

# SCOL: Style Code Orchestration in Latent Space for Proactive Face Swapping Defense (Appendix)

In this document, we first present the results of identity-obfuscating fusion, including visual analysis of identity retrieval and evaluation on adversarial dodging (untargeted) attacks in face recognition (Section A). Next, we conduct experiments to determine the optimal hyperparameters, including the weighting factor of the id-lock loss ( $\lambda_{id-lock}$ ), the latent component decision threshold ( $\tau$ ), and attack intensity ( $\epsilon$ ) (Section B). Then, we provide additional qualitative results to further demonstrate the effectiveness of our method (Section C). Finally, we discuss the generalizability, limitations, practical applicability, and future directions (Section D).

## A Effect of Identity Obfuscation

### A.1 Visual Analysis of Identity Retrieval Results

To validate the effectiveness of the proposed identity-obfuscating fusion, we analyze the retrieval results based on images generated from both the original and fused latent codes. Specifically, the latent codes are fed into an optimized generator to produce images, which are subsequently used for identity retrieval.

Fig. 1 presents the top-5 retrievals, where the numbers above each retrieved image indicate identity similarity to the query image. Greens denote images belonging to the same identity as the query, while reds indicate different identities. In the first row, the retrieval results are obtained using the original image as the probe. The high similarity scores and consistent identity labels show that the retrieval process correctly associates the query with its corresponding identity. The second row shows the retrievals when using an image generated from the fused latent code as the probe. Compared to the first row, the retrieved images generally exhibit lower similarity scores and the top-5 matches belong to diverse identities. This indicates that identity information has been dispersed via the fusion process, demonstrating the obfuscation effect. The third row presents the retrieval results for an image generated using the proposed optimized generator, with the generated image used as the probe. While the generated image closely resembles the input image in terms of visual appearance, its retrieval results remain consistent with those of the fused latent code-based image.

These results illustrate that style code-wise fusion effectively obfuscates identity. Furthermore, the proposed generator optimization ensures the generated image aligns closely

with the visual appearance of the original image while preserving the obfuscated identity state.

### A.2 Performance Evaluation of Dodging Attack in Face Recognition

To quantitatively evaluate identity obfuscation, we assessed dodging attack performance in various face recognition models. Expanding on the results provided in the CLIP2Protect [Shamshad *et al.*, 2023], we included our method to compare its effectiveness in the same experimental settings. The evaluation was conducted on the LFW dataset [Huang *et al.*, 2008], using the untargeted identity success rate (Rank-1-U) as the metric. Rank-1-U measures the rate at which the top-1 retrieved candidate does not match the identity of the original image, indicating successful identity obfuscation.

Table 1 summarizes the results in four face recognition models: IRSE50 [Hu *et al.*, 2018], IR152 [Deng *et al.*, 2019], FaceNet [Schroff *et al.*, 2015], and MobileFace [Chen *et al.*, 2018]. Identity-obfuscating fusion in SCOL outperforms all baselines with an average success rate of 90.4%. SCOL demonstrates strong robustness and generalizability in preventing identity extraction across various face recognition models, which are used by face-swapping deepfake generators to create deepfake images.

### A.3 Necessity of the Fused Latent Code

One might question whether the fused latent code is necessary for identity obfuscation, given that the generated image exhibits slight differences from the original input. An alternative approach could involve using a randomly sampled latent code or directly adopting another person’s latent representation. As demonstrated in Fig.2, optimizing a pre-trained StyleGAN[Karras *et al.*, 2020] to completely alter the output using a specific latent code has clear limitations. The fused latent code, resulting from aggregating similar style codes, can successfully represent the input image via generator fine-tuning. However, using a latent code unrelated to the input leads to significant distortion in the image. These results demonstrate that the fused latent code plays a crucial role in balancing identity manipulation and visual fidelity, making it a necessary component of our approach.

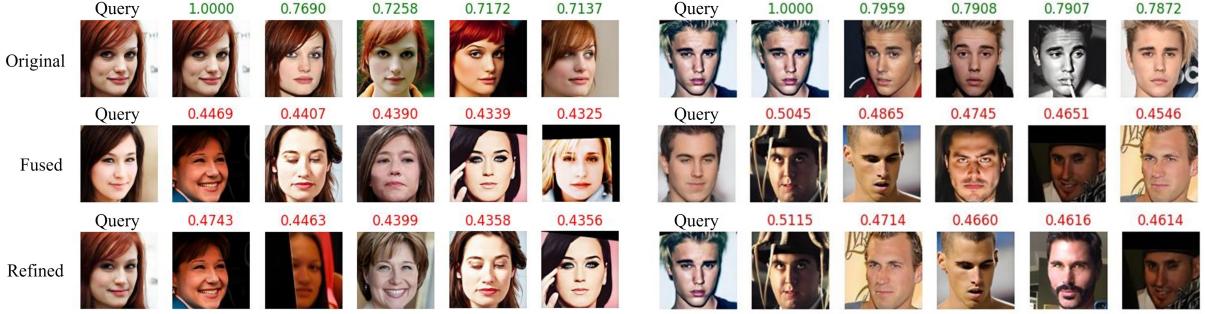


Figure 1: Top-5 identity retrievals for different input images using the VGGFace2-HQ dataset as both the query and gallery. Green indicates matching identities, red denotes mismatches, and the numbers show identity similarity for each output.

