# Face Anonymization Made Simple (Supplementary Material)

Han-Wei Kung[1] Tuomas Varanka[2] Sanjay Saha[3] Terence Sim[3] Nicu Sebe[1]

[1]University of Trento [2]University of Oulu [3]National University of Singapore

## Anonymization Degree vs. Identity Distance

Our anonymization approach features a single adjustable parameter, $d$, which controls the degree of anonymization. Figure 1 illustrates how the identity cosine distance, measured using the FaceNet [12] recognition model, changes with varying degrees of anonymization. This analysis was conducted using dozens of identities and seeds from our CelebA-HQ [7] and FFHQ [8] test sets. We selected 50 identities from each dataset. For each identity, we applied six different degrees of anonymization ($d$ values of 0.3, 0.6, 0.9, 1.2, 1.4, 1.5). For each anonymization degree, we created 10 variations using 10 different seed values. The plot displays the average cosine distance of these variations for each anonymization degree. It reveals a clear trend: as the degree of anonymization increases, the identity distance between the anonymized and original images grows wider.

## Optimizing Face Anonymization

Increasing the anonymization degree parameter, $d$, results in a greater divergence from the original identity, but beyond a certain range, it can hinder the model's ability to generate realistic faces. We followed prior research methodologies [1, 4, 6] and used face detection to assess the validity of synthesized faces. We applied six different anonymization levels ($d$ values of 0.3, 0.6, 0.9, 1.2, 1.4, and 1.5) to 250 facial images from our CelebA-HQ [7] test set and evaluated the detection rate with two face detectors, RetinaFace [2] and Dlib [9]. As shown in Fig. 2, when $d$ reaches 1.5, the face detectors begin to flag invalid faces among the 250 generated, indicating that $d$ values above 1.5 are unsuitable for maintaining realistic outputs. Figure 2 also includes an example illustrating the unrealistic faces generated by our model at higher $d$ values.

While higher $d$ values create more distinct face shapes for anonymization, they compromise the preservation of non-identity related facial attributes. We again applied the same six levels of anonymization to 250 facial images from our CelebA-HQ [7] test set and measured the attribute distances for face shape, pose, gaze, and expression. Figure 3a reveals that higher $d$ values produce more distinctive face shapes, but this comes at the cost of preserving non-identity-related attributes. This trend is further illustrated in Figs. 3b to 3d, aligning with our expectations.

We also evaluated the attribute distance performance of two state-of-the-art anonymization methods, FALCO [1] and DP2 [5], on the same 250 test images, presenting their

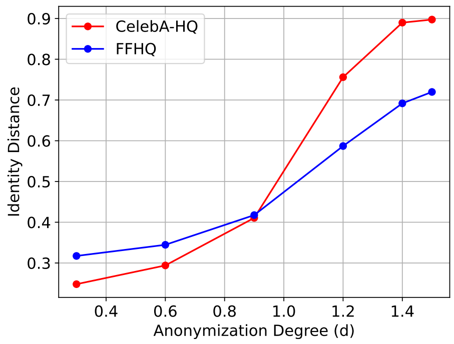


Figure 1. Relationship between degree of anonymization and identity distance.

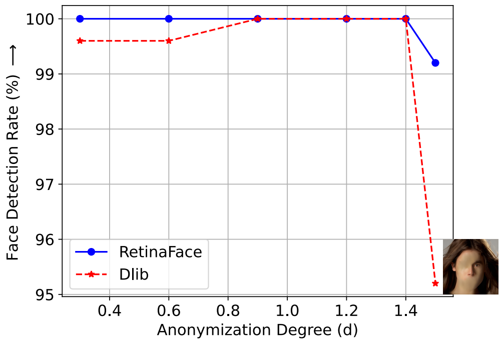
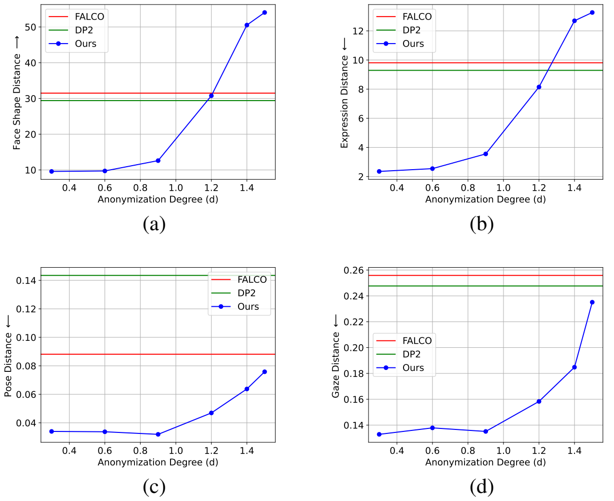


Figure 2. Face detection rates at various anonymization degrees.

|  | Identity Distance |  |  |  | Attribute Distance |  |  |  |  |  | Image Quality |  |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
|  | Re-ID ↑ |  | Shape ↓ |  | Pose ↓ |  | Gaze ↓ |  | Expression ↓ |  | Face IQA ↑ |  |
|  | CelebA-HQ | FFHQ | CelebA-HQ | FFHQ | CelebA-HQ | FFHQ | CelebA-HQ | FFHQ | CelebA-HQ | FFHQ | CelebA-HQ | FFHQ |
| DiffSwap [14] | 0.114 | 0.162 | 16.590 | 17.914 | 0.034 | 0.041 | 0.151 | 0.164 | **2.682** | **3.195** | 0.549 | 0.542 |
| BlendFace [13] | 0.693 | 0.642 | 13.234 | 16.497 | **0.028** | **0.036** | **0.120** | **0.148** | 3.170 | 4.102 | 0.527 | 0.511 |
| InSwapper [3] | **0.871** | **0.830** | **11.558** | **14.300** | 0.035 | 0.042 | 0.158 | 0.177 | 4.197 | 4.872 | 0.364 | 0.371 |
| Ours | 0.566 | 0.310 | 17.211 | 22.312 | 0.036 | 0.043 | 0.139 | 0.149 | 4.067 | 4.745 | **0.728** | **0.720** |

Table 1. Quantitative results on the task of face swapping for CelebA-HQ [7] and FFHQ [8] test sets, with the best results highlighted in bold and the second-best results underlined.



