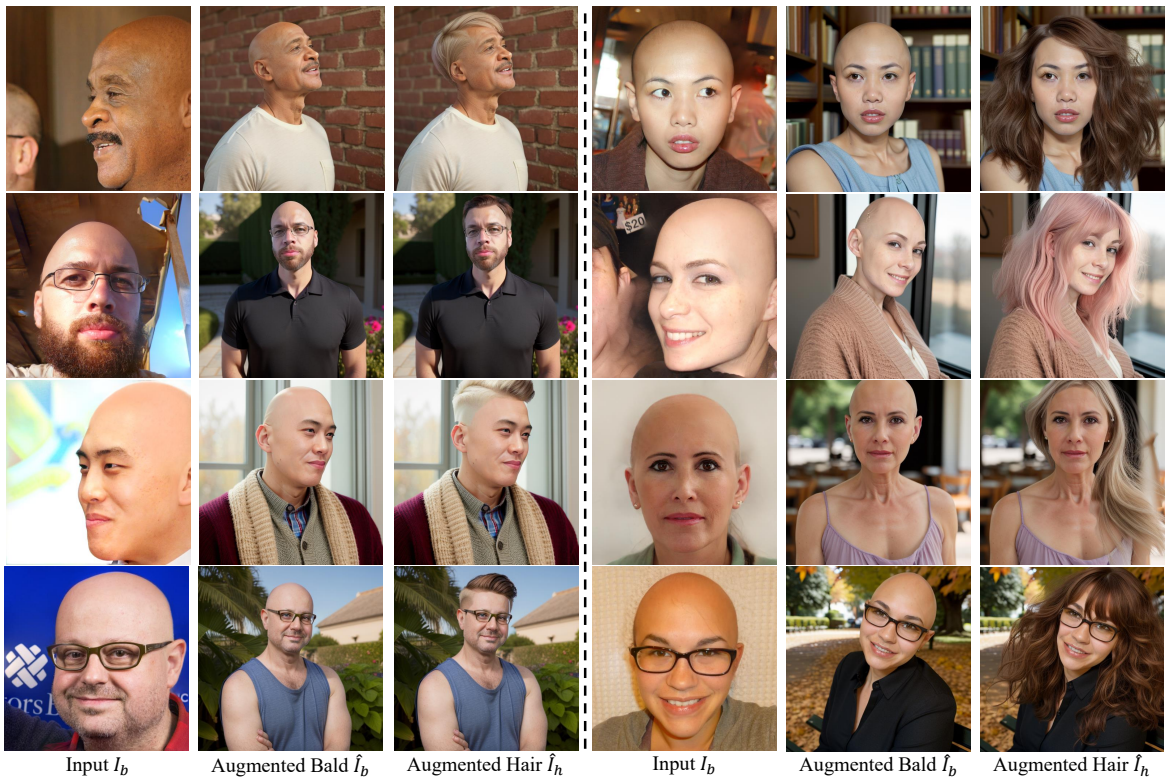

Figure 3. Sampled images from `AnyBald` Dataset.

(2) Preservation of non-hair, and (3) Visual Naturalness. For each criterion, participants are asked to select the single image they considered best. Fig. 8 and Fig. 9 show user study comparison results on CelebA, Fig. 10 and Fig. 11 on DeepFashion2, and Fig. 12 and Fig. 13 on additional in-the-wild images collected from the internet.

4. Additional Application Results

This section demonstrates the applicability of bald images from our `AnyBald` model across three downstream tasks. First, we present the results of multi-subject hair removal in Fig 14. Despite being trained solely on single-subject portraits, our model generalizes well to images containing multiple people. It reasonably identifies and removes hair regions for each individual while preserving non-hair areas such as facial structure and clothing. Also, despite being trained only on realistic human images, our model is still able to remove hair to some extent from cross-domain inputs, such as stylized or animation-like characters, as shown in column 2 of Fig. 14.

Next, we compare the results of 3D face reconstruction using the original input images and the bald images produced by our model. To do this, we employ DECA [2], a widely adopted model for 3D face reconstruction. As shown in Fig 15 (a), it often occludes parts of the face when hair remains on the forehead or sides, which leads to the reconstructed texture containing hair artifacts and degrading visual quality. In contrast, the results using our bald images in Fig 15 (b) produce clean facial textures without hair artifacts, better preserving the original appearance.

Lastly, we present improved hair transfer results by leveraging bald images produced by our `AnyBald`, in comparison to those from the original Stable-Hair. To this end, we utilize the hair transfer module from Stable-Hair, providing it with bald images produced by our model and the bald converter of Stable-Hair, respectively. As shown in Fig. 16, Stable-Hair occasionally produces collapsed body structures visible in the red box, leading to poor hair alignment and unnatural appearance. In contrast, our method generates structurally robust bald images while preserving identity and facial details. This enhances the overall quality of hair transfer, resulting in more natural and realistic results.

5. Limitations And Future Works

Our method sometimes faces limitations in cases of severe external occlusion. As shown in the fourth row in Fig. 11 and the third row in Fig. 12, when the face or upper-body clothing is heavily occluded by hair or hands, our model sometimes hallucinates new garments or introduces arti-

(a) (b)

Figure 4. Analysis on wearing hats. (a) In some cases, the hat is preserved well while naturally removing the underlying hair. (b) In other cases, the hat is partially lost or entirely removed.

facts, instead of accurately reconstructing the original appearance based on the visual cues. In some other cases, like in the first and fourth rows of Fig. 8, the model occasionally generates unintended black artifacts when the background is overly bright or entirely white, which shows difficulty in handling uniformly colored regions during background completion. Furthermore, our model struggles to properly handle accessories worn on the head, such as hats or hairbands. As shown in Fig. 4(a), when a subject wears a hat, the model handles only the visible hair beneath it, whereas parts of the hat are removed or entirely disappear in (b). This uncertainty likely arises from the absence of training data containing such accessories. Overall, these cases highlight the difficulty of recovering severely occluded or ambiguous regions with perceptual consistency, which remains an important direction for future research.