Face recognition models		IRSE50	IR152	FaceNet	MobileFace	Average
Method	MI-FGSM [Dong <i>et al.</i> , 2018]	70.2	58.4	59.2	68.0	63.9
	TI-DIM [Dong <i>et al.</i> , 2019]	79.0	67.4	74.4	79.2	75.0
	TIP-IM [Yang <i>et al.</i> , 2021]	81.4	71.8	76.0	82.2	77.8
	CLIP2Protect [Shamshad <i>et al.</i> , 2023]	86.6	73.4	83.8	85.0	82.2
SCOL (w/o Identity inversion attack)		<b>93.8</b>	<b>85.6</b>	<b>91.4</b>	<b>90.7</b>	<b>90.4</b>

Table 1: Protection success rate of black-box dodging attacks under the face recognition on LFW dataset.

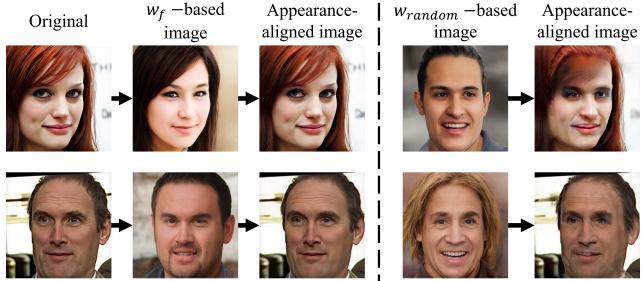


Figure 2: Comparison of optimized generator outputs using the fused latent code and a random sampled latent code.

## B Hyperparameter Tuning

### B.1 ID-Lock Loss

Table 2 shows the identity score matching (ISM) [Van Le *et al.*, 2023] and PSNR results for images generated with varying  $\lambda_{id-loss}$  values. As  $\lambda_{id-loss}$  increases, ISM decreases, and PSNR also drops. Although  $\lambda_{id-loss} = 0.5$  yields a higher PSNR and a lower ISM of 0.509, it was not selected due to the tendency of the top-1 retrieval result to return the same identity. Therefore, the balance between identity obfuscation and image quality led to the selection of  $\lambda_{id-loss} = 1$ . At  $\lambda_{id-loss} = 1$ , ISM reduces significantly while PSNR is 32.01, corresponding to an average pixel difference of 6 pixels in an 8-bit image, which is difficult to distinguish by the human eye.

### B.2 Latent Component Decision Threshold ( $\tau$ )

Table 3 presents the results of an experiment aimed at determining the optimal value of the latent component decision threshold ( $\tau$ ). When  $\tau$  is set too low, a large portion of the la-

$\lambda_{id-loss}$	0	0.1	0.5	1	5
	Source metric	ISM $\downarrow$	PSNR $\uparrow$	ISM $\downarrow$	PSNR $\uparrow$
0.000	1.000	0.744	0.509	0.158	0.151
0.5045	0.5045	35.70	35.26	33.25	32.01

Table 2: Experimental results evaluating the effect of different values of  $\lambda_{id-loss}$  on source metrics on the VGGFace2-HQ dataset.

tent code is replaced by the original input, resulting in a significant decrease in the performance of identity obfuscation and a smaller error for identity inversion attacks. As a result, both attack efficiency and identity obfuscation are compromised. In contrast, as  $\tau$  increases, more components in the latent code are considered identity related, reducing the number of components replaced by the original input. This leads to a slight decrease in PSNR, but the effectiveness of identity inversion attacks increases. However, if  $\tau$  is set too high, unnecessary errors such as identity-irrelevant difference are included for adversarial attacks, making the attack inefficient. Therefore,  $\tau = 3$  is chosen as the optimal value, as it results in the highest defense success rate (DSR) [Qu *et al.*, 2024] and the lowest PSNR from the perspective of deepfake metrics. This value is particularly effective because it focuses on the latent components closely tied to identity, allowing for the calculation of adversarial error based on these components alone. By targeting these specific latent features, the attack becomes more efficient, requiring less perturbation to achieve a more impactful result. This leads to a more effective attack with minimal modification.

### B.3 Adversarial Attack Intensity ( $\epsilon$ )

Table 4 presents the quality metrics for both the protected sources and their corresponding deepfake results under varying adversarial attack intensities ( $\epsilon$ ). As  $\epsilon$  increases, the difference between the deepfake results generated from the original image and the protected image tends to grow. However,

	$\tau$	1	2	3	4	5
Deepfake metric	DSR $\uparrow$	63.5	81.5	95.5	94.0	91.5
	PSNR $\downarrow$	26.54	21.89	20.68	20.94	21.35

Table 3: Experimental results evaluating the effect of different values of the latent component decision threshold ( $\tau$ ) on deepfake metrics, using the VGGFace2-HQ dataset and FaceSwapper face-swapping deepfake generator.