(a) (b) (c) (d)

Figure 3. Changes in facial features at different anonymization degrees.

results in  Fig. 3. For the FALCO [1] method, we set its identity loss margin value to 0. This configuration maximizes the identity difference between the anonymized result and the original image, but it compromises the preservation of non-identity facial attributes. Figure 3a reveals that when the $d$ value exceeds 1.2, our method begins to outperform these methods in producing more distinctive face shapes. Additionally, our method may continue to outperform them in preserving facial expressions until $d$ reaches approximately 1.3, as illustrated in Fig. 3b. Based on our empirical results, we recommend an optimal $d$ range of 1.2 to 1.3 to achieve the ideal balance between identity obfuscation and attribute preservation. Within this range, our method also demonstrates superior performance compared to current state-of-the-art anonymization techniques.

## Dataset Preparation for Model Training

For training our model, we used three datasets: CelebRef-HQ [11], CelebA-HQ [7], and FFHQ [8]. We used all 10,555 images from CelebRef-HQ [11], which contains 1,005 identities with multiple images per identity showing varied expressions and angles. For CelebA-HQ [7] and FFHQ [8], we employed face recognition to identify same-person images, selecting 6,203 images (2,506 identities) from CelebA-HQ [7] and 7,816 images (2,887 identities) from FFHQ [8]. For each identity, we randomly chose two images: one as the source and one as ground truth. We then created a synthesized driving image by using a state-of-the-art face swapping model [3] to replace the face in the ground truth image with another person's face. This process resulted in 153,414 source-driving image pairs for training: 49,518 from CelebA-HQ [7], 42,188 from CelebRef-HQ [11], and 61,708 from FFHQ [8].

## Qualitative Results of Anonymization

Figures 4 to 9 present additional qualitative results of our anonymization technique. We showcase these results using images from our test sets in FFHQ [8] and CelebA-HQ [7] databases. We also compare our method's performance against the same set of anonymization techniques [1, 5, 10] discussed earlier in our paper.

## Face Swapping Results

Although face swapping is not the primary focus of our research, our model initially develops this capability as part of its anonymization process. To demonstrate its effectiveness, we present additional face swapping examples using the FFHQ [8] and CelebA-HQ [7] datasets in  Figs. 10 to 13. We also compare our results to established face swapping benchmarks [3, 13, 14] discussed in our paper. These comparisons showcase our model's superior ability to generate high-quality facial images.

Furthermore, Tab. 1 provides quantitative results for the face swapping tasks. These results indicate that our model achieves superior Image Quality Assessment (IQA) scores across both datasets. While both DiffSwap [14] and our model can natively generate high-resolution images at 512 × 512, our model achieves an IQA score that is more than 30% higher than DiffSwap's [14]. This improvement is likely due to our use of ReferenceNet, which encodes fine-grained features and enables our model to produce higher quality facial images.

## Societal Impact of AI-Generated Faces

AI-generated faces present a dual challenge in our digital world. While they can enhance privacy by offering anonymity, they also create opportunities for malicious activities. Scammers might use these synthetic identities to produce more convincing deceptions, potentially eroding trust in online interactions and media. To address these risks, a comprehensive strategy is essential. This includes technological solutions such as advanced watermarking and AI detection systems, along with legal frameworks regulating the use of synthetic faces. Additionally, raising public awareness about this technology and its potential misuse is crucial. Establishing clear industry standards for the ethical creation and application of AI-generated faces will help balance their benefits while protecting social trust. A coordinated effort across technological, legal, and educational fronts is vital for maximizing the positive potential of this innovation while minimizing its societal drawbacks.

## References

[1] Simone Barattin, Christos Tzelepis, Ioannis Patras, and Nicu Sebe. Attribute-preserving face dataset anonymization via latent code optimization. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 8001–8010, 2023. 1, 2, 4, 5, 6

[2] Jiankang Deng, Jia Guo, Evangelos Ververas, Irene Kotsia, and Stefanos Zafeiriou. Retinaface: Single-shot multi-level face localisation in the wild. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 5203–5212, 2020. 1

[3] Jia Guo, Jiankang Deng, Xiang An, Jack Yu, and Baris Gecer. InsightFace Swapper. `https://github.com/deepinsight/insightface/tree/master/examples/in_swapper/`. 2, 10, 11, 12, 13

[4] Majed El Helou, Doruk Cetin, Petar Stamenkovic, and Fabio Zund. Vera: Versatile anonymization fit for clinical facial images. *arXiv preprint arXiv:2312.02124*, 2023. 1

[5] Håkon Hukkelås and Frank Lindseth. Deepprivacy2: Towards realistic full-body anonymization. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pages 1329–1338, 2023. 1, 2, 4, 5, 6, 7, 8, 9

[6] Håkon Hukkelås, Rudolf Mester, and Frank Lindseth. Deepprivacy: A generative adversarial network for face anonymization. In *International symposium on visual computing*, pages 565–578. Springer, 2019. 1

[7] Tero Karras, Timo Aila, Samuli Laine, and Jaakko Lehtinen. Progressive growing of gans for improved quality, stability, and variation. *arXiv preprint arXiv:1710.10196*, 2017. 1, 2, 4, 5, 6, 10, 11

[8] Tero Karras, Samuli Laine, and Timo Aila. A style-based generator architecture for generative adversarial networks. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 4401–4410, 2019. 1, 2, 7, 8, 9, 12, 13

[9] Davis E King. Dlib-ml: A machine learning toolkit. *The Journal of Machine Learning Research*, 10:1755–1758, 2009. 1

[10] Dongze Li, Wei Wang, Kang Zhao, Jing Dong, and Tieniu Tan. Riddle: Reversible and diversified de-identification with latent encryptor. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 8093–8102, 2023. 2, 7, 8, 9

[11] Xiaoming Li, Shiguang Zhang, Shangchen Zhou, Lei Zhang, and Wangmeng Zuo. Learning dual memory dictionaries for blind face restoration. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(5):5904–5917, 2022. 2

[12] Florian Schroff, Dmitry Kalenichenko, and James Philbin. Facenet: A unified embedding for face recognition and clustering. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 815–823, 2015. 1

[13] Kaede Shiohara, Xingchao Yang, and Takafumi Taketomi. Blendface: Re-designing identity encoders for face-swapping. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 7634–7644, 2023. 2, 10, 11, 12, 13

[14] Wenliang Zhao, Yongming Rao, Weikang Shi, Zuyan Liu, Jie Zhou, and Jiwen Lu. Diffswap: High-fidelity and controllable face swapping via 3d-aware masked diffusion. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 8568–8577, 2023. 2, 10, 11, 12, 13