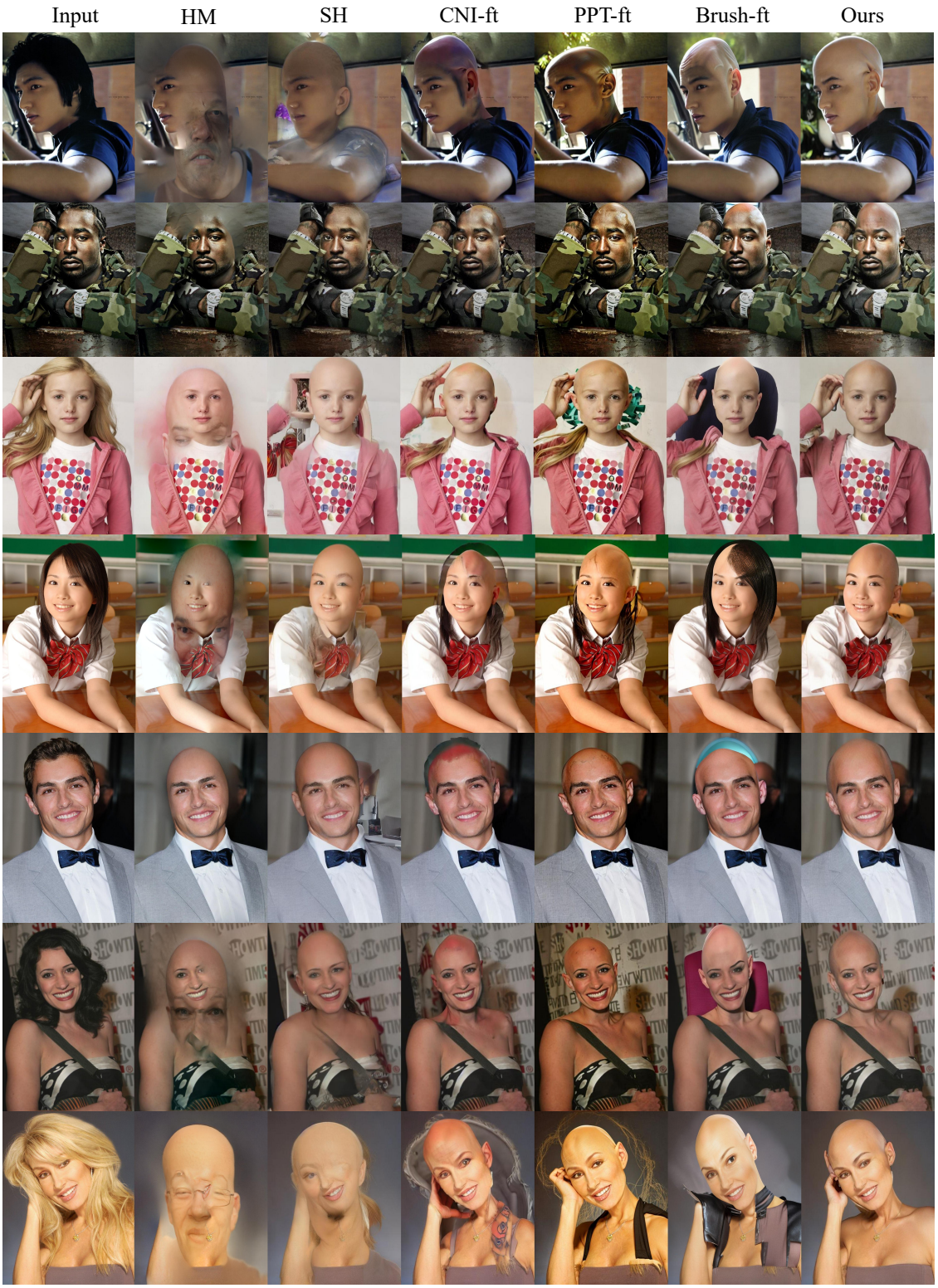

Figure 5. Additional qualitative results on CelebA dataset.

