Ano-Skin: Clinical Feature-Aware Diffusion Model for Dermatological Image Anonymization

Yeon Gyu Han[1,2] Jung Im Na[3] Seong Hwan Kim[4] Dongheon Lee[1,5,6 *]
[1] MODULABS [2] Department of Biomedical Engineering, Chungnam National University
[3] Department of Dermatology, Seoul National University Bundang Hospital
[4] Department of Plastic and Reconstructive Surgery, Kangnam Sacred Hospital
[5] Department of Radiology, Seoul National University College of Medicine
[6] Department of Radiology, Seoul National University Hospital

dusrb37@gmail.com vividna@gmail.com kalosmanus@naver.com
dhlee.jubilee@gmail.com

Abstract

*Medical image anonymization requires effectively balancing patient privacy with clinical feature preservation, yet existing methods either compromise privacy or obscure critical disease information. We propose "**Ano-Skin,**" a framework based on Stable Diffusion-v2 Inpainting with three key contributions: (1) Focused Feature Enhancement loss, $\mathcal{L}_{FFE}$ for preserving disease-specific characteristics, (2) Disease Difference loss, $\mathcal{L}_{DIFF}$ to maintain distinct visual patterns between skin conditions, and (3) Simple Preference Optimization (SimPO) for seamless integration between preserved pathology and anonymized regions. Our method enables flexible control through mask and text-based prompting while maintaining high clinical utility across diverse skin conditions. Evaluated on 6,000 dermatological images, Ano-Skin significantly outperforms existing methods in disease classification performance (94.7% AUC retention), anonymization success (100%), and clinical assessment (4.5/5 by dermatologists). This work advances medical data sharing by resolving the traditional trade-off between privacy protection and diagnostic value in dermatological imaging.*

1. Introduction

Medical image anonymization is crucial for healthcare data sharing, education, and privacy protection, particularly in dermatology where patient faces and skin lesions often appear in the same frame [1-3]. Despite this importance, current approaches fail to simultaneously preserve clinically valuable disease features while effectively anonymizing patient identity.

Traditional anonymization methods that apply digital masks or blurring to facial regions [4, 5] have significant limitations: they often retain identifying features that compromise privacy and, more critically, completely obscure disease information on facial regions. This results in substantial loss of clinically valuable data, especially for dermatological conditions that primarily manifest on the face.

However, although face generation and editing can be used for anonymization, issues related to AI fairness still remain. Much of the work on face synthesis and editing has relied on public datasets, such as CelebA [6] and FFHQ [7], which include a limited range of Asian faces. Consequently, these models fail to accurately represent Asian facial features, such as generally smaller eyes, among other characteristics.

Recent advances in generative models have enabled the synthesis of skin lesion images [8, 9], but these approaches focus primarily on lesion generation rather than anonymization. They face two critical limitations: First, they fail to accurately capture disease-specific characteristics (such as the scaling patterns of psoriasis versus the erythematous patches of atopic dermatitis). Second, not being inpainting models, they cannot create natural transitions between preserved disease features and anonymized facial regions, resulting in visually inconsistent images that diminish clinical utility [2].

We introduce **Ano-Skin**, a novel framework based on Stable Diffusion v2 Inpainting [10] model that addresses these limitations through a specialized architecture designed for dermatological image anonymization. Our technical approach incorporates three primary mechanisms: (1) Focused Feature Enhancement (FFE) loss, $\mathcal{L}_{FFE}$, that operates through critical feature masking to identify and enhance disease-specific details, (2) Disease Difference (DIFF) loss, $\mathcal{L}_{DIFF}$, that maximizes angular separation in the embedding space between different skin conditions, and (3) Simple Preference Optimization [11] that ensures

* Corresponding Author.