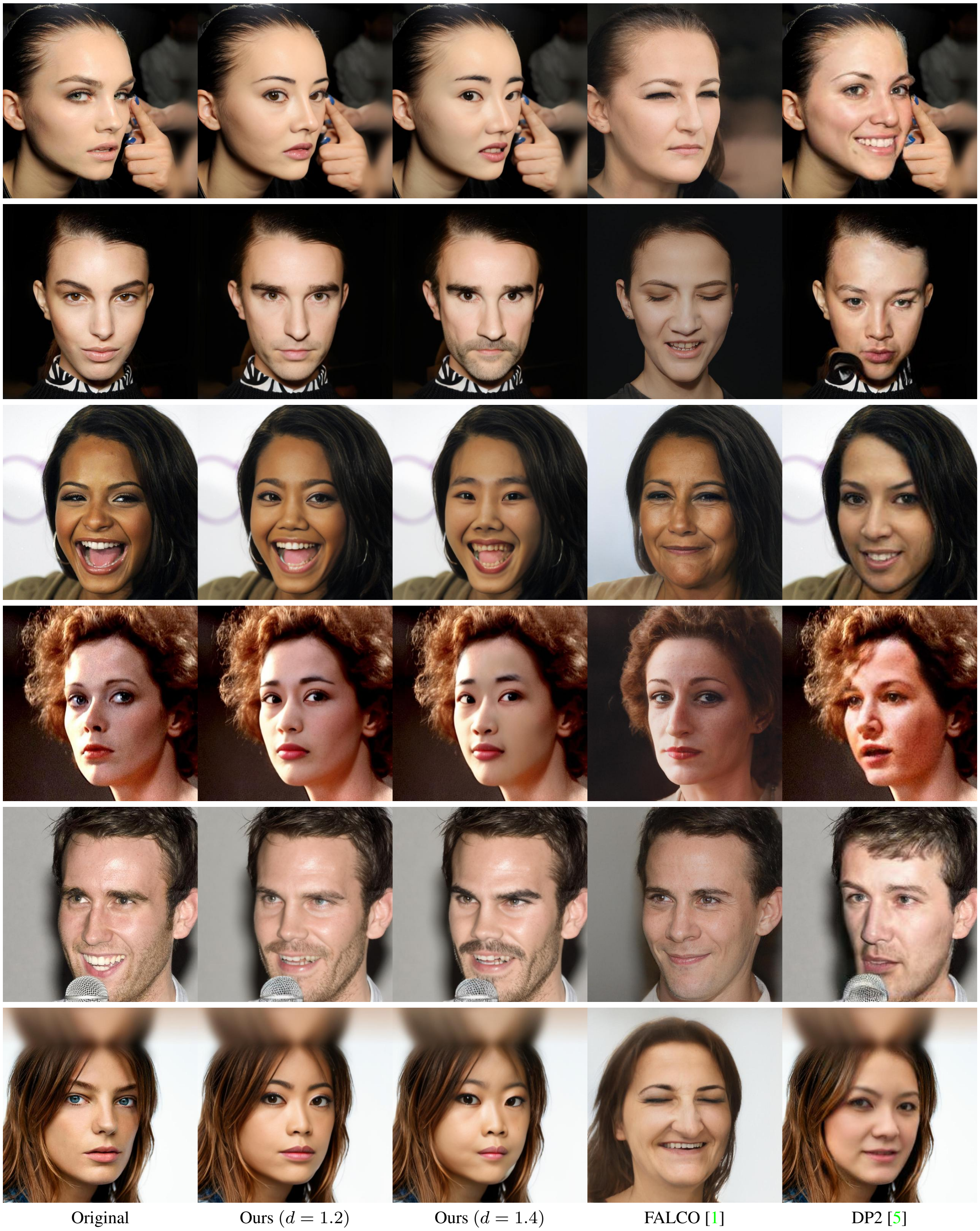


Figure 4. Qualitative results on the task of face anonymization for CelebA-HQ [7] test set.