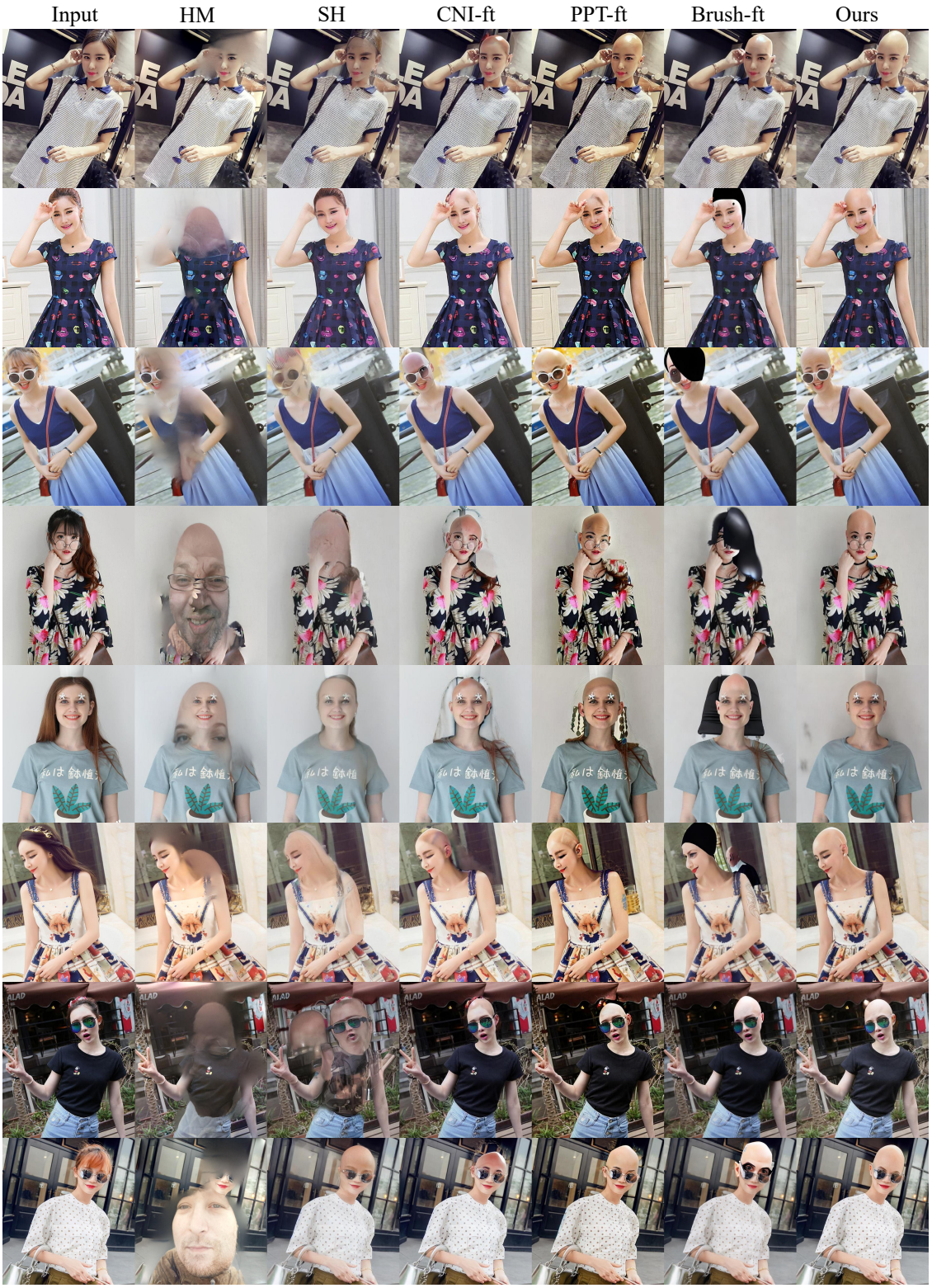

Figure 6. Additional qualitative results on DeepFashion2 Dataset.

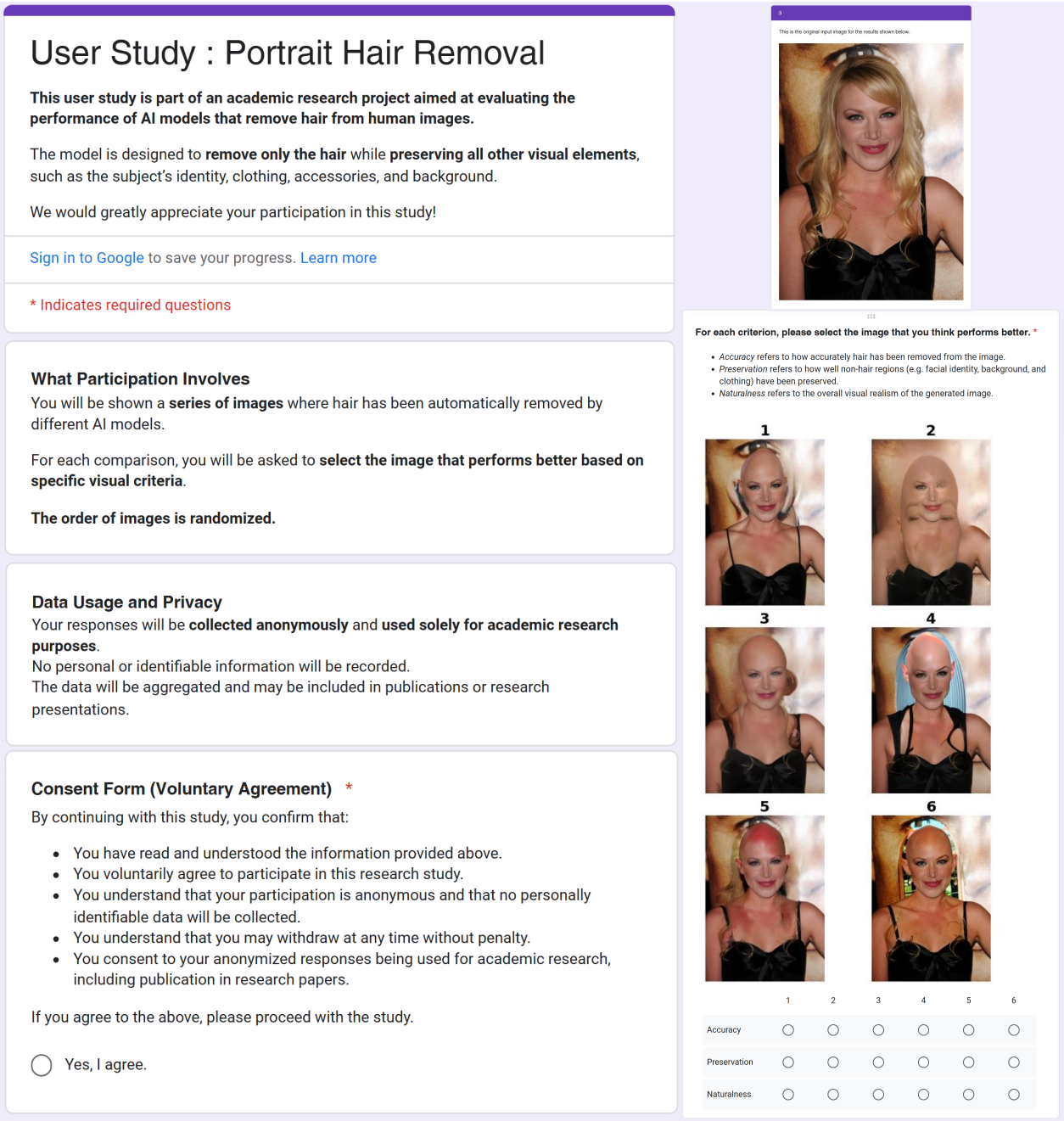

Figure 7. Overview of the user study format.