	$\epsilon$	4	8	16
Source metric-PSNR( $\uparrow$ )		33.35	32.01	29.89
Deepfake metric-PSNR( $\downarrow$ )		21.96	20.68	20.60

Table 4: Comparison of quality metrics for protected sources and deepfake results at varying  $\epsilon$  values, using the VGGFace2-HQ dataset and the FaceSwapper face-swapping deepfake generator.

this improvement comes at the expense of source image quality. When  $\epsilon=4$  is applied, the source image quality remains high, but the defense performance is less effective. Given that the deepfake PSNR values for  $\epsilon=8$  and 16 are nearly identical,  $\epsilon=8$  is selected as the optimal choice, as it provides strong protection against deepfake generation while preserving visual fidelity in the protected source.

## C Additional Qualitative Results

### C.1 Effect of Identity Inversion Attack

As shown in Fig. 3, identity obfuscation alone can significantly alter deepfake results. However, it may still retain some recognizable facial attributes, such as gender or eye-brow shape. This raises concerns about the completeness of the defense. To address this, we apply the identity inversion attack, which forces a complete transformation of facial features. This ensures that the generated deepfake presents a completely different and unrecognizable appearance. The additional step removes any lingering traces of the original identity, providing a more robust defense.

### C.2 Proactive Defense Application in The-Wild Images

To further assess the practical effectiveness of the proposed defense model, SCOL was applied to images with challenging real-world conditions, as shown in Fig. 4. These images, sourced from the VGGFace2-HQ dataset, include occlusion, varying lighting, and dynamic poses. Deepfake face-swapping was performed using FaceSwapper [Li *et al.*, 2024]. To ensure the effective application of SCOL, the pre-processing step requires only facial alignment, typically performed using a module such as MTCNN [Zhang *et al.*, 2016]. Once the faces are properly aligned, SCOL can be successfully applied. The results demonstrate SCOL to effectively counteracts face-swapping deepfakes in these complex scenarios.

### C.3 Visual Comparison of Proactive Defenses Across Different Deepfake Generators

Figures 5-6 show the results of various proactive deepfake defense models, evaluated on the VGGFace2-HQ dataset [Chen *et al.*, 2023], alongside the outcomes of different deepfake

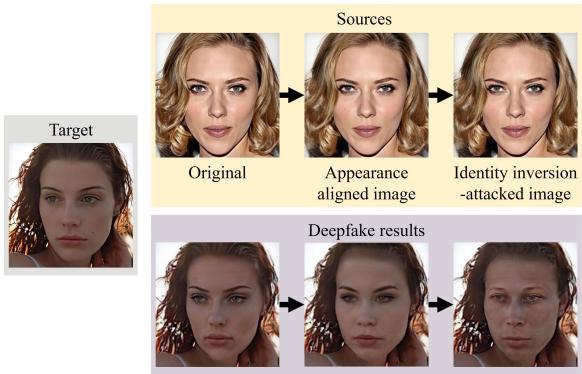


Figure 3: Effect of the identity inversion attack, showing how the resulting deepfake varies significantly despite no visible change to the source image.

models. Figures 7-9 present the experimental results conducted using the CelebA-HQ dataset [Karras, 2017]. In these experiments, the defense models TIP-IM [Yang *et al.*, 2021], CMUA-Watermark [Huang *et al.*, 2022], Anti-Forgery [Wang *et al.*, 2022], CLIP2Protect [Shamshad *et al.*, 2023], DiffAM [Sun *et al.*, 2024], and DF-RAP [Qu *et al.*, 2024] were used as baseline comparisons, while SimSwap [Chen *et al.*, 2020], FaceDancer [Rosberg *et al.*, 2023], BlendFace [Shiohara *et al.*, 2023], and FaceSwapper [Li *et al.*, 2024] served as the deepfake models. The proposed model, SCOL, effectively defends against all considered deepfake models. The results demonstrate its generalizability in defending against face-swapping deepfake methods.

## D Discussion

### D.1 Generalizability

Unlike traditional methods, SCOL does not rely on pre-trained face recognition models. Instead, it uses style code-wise fusion in the latent space to obfuscate identity, making it robust against various face-swapping deepfake models. This approach enhances the generalizability of SCOL, allowing it to protect identity effectively across diverse face-sapping deepfake methods.

### D.2 Complexity

In this section, we discuss the computational complexity of the proposed SCOL defense method in comparison to existing proactive defense models such as CLIP2Protect, DiffAM, and DF-RAP. Since SCOL, CLIP2Protect, and DiffAM all employ optimization-based approaches, they share relatively high computational demands, particularly in terms of FLOPs and inference time.