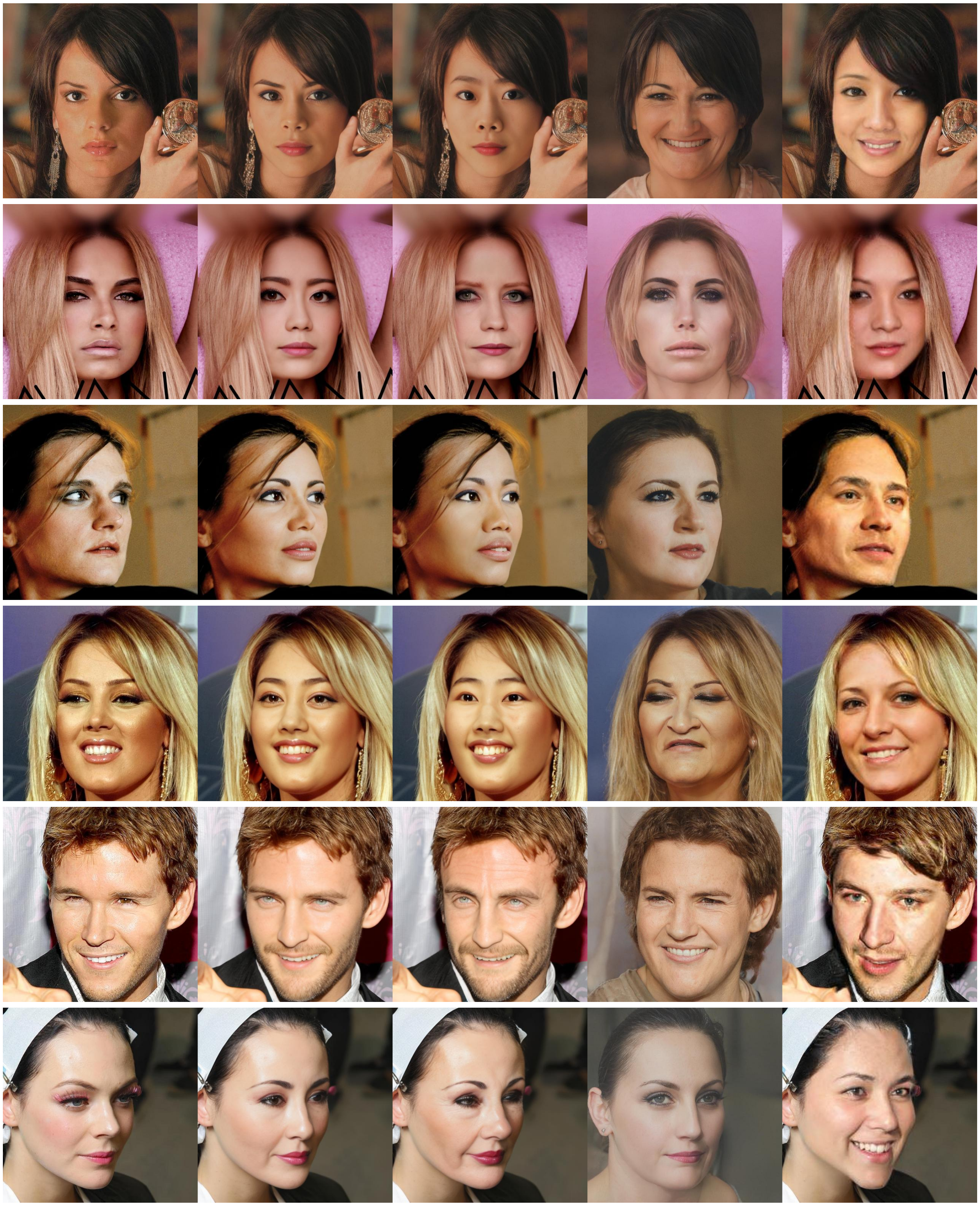
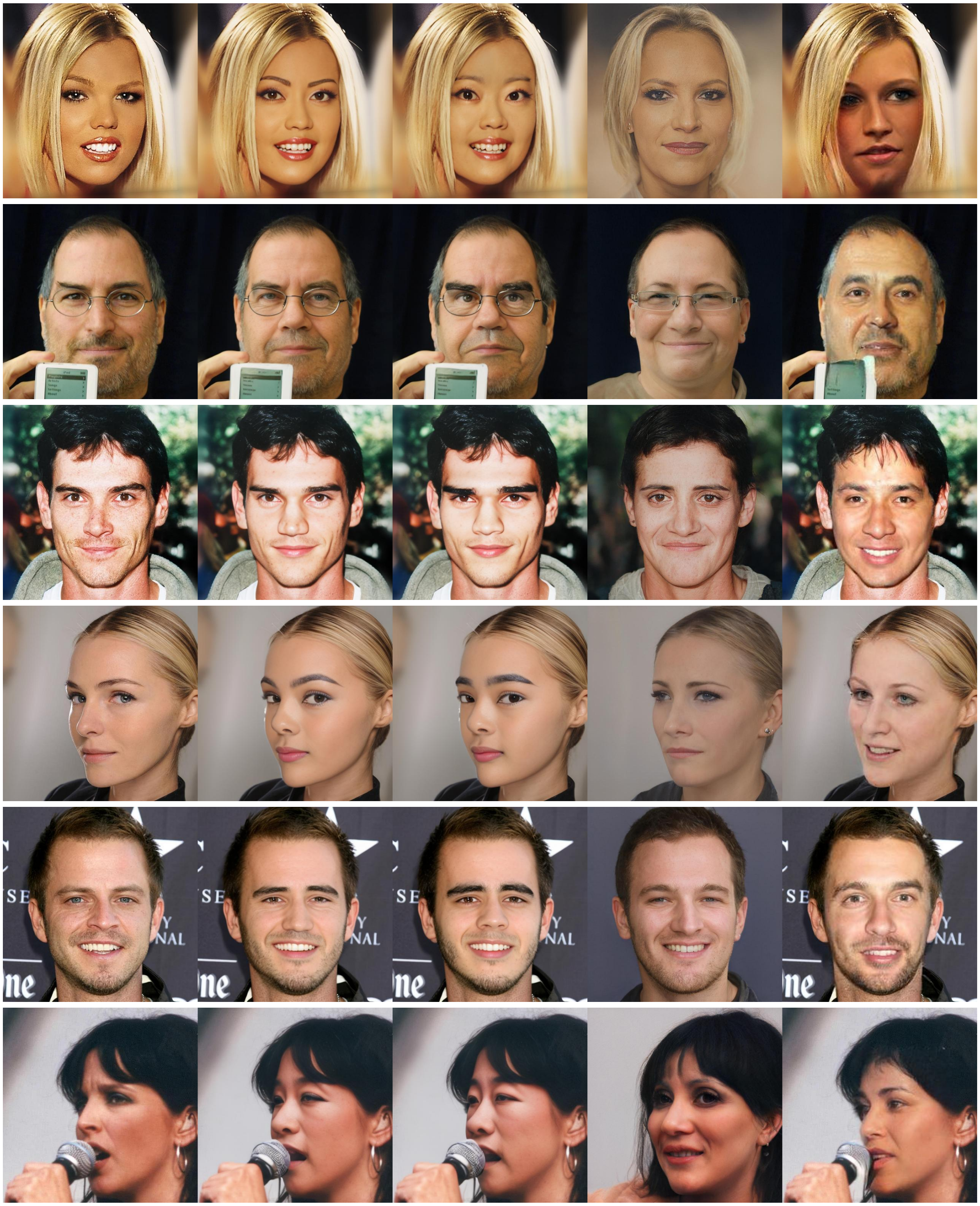


Figure 5. Qualitative results on the task of face anonymization for CelebA-HQ [7] test set.



Original Ours ($(d = 1.2)$) Ours ($(d = 1.4)$) FALCO [1] DP2 [5]

Figure 6. Qualitative results on the task of face anonymization for CelebA-HQ [7] test set.