Category	Gender	Prompt
Head Pose	Male	his head looking over his left shoulder, glancing backward over his shoulder, with a soft expression
		his head looking up toward the light, eyes following the light above, with a confident gaze
		his head gazing down gently, eyes lowered with a serene expression, with a focused gaze
		his head facing forward, looking directly at the viewer, with a subtle grin
		his head head tilted downward, eyes looking up, eyes lifted gently upward, with a gentle smile
		his head turned to the right, gazing softly to the right, with a contemplative look
		his head glancing sideways with curiosity, eyes shifting sideways with interest, with a warm smile
		his head looking over his right shoulder, glancing back with a soft gaze, with a serene expression
	Female	her head chin slightly raised, gazing slightly upward with confidence, with a playful smirk
		her head head tilted downward, eyes looking up, eyes lifted gently upward, with a gentle smile
		her head turned to the right, gazing softly to the right, with a subtle grin
		her head glancing sideways with curiosity, eyes shifting sideways with interest, with a focused gaze
		her head turned to the left, gazing softly to the left, with a contemplative look
		her head looking over her right shoulder, glancing back with a soft gaze, with a gentle smile
		her head facing forward, looking directly at the viewer, with a confident gaze
		her head looking over her left shoulder, glancing backward over her shoulder, with a serene expression
Clothing	Male	in a long-sleeve casual pullover
		in a cozy knit sweater
		wearing a fitted black shirt
		wearing a soft cardigan
		in a breathable linen wrap top
		wearing a minimal cotton shirt
		in a tailored blazer over a shirt
		wearing a simple turtleneck
	Female	wearing a fitted black blouse
		wearing a sleeveless tank top
		wearing a flowy summer dress
		in a breathable linen wrap top
		in a minimal cotton shirt
		in a cozy knit sweater
		wearing a simple turtleneck
		dressed in a soft cardigan
Background	Male	on a park bench under autumn leaves, subtle morning light.
		at a wooden table in a rustic kitchen, soft daylight illuminating his scalp.
		in a greenhouse surrounded by ferns, gentle window light on his face.
		in front of a muted gray backdrop, filtered sunlight through leaves.
		at a wooden table in a rustic kitchen, ambient lighting creating a calm mood.
		inside a quiet library, subtle morning light.
		on a rooftop at dusk, glow of sunset behind him.
		in a peaceful garden courtyard, glow of sunset behind him.
	Female	by a large window in a home office, natural light from the side.
		on a park bench under autumn leaves, diffused light from the sky.
		in front of a muted gray backdrop, soft daylight illuminating her scalp.
		on a rooftop at dusk, diffused light from the sky.
		in front of a muted gray backdrop, subtle morning light.
		standing in front of a brick wall, glow of sunset behind her.
		on a park bench under autumn leaves, diffused light from the sky.
		inside a quiet library, gentle window light on her face.

Table 1. Sample prompts for bald augmentation.

Gender	Prompt
Male	medium-length straight hair with middle part, with bangs, in pastel pink, with soft glow
	buzz cut tousled hair with side part, no bangs, in brown, with a fade
	shoulder-length wavy hair with side part, with side-swept bangs, in silver gray, with a slight wave
	waist-length straight hair with middle part, with curtain bangs, in platinum blonde, with layered texture
	buzz cut slicked back hair with deep side part, with side-swept bangs, in golden brown, with a slight wave
	medium-length wavy hair no part, with bangs, in ombre, with an undercut
	short curly hair with side part, no bangs, in brown, with layered texture
	long tousled hair no part, with bangs, in black, with layered texture
	crew cut spiky hair with messy part, with curtain bangs, in silver gray, with an undercut
	long natural curls hair with middle part, with bangs, in golden brown, with soft glow
	short tousled hair with middle part, with short bangs, in black, with volume
	shoulder-length slicked back hair with middle part, with bangs, in silver gray, with a fade
	medium-length straight hair with messy part, with side-swept bangs, in black, with an undercut
	buzz cut slicked back hair with middle part, with curtain bangs, in golden brown, with a fade
	buzz cut spiky hair with messy part, no bangs, in ash blonde, with shaved sides
	crew cut natural curls hair with messy part, with side-swept bangs, in pastel pink, under natural sunlight
	long curly hair with side part, no bangs, in red, with a fade
	waist-length curly hair with middle part, with side-swept bangs, in ombre, with layered texture
	medium-length tousled hair with side part, with curtain bangs, in chestnut brown, with volume
	long wavy hair with messy part, with short bangs, in platinum blonde, with a slight wave
	shoulder-length slicked back hair with deep side part, no bangs, in pastel pink, with soft glow
	long tousled hair no part, no bangs, in chestnut brown, with highlights
	short natural curls hair with deep side part, with side-swept bangs, in golden brown, with an undercut
	long spiky hair no part, with side-swept bangs, in ombre, with shaved sides
	crew cut tousled hair with middle part, with curtain bangs, in silver gray, with a slight wave
Female	waist-length wavy hair with side part, with bangs, in silver gray, with hair clips
	very long natural curls hair no part, with bangs, in black, with hair clips
	long straight hair with side part, with bangs, in chestnut brown, under studio lighting
	long curly hair no part, with side-swept bangs, in silver gray, under natural sunlight
	long wavy hair with middle part, with bangs, in chestnut brown, softly glowing
	shoulder-length curly hair with side part, with side-swept bangs, in ash blonde
	very long wavy hair with side part, with side-swept bangs, in brown, softly glowing
	short loose waves hair no part, no bangs, in ash blonde, under studio lighting
	waist-length layered hair with side part, no bangs, in brown, with hair clips
	medium-length straight hair with middle part, no bangs, in ombre, with hair clip
	shoulder-length loose waves hair with middle part, with bangs, in pastel pink, with hair clips
	long tousled hair with side part, no bangs, in ash blonde, under natural sunlight
	long natural curls hair with middle part, with side-swept bangs, in red
	short wavy hair with middle part, no bangs, in chestnut brown, under natural sunlight
	shoulder-length wavy hair with middle part, no bangs, in silver gray, softly glowing
	short loose waves hair with middle part, with side-swept bangs, in blonde, under studio lighting
	shoulder-length layered hair with middle part, no bangs, in blonde, under natural sunlight
	short natural curls hair with middle part, with side-swept bangs, in silver gray, under studio lighting
	medium-length straight hair no part, no bangs, in chestnut brown, under natural sunlight
	very long natural curls hair no part, with bangs, in silver gray, under natural sunlight
	long wavy hair with middle part, with bangs, in ombre, under studio lighting
	short curly hair no part, no bangs, in ash blonde, with hair clips
	long loose waves hair with middle part, with side-swept bangs, in red, with hair clips
	very long layered hair with middle part, no bangs, in blonde, under natural sunlight
	shoulder-length layered hair with middle part, no bangs, in ombre, with hair clips