Proactive defense method	FLOPs (G) $\downarrow$	Param (M) $\downarrow$	Inference time (s) $\downarrow$
CLIP2Protect	18756	423	70
DiffAM	1120	227	12
DF-RAP	332	201	4
SCOL	15229	297	47

Table 5: Complexity of deepfake proactive defense methods on the VGGFace2-HQ dataset.

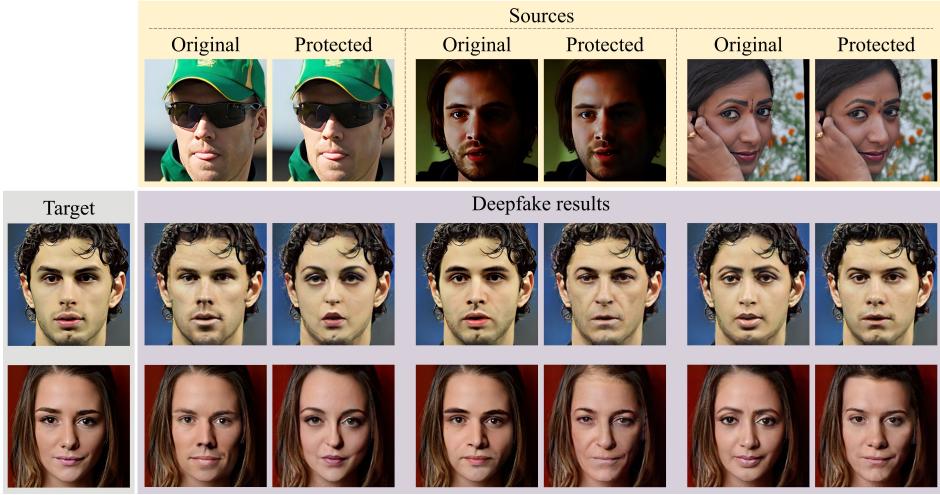


Figure 4: Application of SCOL to images with challenging conditions, such as occlusion, dynamic poses, and varying lighting, using the VGGFace2-HQ dataset and the FaceSwapper face-swapping deepfake generator. SCOL effectively counteracts face-swapping deepfakes in these complex scenarios.

However, SCOL benefits from a simpler architecture compared to CLIP2Protect. The CLIP2Protect consists of generator tuning and adversarial optimization processes, they also incorporate heavy computations, including the use of CLIP model and various facial recognition models. SCOL leverages a combination of GAN inversion and StyleGAN, resulting in fewer parameters and a more streamlined process. In the case of DiffAM, adversarial makeup transfer requires fine-tuning a pre-trained diffusion model, which entails an additional 19 minutes of retraining.

Despite this, SCOL’s inference time remains relatively high due to the optimization-based fine-tuning performed at inference time, where the generator is updated, adding to the computational load. In comparison, methods like CLIP2Protect and DiffAM, which also rely on optimization-based techniques, tend to exhibit similar or higher computational complexity, particularly in terms of FLOPs. DF-RAP, based on a pre-trained generative model, injects adversarial perturbations, resulting in lower FLOPs and fewer parameters, with a efficient inference time. However, its defense effectiveness is limited to specific deepfake models. Nevertheless, SCOL’s inference time, while slower than nonoptimization-based methods, is justified by its effectiveness in defending against deepfake attacks. SCOL demonstrates superior defense capabilities, maintaining robust protection against face-swapping deepfake models.

#### Inference time

SCOL consists of three stages and takes less than one minutes on a single Nvidia GeForce RTX 3090. The inversion time takes less than one second, since we use encoder-based gan inversion model [Tov *et al.*, 2021]. The process of finding the most similar and least similar style codes from a prebuilt gallery involves simple dot product operations. The number of operations is  $N \times 512 \times 18$ , where  $N$  is the number of stored images. This process is completed within one second. The generator tuning takes about 45 seconds as it computes

the loss while comparing identities, while latent component inspection and identity inversion attacks take less than one second.

#### D.3 Limitation

The SCOL model employs GAN inversion to extract latent codes from facial images, where accurate facial alignment is crucial for obtaining reliable latent representations. However, face detection may fail, resulting in a reduction in the accuracy of style code extraction. This, in turn, undermines SCOL’s effectiveness, potentially reducing its robustness as a defense against deepfake face swapping. Furthermore, the GAN inversion model operates on  $256 \times 256$  resolution inputs, requiring all images to be resized accordingly. For lower-resolution images, upscaling is applied before latent code extraction, which can lead to a loss of fine-grained identity features. Consequently, the precision of style code retrieval and fusion may be compromised, affecting SCOL’s ability to obfuscate identity while preserving visual consistency.

Another limitation of SCOL arises from identity inversion attacks, which rely on gradient-based methods. As noted by Qu *et al.* [2024], pre-defense algorithms that inject adversarial perturbations are sensitive to lossy compression, such as JPEG compression commonly used in online social networks. As compression artifacts can alter pixel-level perturbations, potentially neutralizing the effects of adversarial modifications.