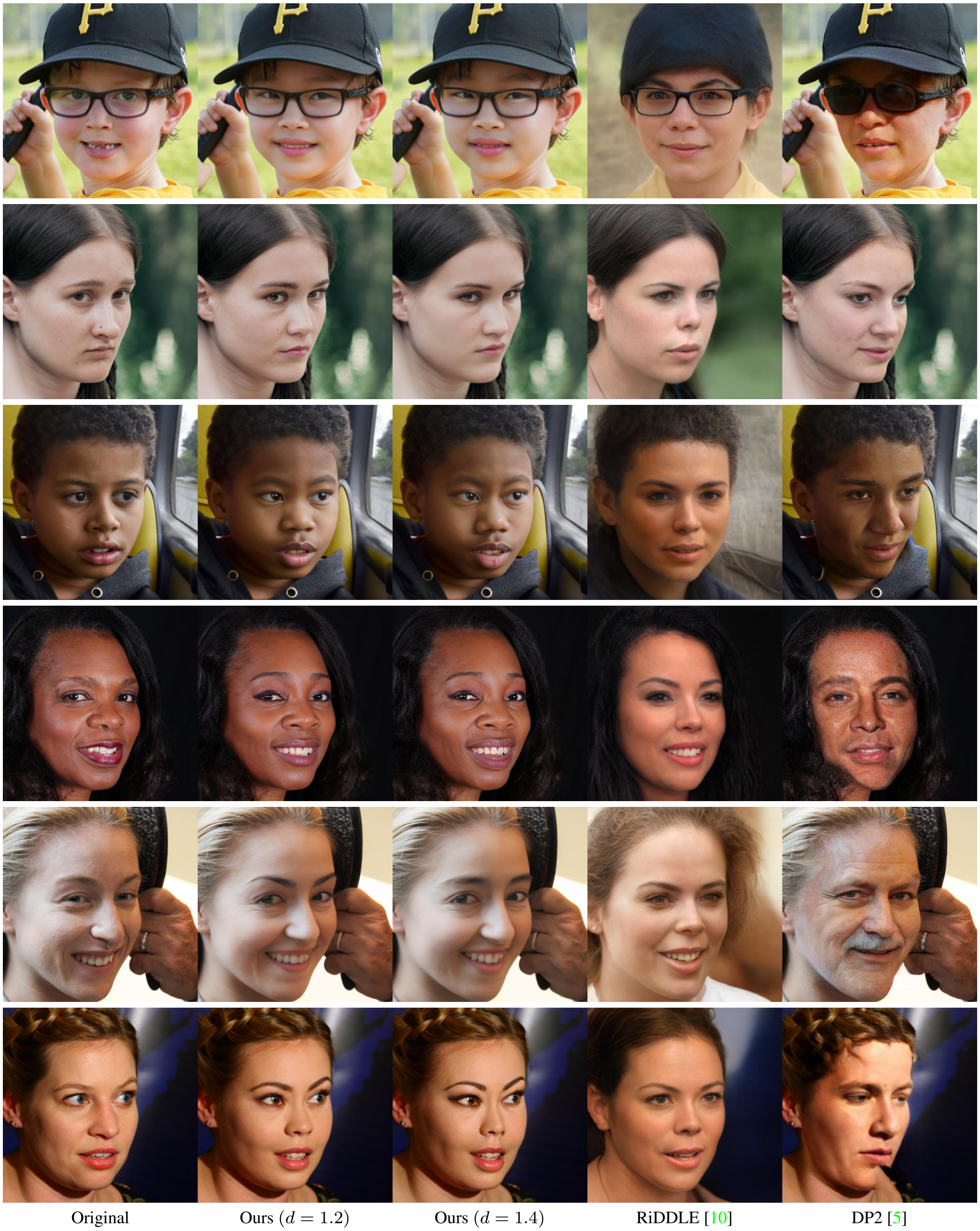


Figure 7. Qualitative results on the task of face anonymization for FFHQ [8] test set.

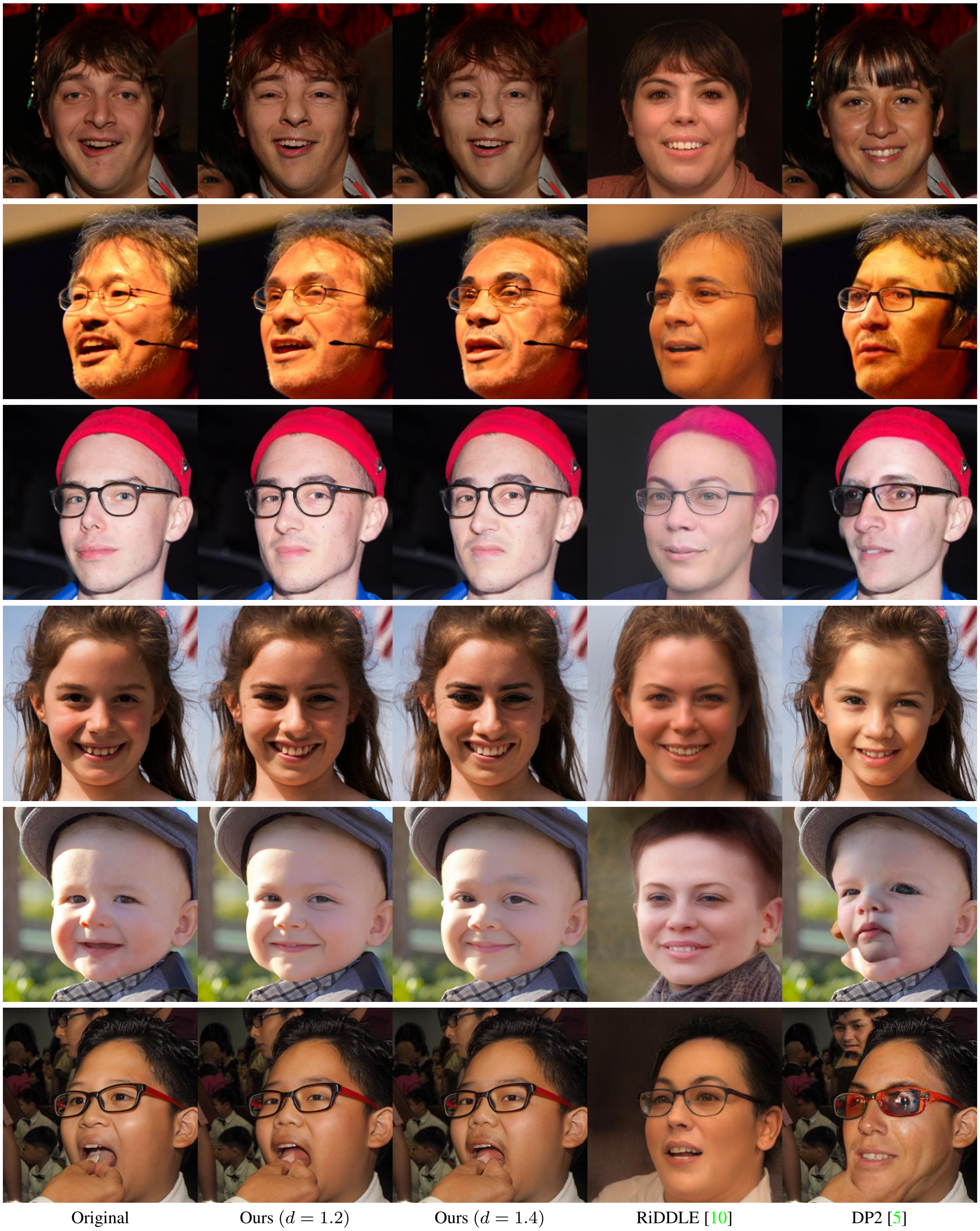


Figure 8. Qualitative results on the task of face anonymization for FFHQ [8] test set.