Table 2. Sample prompts for style augmentation.

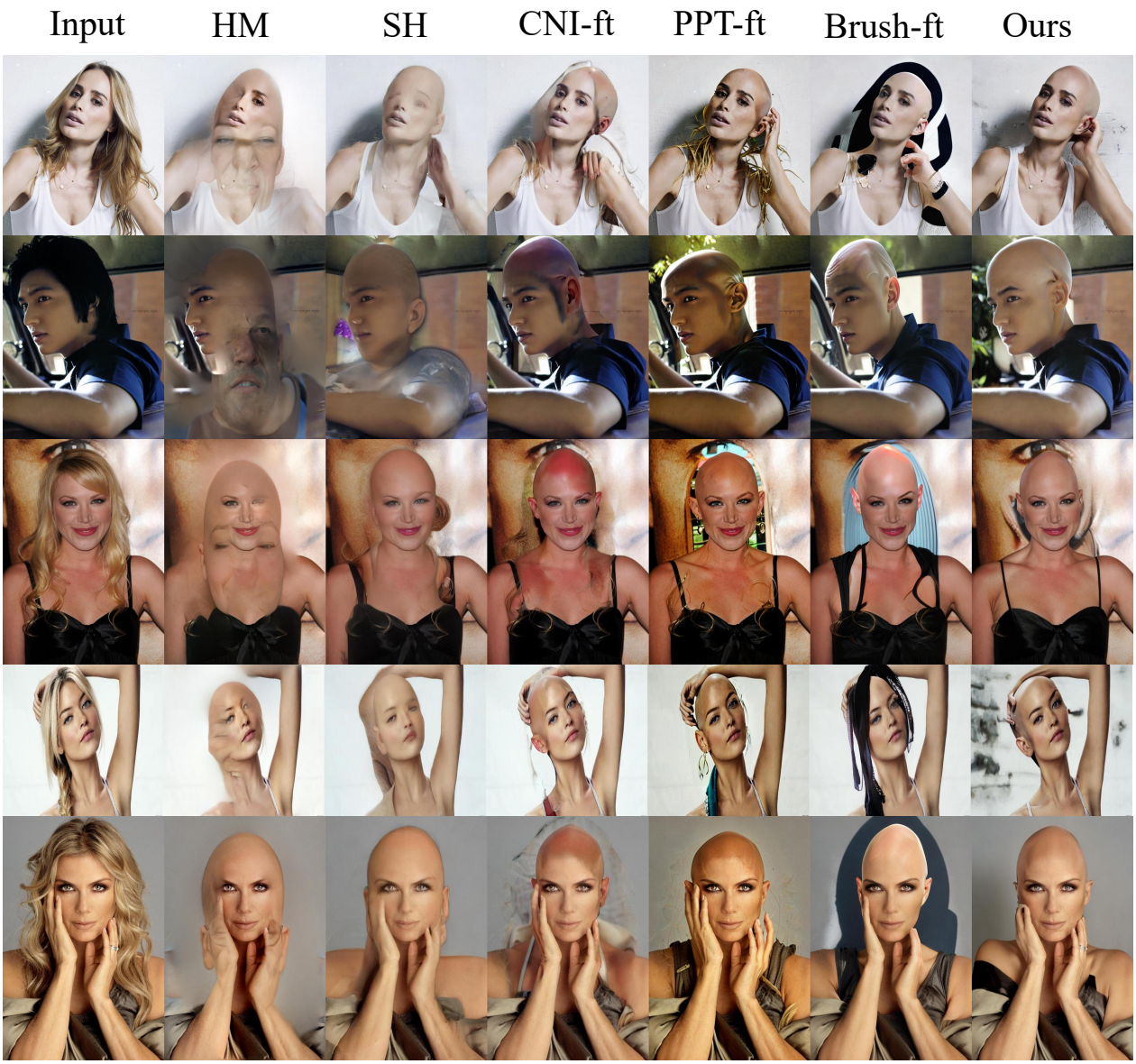

Figure 8. Qualitative comparison of CelebA images used in the user study.

Figure 9. Qualitative comparison of CelebA images used in the user study.