#### D.4 Applicability

The SCOL framework encompasses a computationally intensive pipeline, including style code retrieval, fusion, generator-based transformation, and adversarial attack application. Running SCOL directly on end-user devices would be impractical due to its computational demands, making an external online service or a specialized application a more viable

solution. Notably, server-side processing is already a common practice in many photo-editing and video-generation services, effectively alleviating computational loads from user devices.

In this scenario, users can upload their images to an SCOL-enabled service prior to sharing them on social media platforms. This external service automatically applies SCOL, generating a protected image that users can then share on platforms such as Instagram or YouTube. This approach offloads the computational burden from personal devices while providing consistent and scalable deepfake protection. Additionally, integrating SCOL with user-friendly editing workflows, similar to those found in current image or video-editing applications, helps maintain a seamless process and encourages broader adoption.

SCOL relies on accurate face detection and alignment for effective style code extraction. While this dependency might seem like a limitation, it is important to note that deepfake generation models share the same requirement. If SCOL fails to detect and process a face due to occlusions, extreme poses, or poor lighting, the deepfake model will likely encounter similar challenges, reducing the likelihood of successful manipulation.

Ultimately, SCOL’s real-world effectiveness is inherently linked to the same fundamental constraints faced by deepfake models. Since both attackers and defenders depend on similar preprocessing steps, SCOL stands as a viable and practical solution for proactive deepfake mitigation in social media platforms and digital identity protection systems.

## D.5 Future work

To address limitations related to the lossy compression sensitivity, future work will explore adversarial attacks in latent code that make fine-grained adjustments to facial attributes. These attacks aim to preserve defense effectiveness against identity extraction while ensuring robustness to compression. By carefully adjusting the latent code, we can obscure identity while preserving the visual fidelity of the original image. This approach enhances robustness across diverse environments and applications, increasing its resilience to compression artifacts and adversarial attacks.

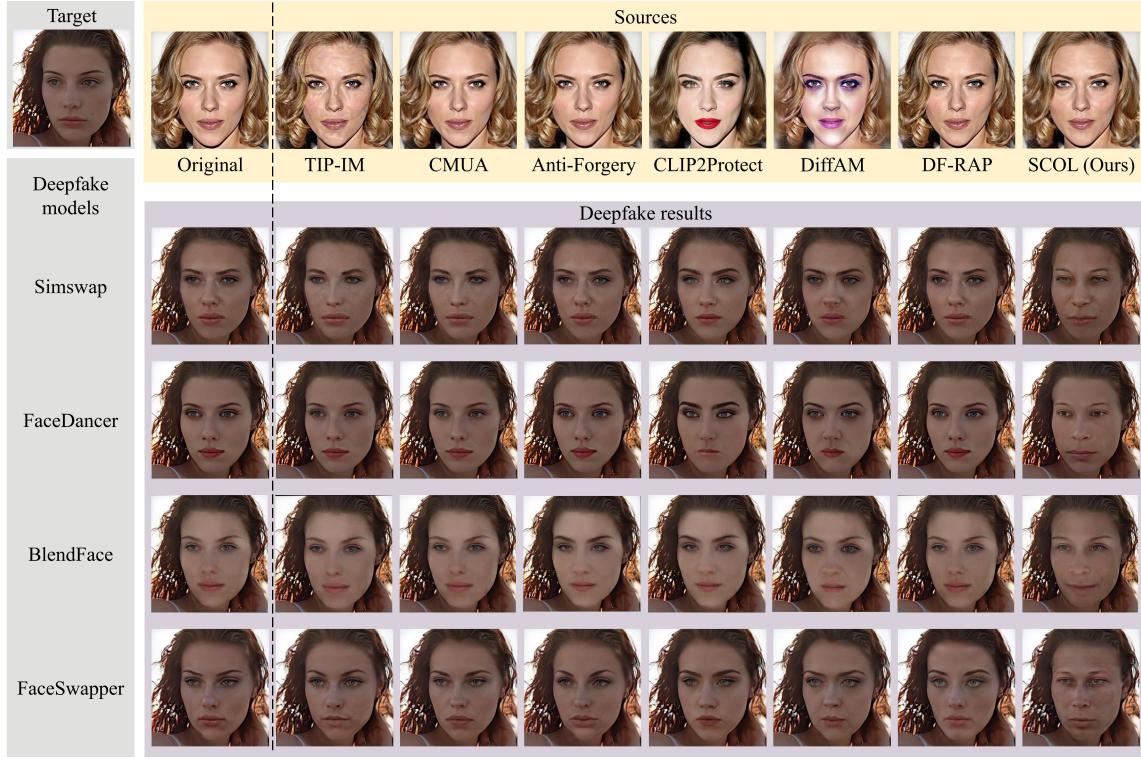


Figure 5: Visualizations of protected female source and target faces on VGGFace2-HQ, generated by various proactive defense models. Alongside, face-swapping results from four deepfake models, demonstrating the generalization ability of our model.