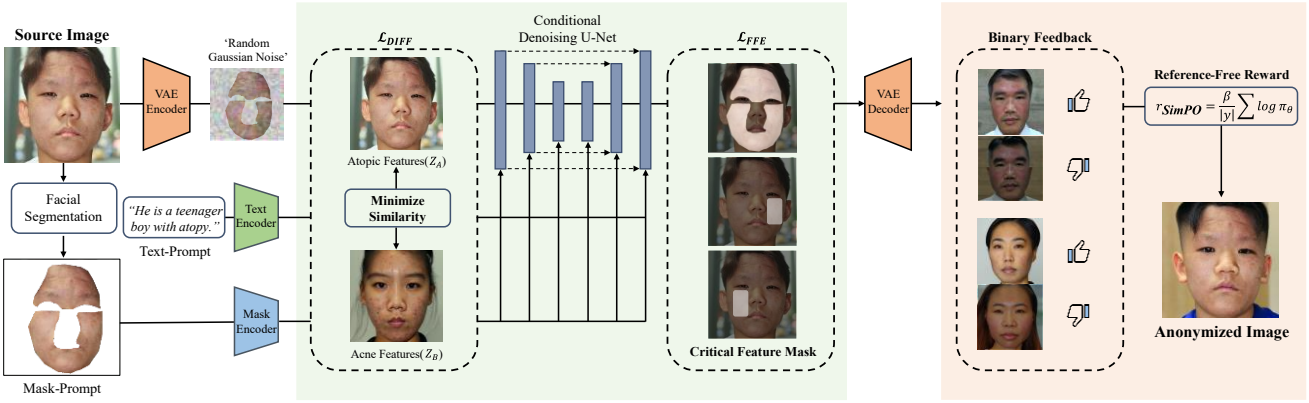

Figure 1: The overview of ***Ano-Skin***. The disease-preserving facial segmentation model generates masks that designate which clinical features to preserve during anonymization. The Stable Diffusion v2 Inpainting model serves as our baseline, fine-tuned on dermatological images. We apply $\mathcal{L}_{FFE}$ to enhance disease-specific details (like the atopic dermatitis features shown) while using $\mathcal{L}_{DIFF}$ to maintain distinct characteristics between different skin conditions (shown as atopic vs acne features). Finally, we employ Simple Preference Optimization (**SimPO**), with LoRA Adapter for efficient fine-tuning, producing clinically valuable anonymized images.

seamless integration between preserved pathology and generated facial features.

We evaluated our model using a comprehensive dataset of facial skin disease images, demonstrating that **Ano-Skin** significantly outperforms existing methods in clinical utility and feature preservation accuracy, as assessed by board-certified dermatologists. Our model maintains the diagnostic value of images while effectively anonymizing patient identity across diverse skin conditions and demographics.

Our contributions are as follows:

- We propose the framework that balances the competing requirements of complete facial anonymization and precise preservation of dermatological features across multiple skin conditions (psoriasis, atopic dermatitis, acne, rosacea, and seborrheic dermatitis).
- We develop novel loss function mechanisms that address the unique challenges of dermatological image processing, enabling both fine-grained feature retention and inter-disease differentiation.

2. Related Work

Medical Face Anonymization. Traditional approaches to medical image anonymization have primarily relied on basic image processing techniques. Pixelation and blurring [12] are widely used in clinical publications but often compromise the clinical value of dermatological images by obscuring disease features alongside identifying characteristics. Studies have shown that these methods still pose privacy risks, as machine learning models can sometimes recover identity information from blurred images [12, 13]. More advanced methods employ digital masking techniques [1] that overlay black rectangles on specific facial features have become standard practice in many medical journals. While these methods preserve some clinical information, they fundamentally cannot maintain disease features in masked regions, creating an inherent trade-off between privacy and clinical utility. Additionally, research has demonstrated that partial masking often fails to prevent re-identification, especially when combined with other available information [13]. Recent advances in deep learning have led to GAN-based anonymization methods [14, 15] that replace real faces with synthetic ones. While promising for general medical imaging, these approaches were not designed for dermatological applications and typically replace the entire facial region, eliminating valuable disease information. They also often produce artifacts at the boundaries between preserved and generated regions [12], particularly problematic in dermatology where boundary characteristics are diagnostically significant.