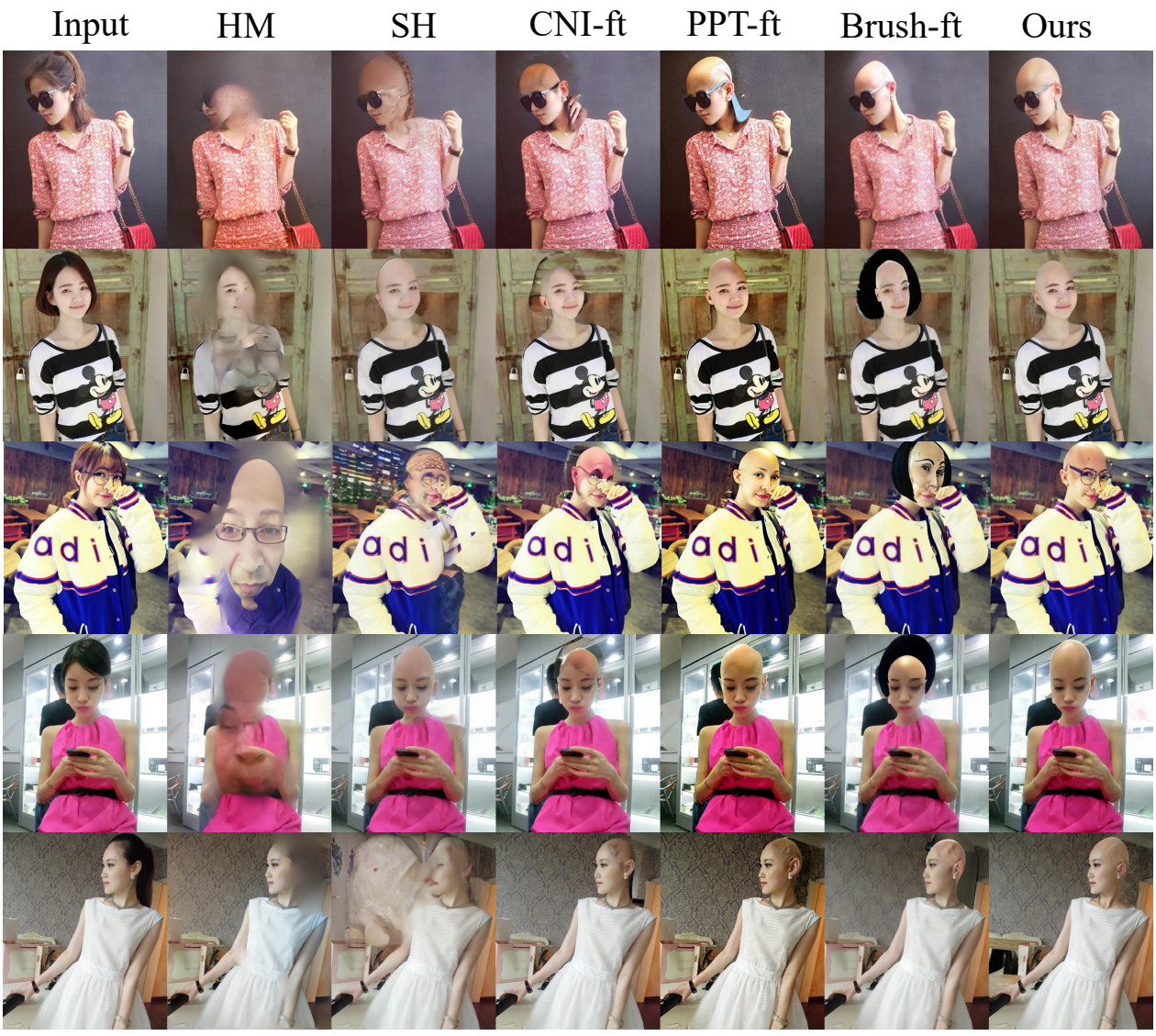

Figure 10. Qualitative comparison of DeepFashion2 images used in the user study.

Figure 11. Qualitative comparison of DeepFashion2 images used in the user study.