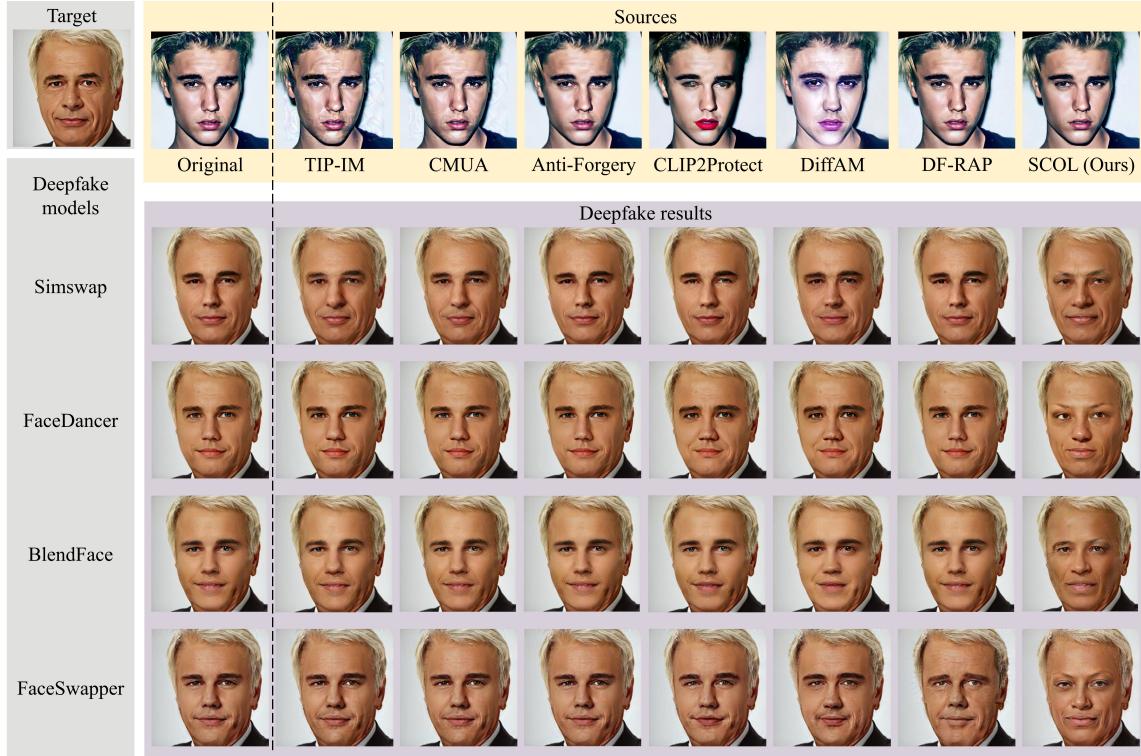


Figure 6: Visualizations of protected male source and target faces on VGGFace2-HQ, generated by various proactive defense models. Alongside, face-swapping results from four deepfake models, demonstrating the generalization ability of our model across different genders.