Skin Disease Synthesis. Early work in skin lesion synthesis utilized statistical modeling and texture synthesis [16], creating simplified representations of common conditions. These methods produced visually plausible results but lacked the detail necessary for clinical applications and were limited to a narrow range of conditions. GAN-based approaches for dermatological image generation [14, 15, 17, 18] have shown promising results for specific conditions like melanoma and common skin lesions. Models such as DermGAN [14] can generate realistic images of isolated lesions but struggle with accurately representing the subtle details that distinguish different skin conditions. Diffusion models [19, 20] have subsequently emerged as a powerful alternative, offering more detailed and diverse generation capabilities compared to GANs. Works such as Derm-T2IM [20] have leveraged diffusion models' probabilistic approach to achieve higher fidelity in texture and color representation of dermatological conditions. Existing generative approaches fail to address the dual requirements of dermatological image anonymization: they cannot

selectively preserve clinically valuable disease-specific characteristics (such as the scaling patterns of psoriasis versus the erythematous patches of atopic dermatitis) while effectively anonymizing patient identity, and lacking inpainting capabilities, they cannot create natural transitions between preserved pathology and anonymized facial regions, resulting in visually inconsistent boundaries that significantly diminish clinical utility for diagnostic and educational purposes.

3. Method

Our framework addresses the critical challenge of dermatological image anonymization through a specialized architecture designed to preserve disease-specific features while protecting patient identity. We build upon Stable Diffusion v2 Inpainting with three key technical innovations: precise disease-preserving segmentation, clinical feature enhancement with dual loss functions, and Simple Preference Optimization for Diffusion model. Figure 1 illustrates our overall approach.

3.1. Disease-Preserving Segmentation

For precise disease feature preservation, we employ PointRend [21] trained on 17,697 annotated facial images from A hospital (mIoU: 0.92). This approach efficiently segments both facial features and lesion boundaries—critical for dermatological diagnosis—creating masks that designate which disease regions to preserve while anonymizing identifiable facial features.

3.2. Feature-Preserving Fine-Tuning

Focused Feature Enhancement (FFE) Loss. We propose the Focused Feature Enhancement (FFE) loss $\mathcal{L}_{FFE}$, to preserve subtle dermatological features often lost during standard fine-tuning:

$$\mathcal{L}_{FFE} = \lambda \frac{1}{N} \sum_{i=1}^{N} M_c^i (O^i - T^i)^2 \quad (1)$$

where M_c is a critical feature mask that highlights regions requiring attention, generated by thresholding the error map between predicted output and target image. This approach specifically preserves disease-specific details like scaling patterns in psoriasis and erythematous patches in dermatitis.

Disease Difference Loss. To maintain distinct visual characteristics between different skin conditions, we implement the Disease Differentiation Loss that maximizes angular separation between disease class embeddings:

$$\mathcal{L}_{DIFF} = \frac{1}{|P|} \sum_{(i,j)\in P} \left(1 - \frac{z_i \cdot z_j}{|z_i||z_j|}\right) \quad (2)$$

This prevents "catastrophic forgetting [22]" of disease-specific features during fine-tuning, ensuring that, for example, psoriasis remains visually distinguishable from atopic dermatitis in the model's embedding space.

Simple Preference Optimization for Diffusion model. To enhance image quality and generate natural-looking anonymized faces, we apply Simple Preference Optimization (SimPO) using reference model-free reward formulation and implement it with Low-Rank Adaptation (LoRA). This approach directly optimizes for image quality and region integration based on dermatologist preferences, without requiring complex reward models or extensive computational resources.

4. Experiments

4.1. Experimental Settings

Baselines. We compare our model with (1) traditional methods (blurring, masking), (2) face anonymization models (DeepPrivacy [23], AnonFaces [24]), and (3) ablated versions of our model.