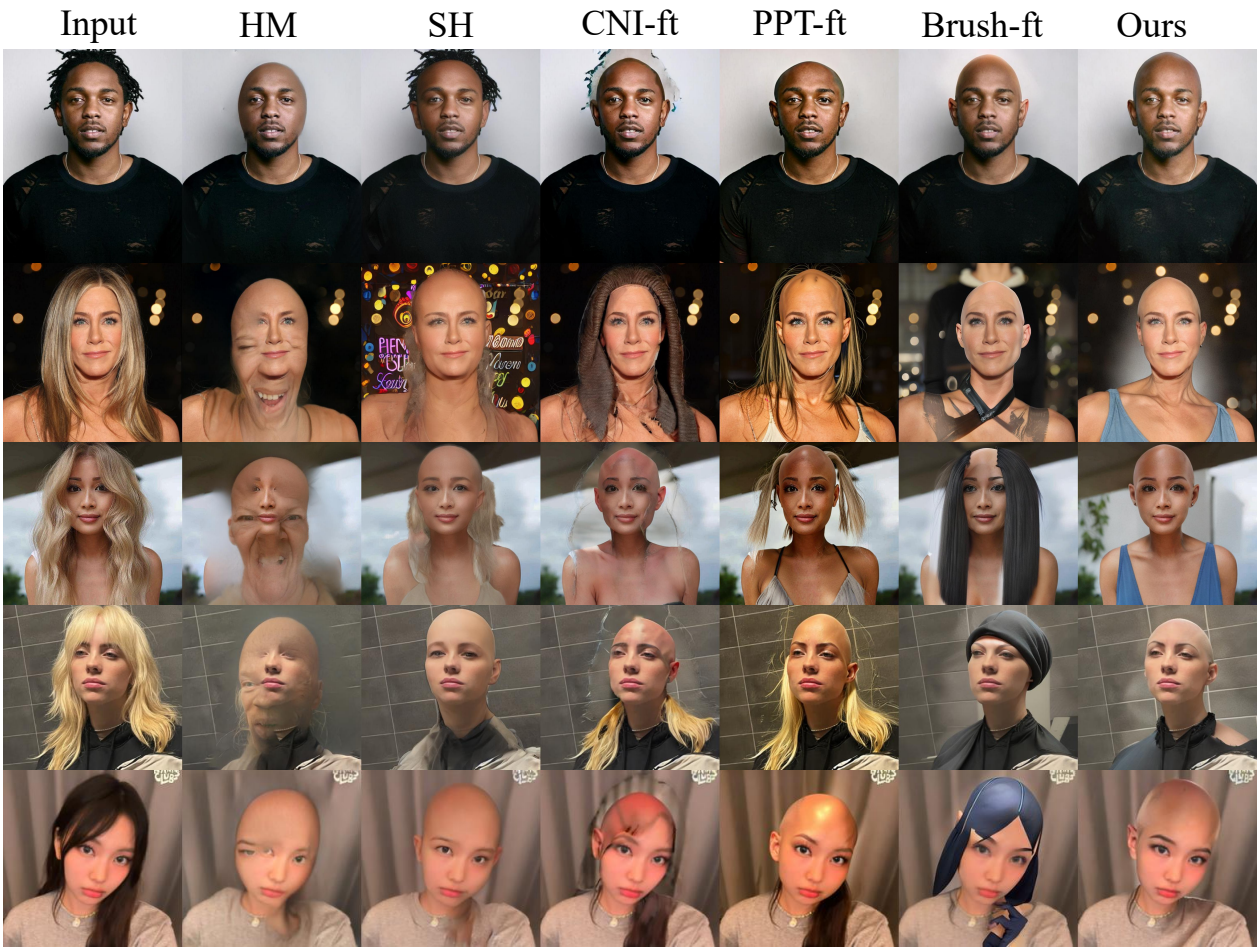

Figure 12. Qualitative comparison of collected internet images used in the user study.

Figure 13. Qualitative comparison of collected internet images used in the user study.

Figure 14. Multi-human hair removal using `AnyBald`.

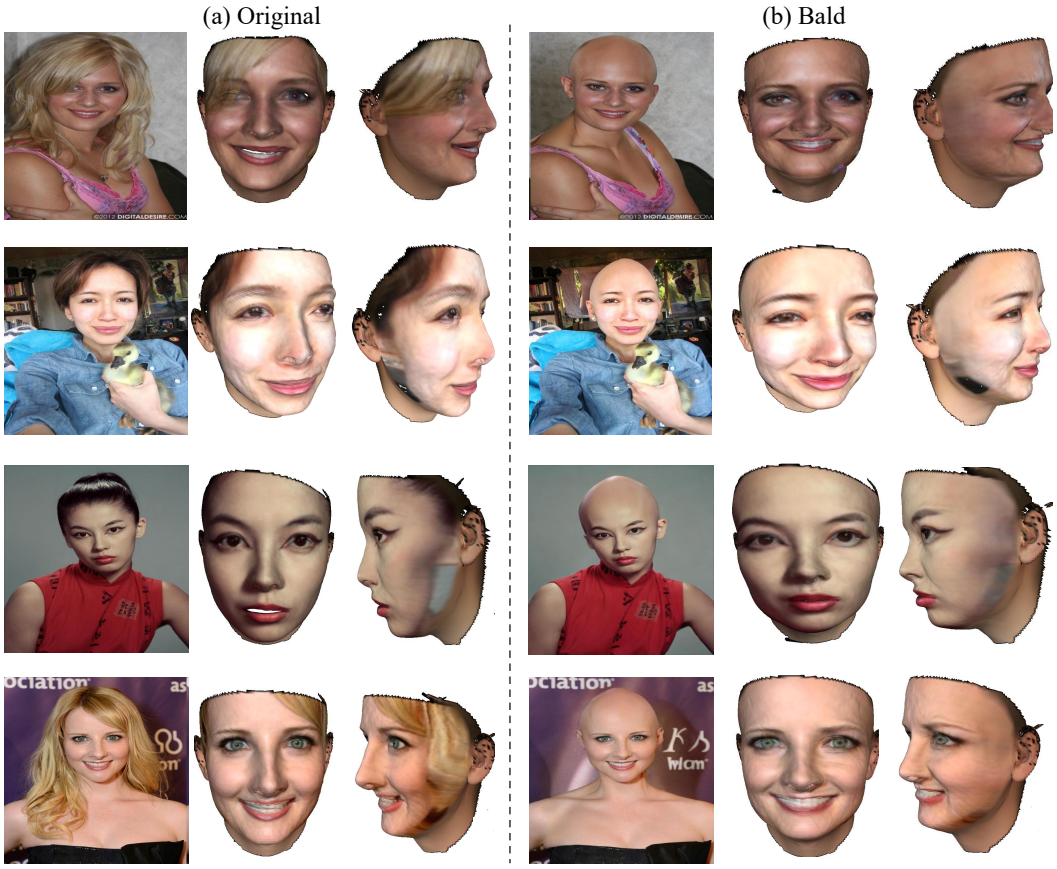

Figure 15. 3D facial reconstruction results with `AnyBald`, compared against the original input images.