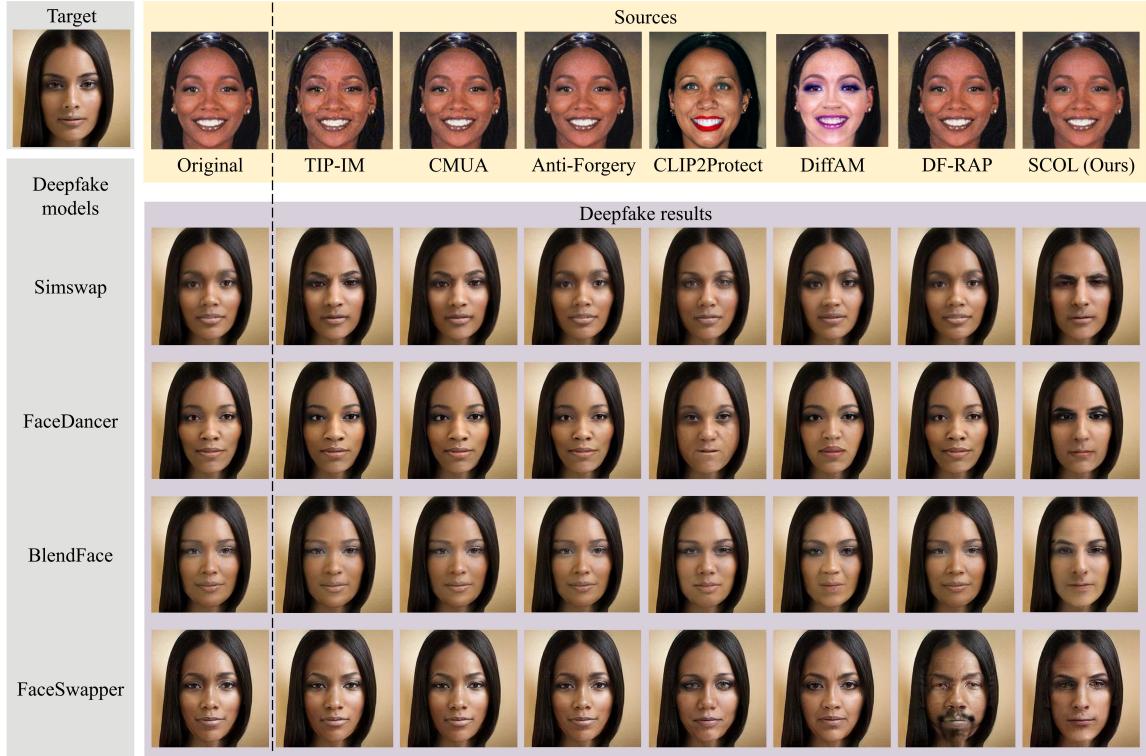


Figure 7: Visualizations of protected source faces on CelebA-HQ, generated by various proactive defense models. Alongside, face-swapping results from four deepfake models, showcasing the robustness of our model to diverse datasets.

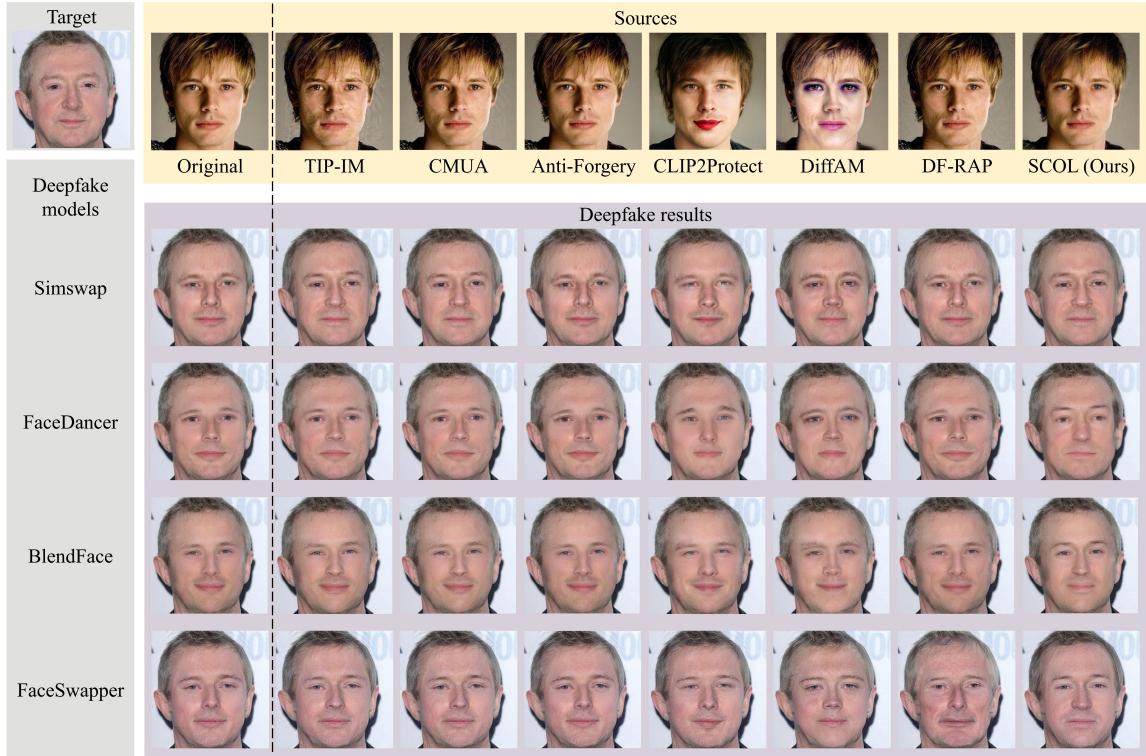


Figure 8: Visualizations of protected source and target faces on CelebA-HQ, where the deepfake results are relatively less effective. Despite this, our model still introduces noticeable changes, demonstrating its robustness even under challenging conditions.



Figure 9: Visualizations of protected male source and target faces on CelebA-HQ, specifically focusing on East Asian individuals. This demonstrates the generalization ability of our model across different ethnicities.