Datasets. We evaluated our method on 6,000 facial skin disease images (psoriasis, atopic dermatitis, acne, rosacea, seborrheic dermatitis, and normal control) collected from three different hospitals. For disease classification, we used an 8:1:1 split, while Stable Diffusion Inpainting was trained with an 8:2 split. We supplemented training with an Asian facial dataset (4,515 images from AI Hub [25]), FFHQ, and CelebA-HQ.

Evaluation. Diagnostic preservation was quantified through Disease Classification Performance (DCP) using a DenseNet-121 [26] classifier to compare AUC scores before and after anonymization. Anonymization Success Accuracy was measured using the InsightFace [27] face recognition model. Image quality was measured using FID, MS-SSIM, and LPIPS metrics. Clinical validation was conducted by two board-certified dermatologists evaluating disease feature preservation and image naturalness. Clinical Score was assessed each image on a 5-point scale (1: poor - 5: excellent)

4.2. Comparison with the Baselines

Table 1 presents our approach outperforming baselines in both disease preservation, image quality and Anonymization Success Accuracy. Clinical validation confirmed 4.5 of our anonymized images maintained diagnostic equivalence to originals. Figure 2 shows Quantitative results of Dermatological Image anonymization.

5. Conclusion

We presented ***Ano-Skin***, a novel framework for dermatological image anonymization that effectively balances patient privacy and clinical utility. Our approach demonstrates significant improvements in preserving disease-specific features while anonymizing patient identity across diverse skin conditions. This work represents an important step toward ethical medical data sharing without compromising clinical value.

Methods	D. AUC ↑	Ano. Acc ↑	FID ↑	MS-SSIM ↓	LPIPS ↓	Clinical
Blurring	54.8	0.89	157.9	0.417	0.472	0.
Masking	61.2	1.00	183.4	0.361	0.354	1.
DeepPrivacy	72.3	0.98	99.8	0.362	0.338	0.
AnonFaces	68.5	0.97	101.3	0.353	0.278	0.
SD-v2-I	83.2	1.00	134.7	0.324	0.286	2.
DL-I	82.7	1.00	94.48	0.312	0.263	2.
SD-XL-I	85.4	1.00	95.24	0.297	0.255	2.
Ours	**94.7**	**1.00**	**94.91**	**0.295**	**0.244**	**4.**

Table 1: Comprehensive Evaluation of Dermatological Image Anonymization Methods.
D: Disease, Ano: Anonymization, SD-v2-I: Stable Diffusion-v2 Inpainting, DL-I: DreamLike Inpainting, SD-XL-I: Stable Diffusion XL Inpainting