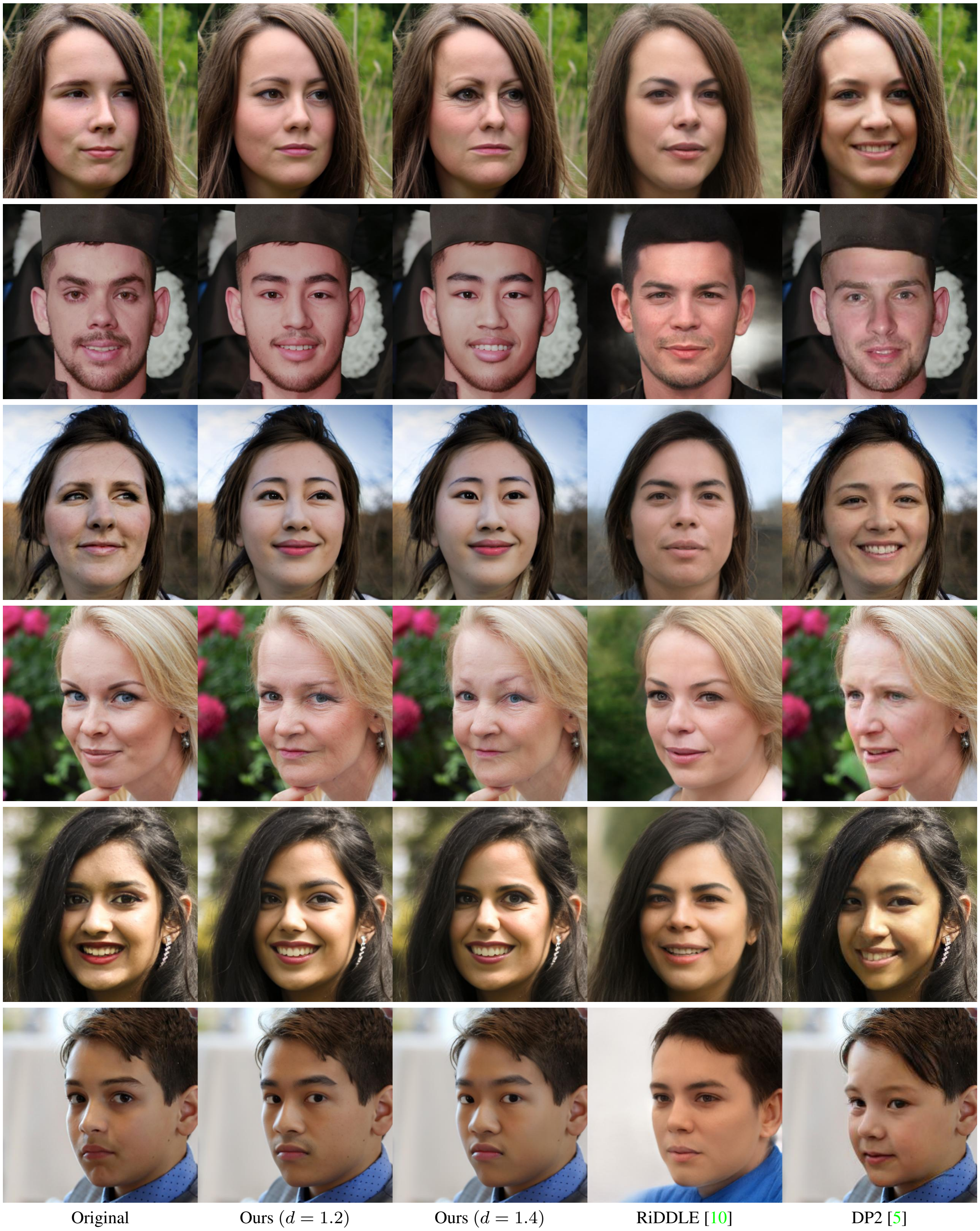


Figure 9. Qualitative results on the task of face anonymization for FFHQ [8] test set.