## References

- [Chen *et al.*, 2018] Sheng Chen, Yang Liu, Xiang Gao, and Zhen Han. Mobilefacenets: Efficient cnns for accurate real-time face verification on mobile devices. In *Biometric Recognition: 13th Chinese Conference, CCBR 2018, Urumqi, China, August 11-12, 2018, Proceedings 13*, pages 428–438. Springer, 2018.
- [Chen *et al.*, 2020] Renwang Chen, Xuanhong Chen, Bingbing Ni, and Yanhao Ge. Simswap: An efficient framework for high fidelity face swapping. In *Proceedings of the 28th ACM international conference on multimedia*, pages 2003–2011, 2020.
- [Chen *et al.*, 2023] Xuanhong Chen, Bingbing Ni, Yutian Liu, Naiyuan Liu, Zhilin Zeng, and Hang Wang. Simswap++: Towards faster and high-quality identity swapping. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2023.
- [Deng *et al.*, 2019] Jiankang Deng, Jia Guo, Niannan Xue, and Stefanos Zafeiriou. Arcface: Additive angular margin loss for deep face recognition. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 4690–4699, 2019.
- [Dong *et al.*, 2018] Yinpeng Dong, Fangzhou Liao, Tianyu Pang, Hang Su, Jun Zhu, Xiaolin Hu, and Jianguo Li. Boosting adversarial attacks with momentum. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 9185–9193, 2018.
- [Dong *et al.*, 2019] Yinpeng Dong, Tianyu Pang, Hang Su, and Jun Zhu. Evading defenses to transferable adversarial examples by translation-invariant attacks. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 4312–4321, 2019.
- [Hu *et al.*, 2018] Jie Hu, Li Shen, and Gang Sun. Squeeze-and-excitation networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 7132–7141, 2018.
- [Huang *et al.*, 2008] Gary B Huang, Marwan Mattar, Tamara Berg, and Eric Learned-Miller. Labeled faces in the wild: A database for studying face recognition in unconstrained environments. In *Workshop on faces in 'Real-Life' Images: detection, alignment, and recognition*, 2008.
- [Huang *et al.*, 2022] Hao Huang, Yongtao Wang, Zhaoyu Chen, Yuze Zhang, Yuheng Li, Zhi Tang, Wei Chu, Jingdong Chen, Weisi Lin, and Kai-Kuang Ma. Cmuawatermark: A cross-model universal adversarial watermark for combating deepfakes. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 36, pages 989–997, 2022.
- [Karras *et al.*, 2020] 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*, pages 8110–8119, 2020.
- [Karras, 2017] Tero Karras. Progressive growing of gans for improved quality, stability, and variation. *arXiv preprint arXiv:1710.10196*, 2017.
- [Li *et al.*, 2024] Qi Li, Weining Wang, Chengzhong Xu, Zhenan Sun, and Ming-Hsuan Yang. Learning disentangled representation for one-shot progressive face swapping. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2024.
- [Qu *et al.*, 2024] Zuomin Qu, Zuping Xi, Wei Lu, Xiangyang Luo, Qian Wang, and Bin Li. Df-rap: A robust adversarial perturbation for defending against deepfakes in real-world social network scenarios. *IEEE Transactions on Information Forensics and Security*, 2024.
- [Rosberg *et al.*, 2023] Felix Rosberg, Eren Erdal Aksoy, Fernando Alonso-Fernandez, and Cristofer Englund. Facedancer: Pose-and occlusion-aware high fidelity face swapping. In *Proceedings of the IEEE/CVF winter conference on applications of computer vision*, pages 3454–3463, 2023.
- [Schroff *et al.*, 2015] 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.
- [Shamshad *et al.*, 2023] Fahad Shamshad, Muzammal Naseer, and Karthik Nandakumar. Clip2protect: Protecting facial privacy using text-guided makeup via adversarial latent search. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 20595–20605, 2023.
- [Shiohara *et al.*, 2023] 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.
- [Sun *et al.*, 2024] Yuhao Sun, Lingyun Yu, Hongtao Xie, Jiaming Li, and Yongdong Zhang. Diffam: Diffusion-based adversarial makeup transfer for facial privacy protection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 24584–24594, 2024.
- [Tov *et al.*, 2021] Omer Tov, Yuval Alaluf, Yotam Nitzan, Or Patashnik, and Daniel Cohen-Or. Designing an encoder for stylegan image manipulation. *ACM Transactions on Graphics (TOG)*, 40(4):1–14, 2021.
- [Van Le *et al.*, 2023] Thanh Van Le, Hao Phung, Thuan Hoang Nguyen, Quan Dao, Ngoc N Tran, and Anh Tran. Anti-dreambooth: Protecting users from personalized text-to-image synthesis. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 2116–2127, 2023.
- [Wang *et al.*, 2022] Run Wang, Ziheng Huang, Zhikai Chen, Li Liu, Jing Chen, and Lina Wang. Anti-forgery: Towards a stealthy and robust deepfake disruption attack via adversarial perceptual-aware perturbations. In *Proceedings of the Thirty-First International Joint Conference on Artificial Intelligence, IJCAI-22*, pages 761–767, 2022.
- [Yang *et al.*, 2021] Xiao Yang, Yinpeng Dong, Tianyu Pang, Hang Su, Jun Zhu, Yuefeng Chen, and Hui Xue. Towards

- face encryption by generating adversarial identity masks. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 3897–3907, 2021.
- [Zhang *et al.*, 2016] K. Zhang, Z. Zhang, Z. Li, and Y. Qiao. Joint face detection and alignment using multitask cascaded convolutional networks. *IEEE Signal Processing Letters*, 23(10):1499–1503, 2016.