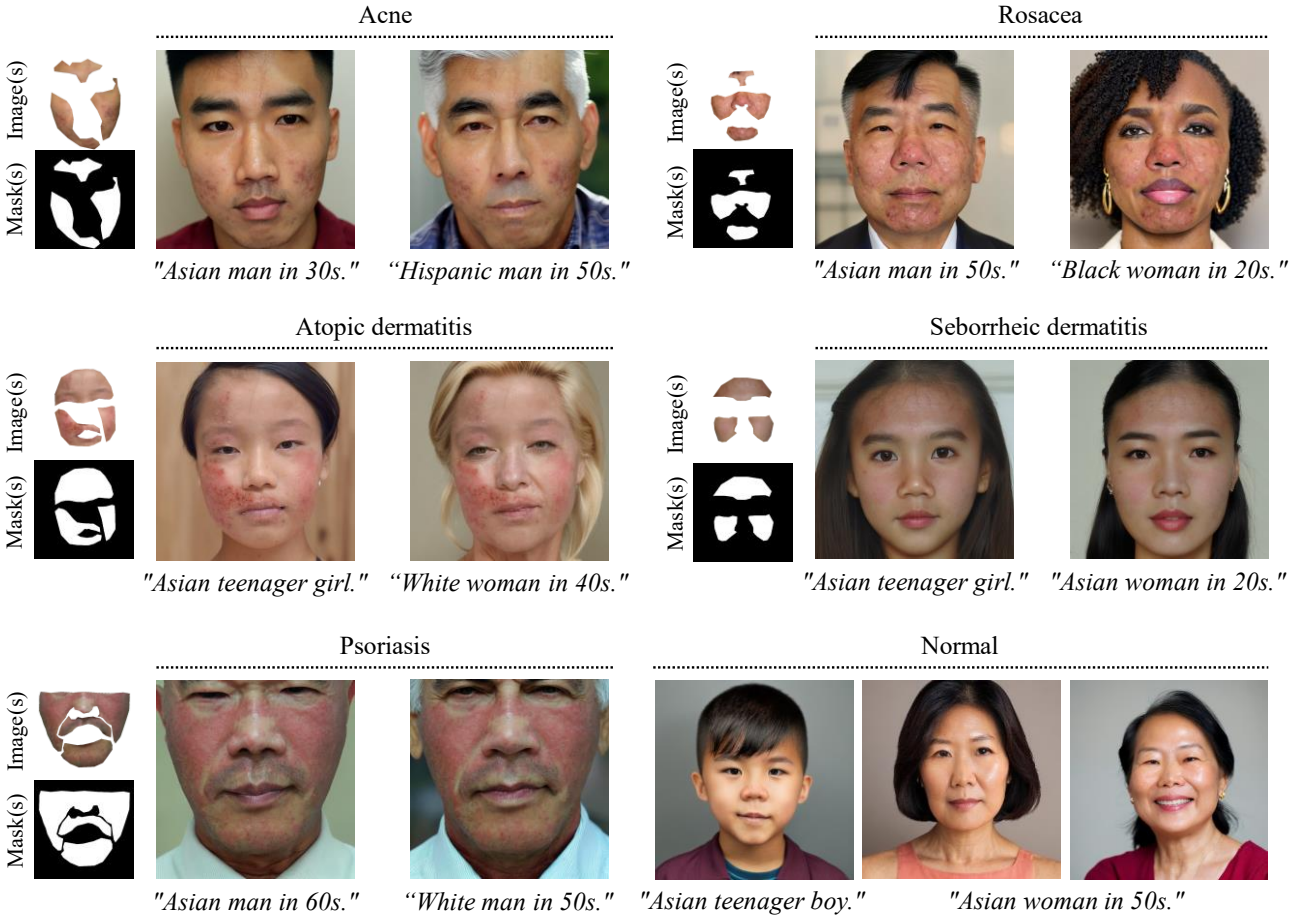

Figure 2: Quantitative results of Dermatological Image anonymization. The results demonstrate effectiveness across different age groups (children, adults, elderly) and genders in both male and female Asian subjects and race groups (Caucasian, Hispanic, Asian, Black). For age-specific conditions like atopic dermatitis and acne that predominantly affect younger populations, our method generates results that appropriately reflect typical age-related manifestations. Through mask-based and text-based prompting strategies, our approach enables precise preservation of disease-specific features while appropriately adjusting facial characteristics to match typical demographic distributions of each condition.

Acknowledgments

This research was supported by Brian Impact Foundation, a non-profit organization dedicated to the advancement of science and technology for all.

References

[1] Y. Yang, S. Irvine, N. Adluru, H. Kang, C. Muralidharan, and V. Singh, "A digital mask to safeguard patient privacy," Nature Medicine, vol. 28, no. 9, pp. 1883–1892, 2022.

[2] M. El Helou, M. Mancini, and R. Timofte, "VerA: Versatile Anonymization Applicable to Clinical Facial Photographs," *arXiv preprint arXiv:2312.09976*, 2023.

[3] M. Rempe, S. Yao, and H. Ling, "De-Identification of Medical Imaging Data: A Comprehensive Tool for Ensuring Patient Privacy," *arXiv preprint arXiv:2410.12402*, 2024.

[4] R. Leyva, B. Meden, P. Peer, and V. Štruc, "Demographic bias effects on face image synthesis," In *Proceedings of CVPR*, 2024.

[5] M. Huber et al., "Bias and diversity in synthetic-based face recognition," In *Proceedings of IEEE/CVF Winter Conference on Applications of Computer Vision*, 2024.