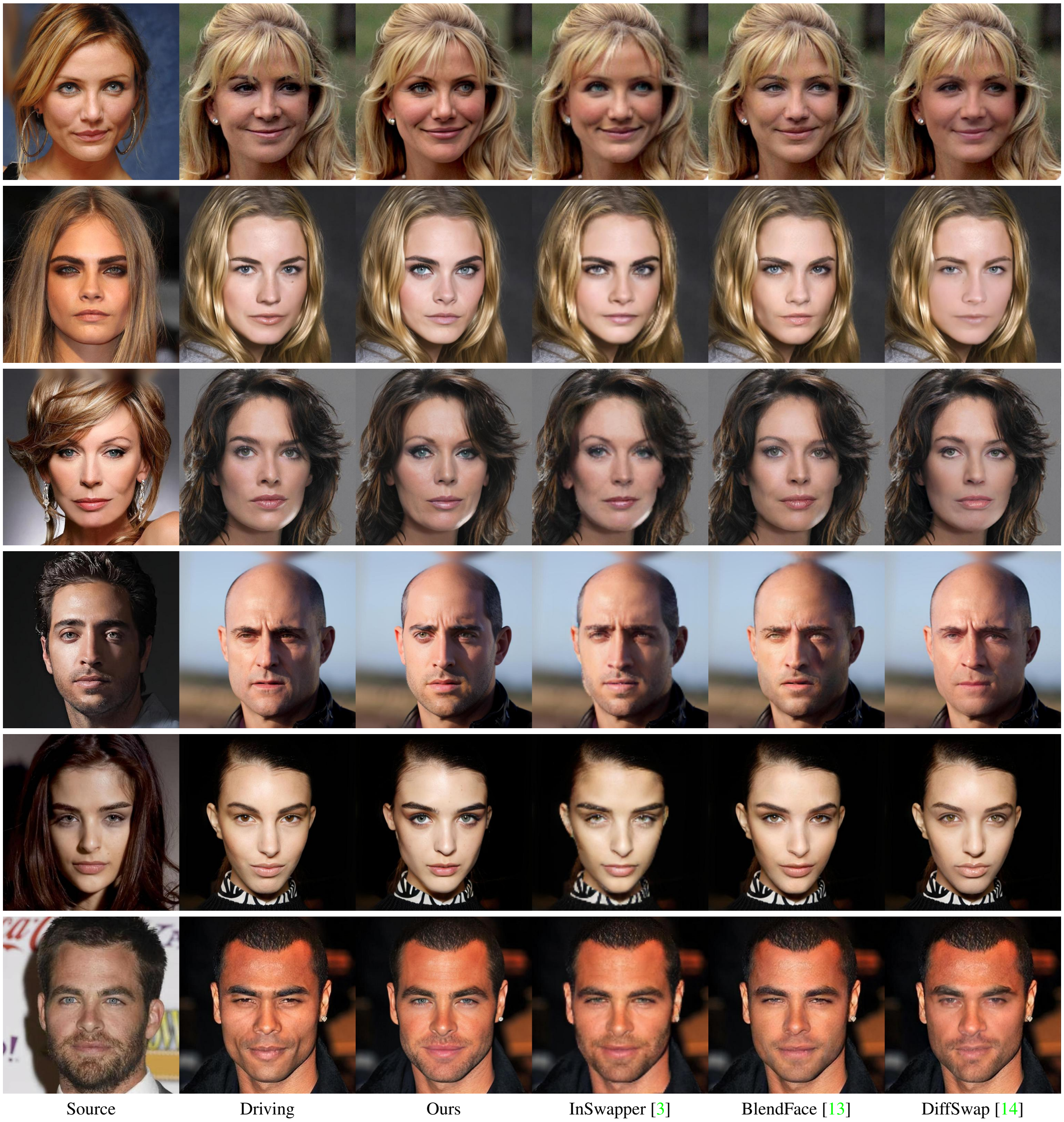


Figure 10. Qualitative results on the task of face swapping for CelebA-HQ [7] test set.

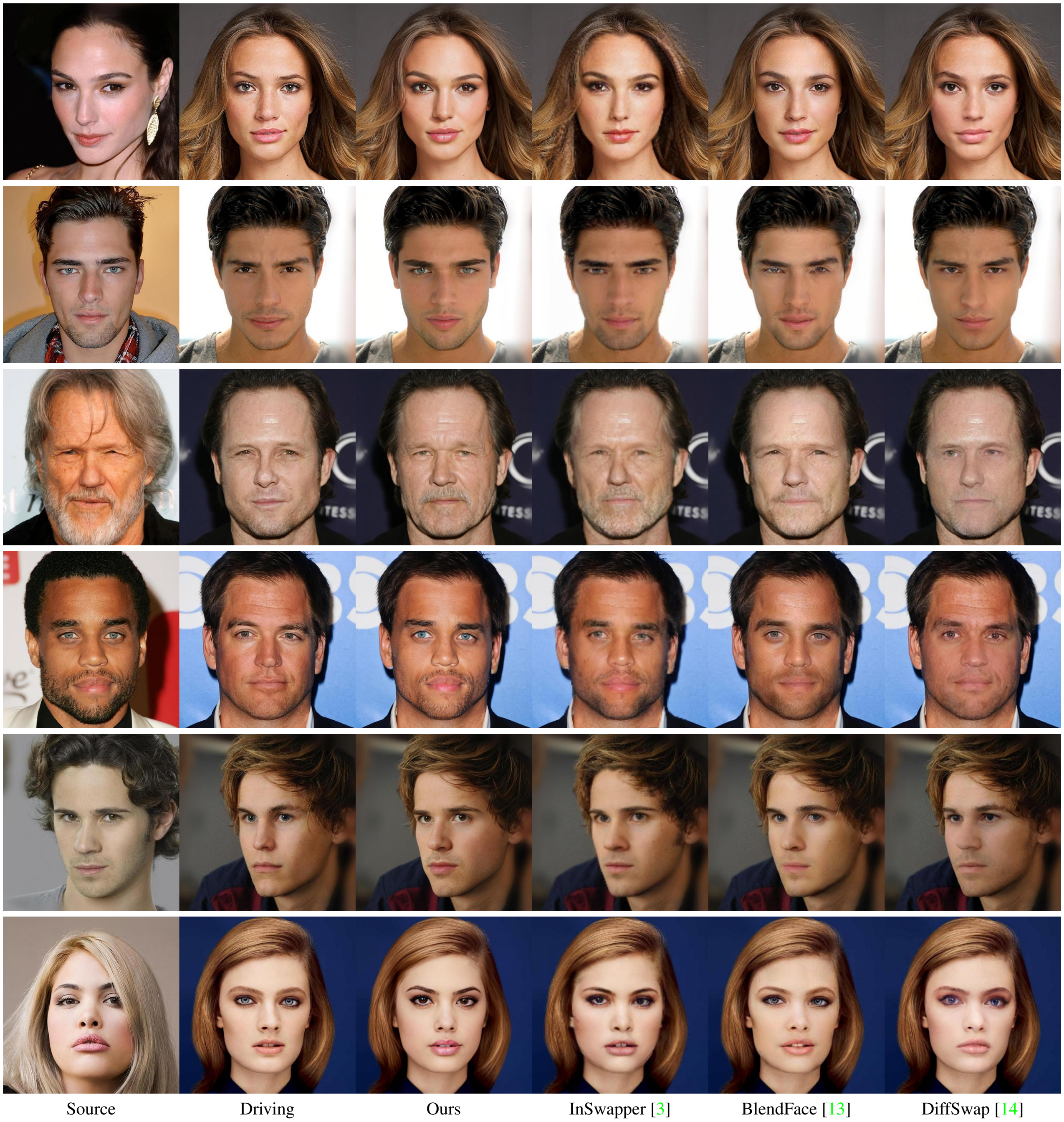


Figure 11. Qualitative results on the task of face swapping for CelebA-HQ [7] test set.