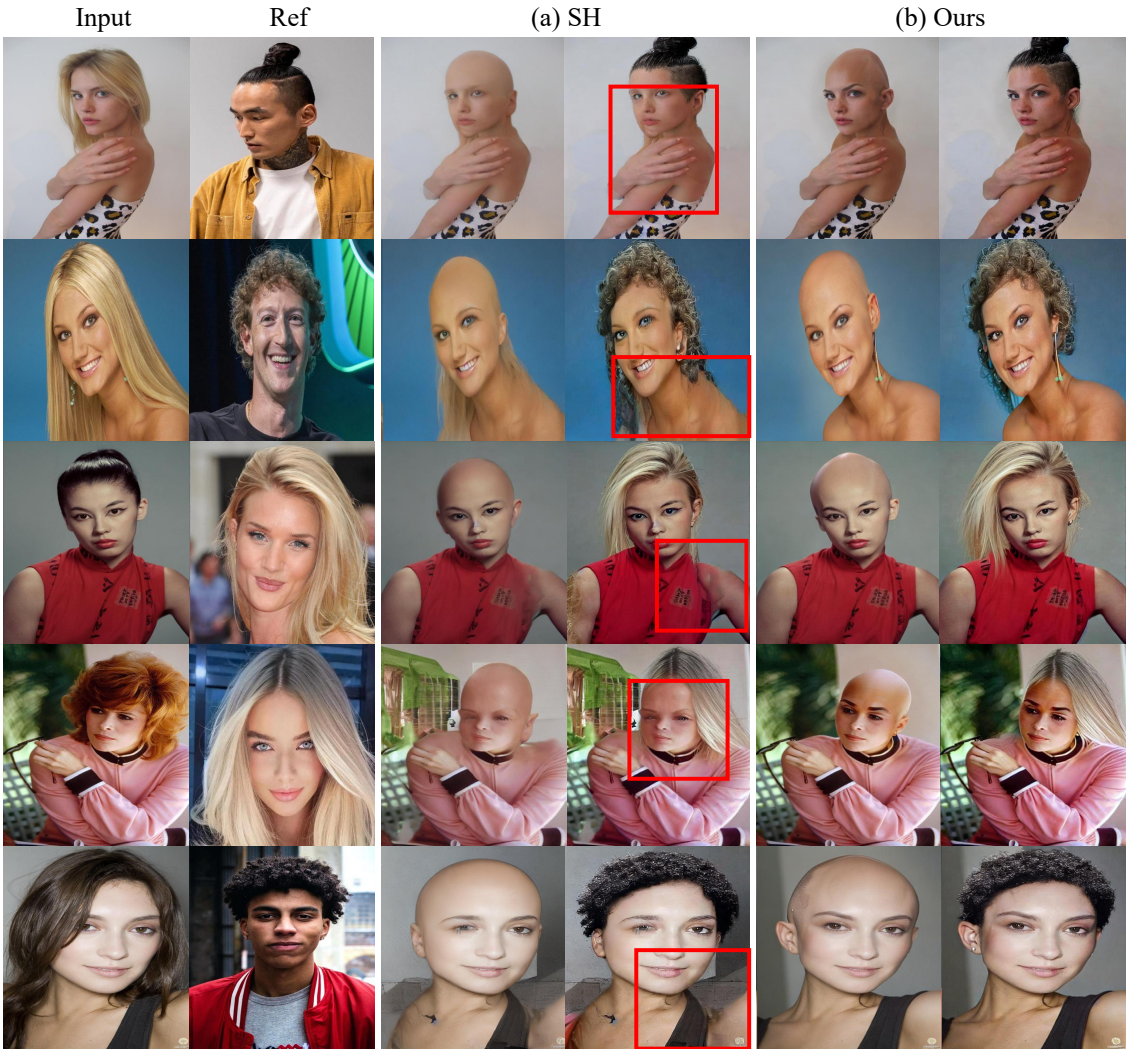

Figure 16. Comparison of hair transfer results between `AnyBald` and Stable-Hair. Results using `AnyBald` achieve more faithful and realistic outputs by leveraging clean bald images as robust intermediates, compared to Stable-Hair.

References

[1] Aiyu Cui, Jay Mahajan, Viraj Shah, Preeti Gomathinayagam, Chang Liu, and Svetlana Lazebnik. Street tryon: Learning in-the-wild virtual try-on from unpaired person images. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 8235–8239, 2024. 2

[2] Yao Feng, Haiwen Feng, Michael J Black, and Timo Bolkart. Learning an animatable detailed 3d face model from in-the-wild images. *ACM Transactions on Graphics (ToG)*, 40(4): 1–13, 2021. 3

[3] Yuying Ge, Ruimao Zhang, Xiaogang Wang, Xiaoou Tang, and Ping Luo. Deepfashion2: A versatile benchmark for detection, pose estimation, segmentation and re-identification of clothing images. In *Proceedings of the IEEE/CVF conference on Computer Vision and Pattern Recognition (CVPR)*, pages 5337–5345, 2019. 1, 2

[4] Xuan Ju, Xian Liu, Xintao Wang, Yuxuan Bian, Ying Shan, and Qiang Xu. Brushnet: A plug-and-play image inpainting model with decomposed dual-branch diffusion. In *European Conference on Computer Vision (ECCV)*, pages 150–168, 2024. 2

[5] Tero Karras, Samuli Laine, Miika Aittala, Janne Hellsten, Jaakko Lehtinen, and Timo Aila. Analyzing and improving the image quality of stylegan. In *Proceedings of the IEEE/CVF conference on Computer Vision and Pattern Recognition (CVPR)*, pages 8110–8119, 2020. 2

[6] Black Forest Labs. Flux.1 kontext: Flow matching for in-context image generation and editing in latent space, 2025. 1

[7] Ziwei Liu, Ping Luo, Xiaogang Wang, and Xiaoou Tang. Deep learning face attributes in the wild. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 3730–3738, 2015. 1, 2

[8] Yiqian Wu, Yong-Liang Yang, and Xiaogang Jin. Hairmapper: Removing hair from portraits using gans. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 4227–4236, 2022. 1, 2

[9] Enze Xie, Wenhai Wang, Zhiding Yu, Anima Anandkumar, Jose M Alvarez, and Ping Luo. Segformer: Simple and efficient design for semantic segmentation with transformers. *Advances in Neural Information Processing Systems (NIPS)*, 34:12077–12090, 2021. 1

[10] Yuxuan Zhang, Qing Zhang, Yiren Song, Jichao Zhang, Hao Tang, and Jiaming Liu. Stable-hair: Real-world hair transfer via diffusion model. In *Proceedings of the AAAI Conference on Artificial Intelligence*, pages 10348–10356, 2025. 2

[11] Junhao Zhuang, Yanhong Zeng, Wenran Liu, Chun Yuan, and Kai Chen. A task is worth one word: Learning with task prompts for high-quality versatile image inpainting. In *European Conference on Computer Vision (ECCV)*, pages 195–211, 2024. 2