[6] Z. Liu, P. Luo, X. Wang, and X. Tang, "Large-scale celebfaces attributes (celeba) dataset," In *Proceedings of ICCV*, pp. 11–15, 2018.

[7] T. Karras, S. Laine, and T. Aila, "A style-based generator architecture for generative adversarial networks," In *Proceedings of CVPR*, pp. 4401–4410, 2018.

[8] A. Bissoto et al., "Skin lesion synthesis with generative adversarial networks," In *Proceedings of OR 2.0 Workshop, MICCAI*, pp. 294–302, 2018.

[9] X. Yi, E. Walia, and P. Babyn, "Generative adversarial network in medical imaging: A review," Medical Image Analysis, vol. 58, p. 101552, 2019.

[10] R. Rombach et al., "High-resolution image synthesis with latent diffusion models," In *Proceedings of CVPR*, pp. 10684–10695, 2022.

[11] Y. Meng, M. Xia, and D. Chen, "Simpo: Simple preference optimization with a reference-free reward," *arXiv preprint arXiv:2405.14734*, 2024.

[12] K. Lander, V. Bruce, and H. Hill, "Evaluating the effectiveness of pixelation and blurring on masking the identity of familiar faces," Applied Cognitive Psychology, vol. 15, no. 7, pp. 101–116, 2001.

[13] J. Todt, S. Hanisch, and T. Strufe, "Fantômas: Understanding Face Anonymization Reversibility," *arXiv preprint arXiv:2210.10651*, 2022.

[14] A. Ghorbani et al., "Dermgan: Synthetic generation of clinical skin images with pathology," In *Proceedings of Machine Learning for Health Workshop, PMLR*, pp. 155–170, 2020.

[15] S. I. Cho et al., "Generation of a melanoma and nevus data set from unstandardized clinical photographs on the internet," JAMA Dermatology, vol. 159, no. 11, pp. 1223–1231, 2023.

[16] R. S. Barros and M. Walter, "Synthesis of human skin pigmentation disorders," Computer Graphics Forum, vol. 36, no. 1, pp. 330–344, 2017.

[17] C. Baur, S. Albarqouni, and N. Navab, "MelanoGANs: High resolution skin lesion synthesis with GANs," *arXiv preprint arXiv:1804.04338*, 2018.

[18] A. Bissoto et al., "Skin lesion synthesis with generative adversarial networks," In *Proceedings of OR 2.0 Workshop, MICCAI*, pp. 294–302, 2018.

[19] R. Zhang et al., "FairSkin: Fair Diffusion for Skin Disease Image Generation," *arXiv preprint arXiv:2410.22551*, 2024.

[20] M. A. Farooq et al., "Derm-t2im: Harnessing synthetic skin lesion data via stable diffusion models for enhanced skin disease classification using vit and cnn," In *Proceedings of IEEE Engineering in Medicine and Biology Conference (EMBC)*, 2024.

[21] A. Kirillov, Y. Wu, K. He, and R. Girshick, "Pointrend: Image segmentation as rendering," In *Proceedings of CVPR*, pp. 9799–9808, 2020.

[22] R. French, "Catastrophic forgetting in connectionist networks," Trends in Cognitive Sciences, vol. 3, no. 4, pp. 128–135, 1999.

[23] H. Hukkelås, R. Mester, and F. Lindseth, "Deepprivacy: A generative adversarial network for face anonymization," In *Proceedings of ISVC*, pp. 565–578, 2019.

[24] M. Le et al., "Anonfaces: Anonymizing faces adjusted to constraints on efficacy and security," In *Proceedings of WPES*, pp. 87–100, 2020.

[25] AI Hub, "AI Hub Website," https://www.aihub.or.kr/, Accessed: 2022-12-05.

[26] G. Huang et al., "Densely connected convolutional networks," In Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, pp. 4700–4708, 2017.

[27] InsightFace Team, "InsightFace: 2D and 3D Face Analysis Project," https://insightface.ai/, Accessed: 2023-09-03.