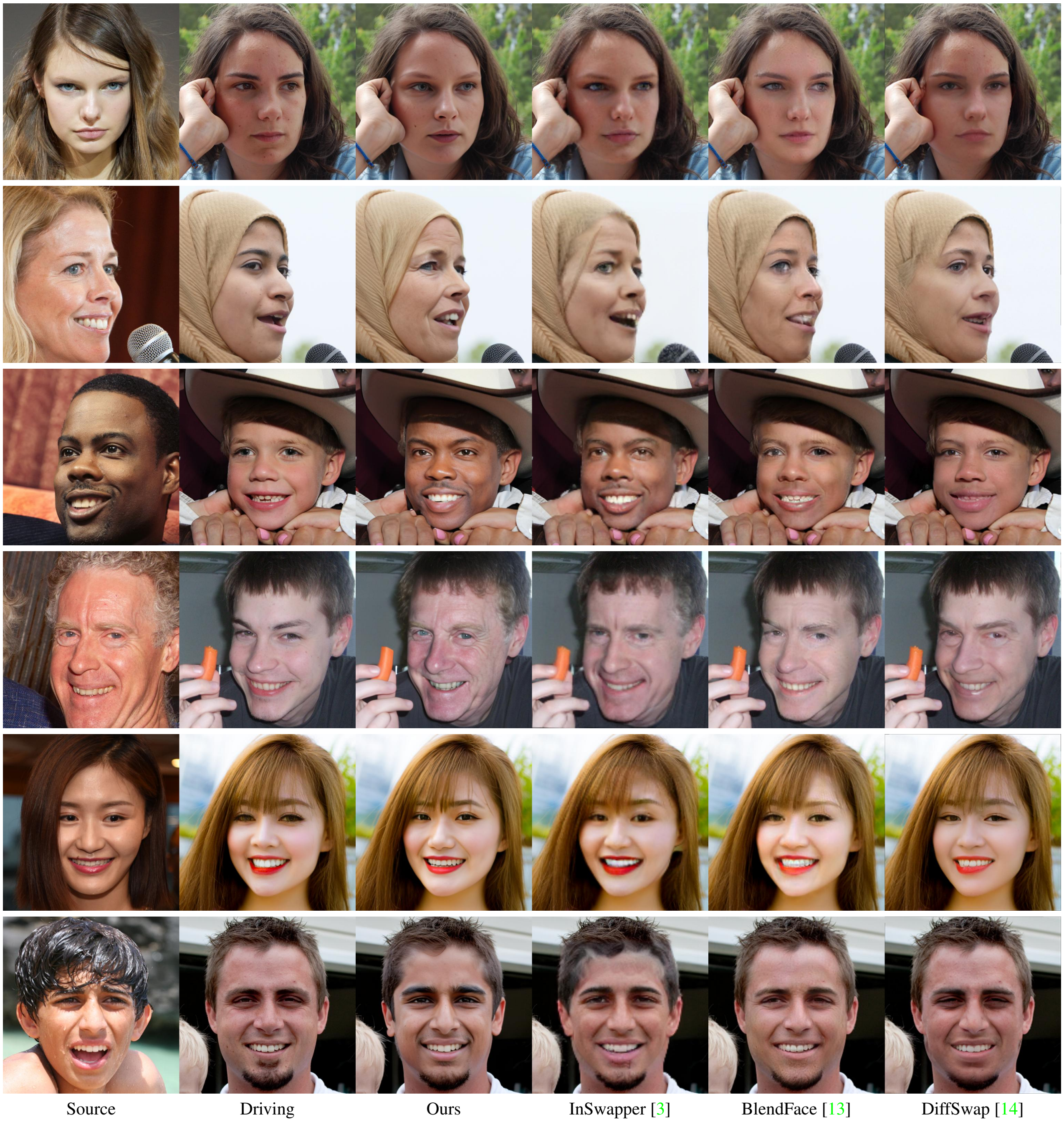
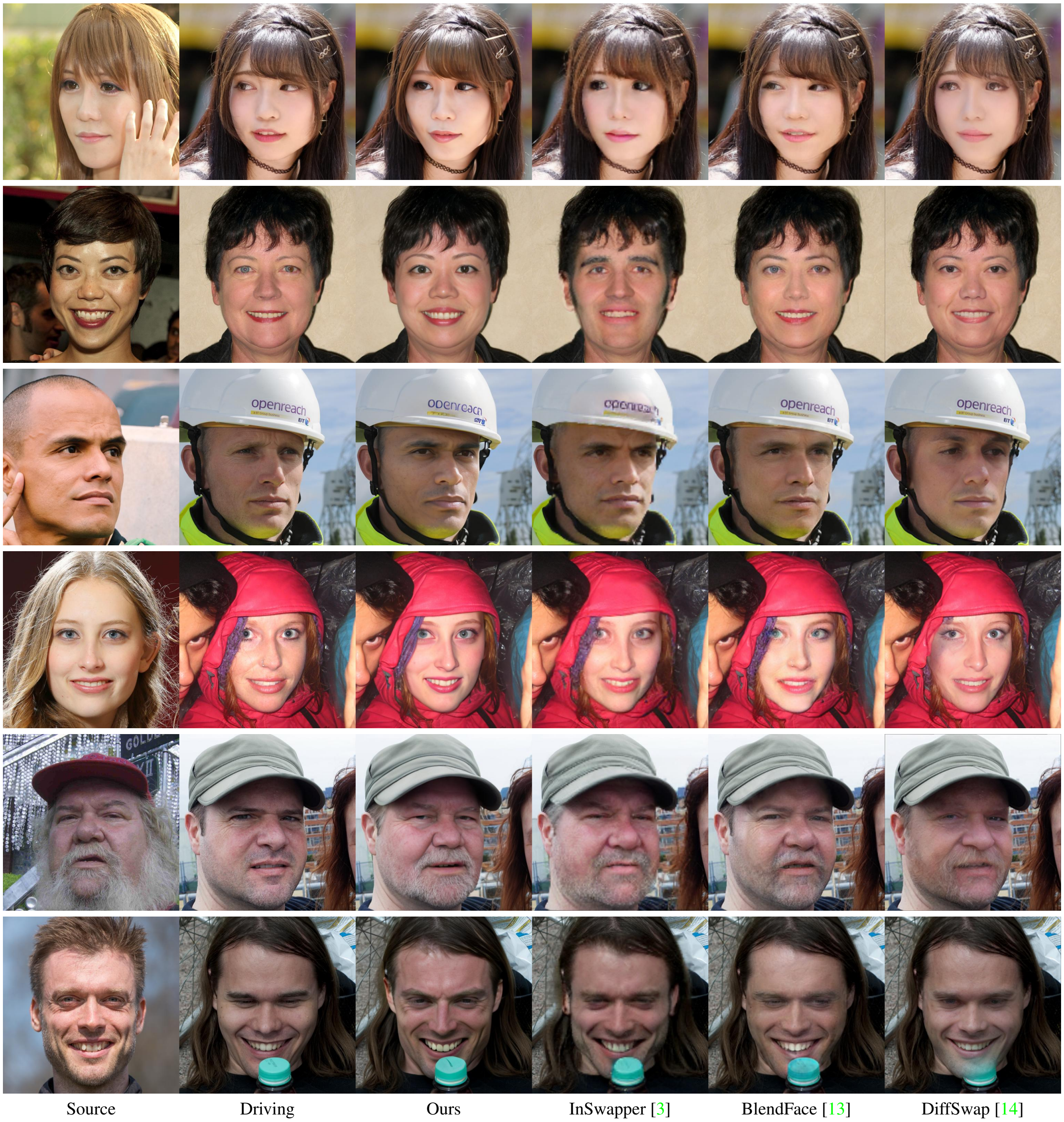


Figure 12. Qualitative results on the task of face swapping for FFHQ [8] test set.



Source Driving Ours InSwapper [3] BlendFace [13] DiffSwap [14]

Figure 13. Qualitative results on the task of face swapping for FFHQ [8] test set.