

A Knowledge Distillation Approach for Diffusion-based Generative Skin Lesion Segmentation

Viet-Hoang Doan^{1,2}, Thanh-Danh Nguyen^{†1,2}, Tuan-Kiet Ngo^{1,2}, and Vinh-Tiep Nguyen^{1,2}

¹University of Information Technology, Ho Chi Minh City, Vietnam

²Vietnam National University, Ho Chi Minh City, Vietnam

23520515@gm.uit.edu.vn, {danhnt, kietnt, tiepnt}@uit.edu.vn, [†]corresponding author

Abstract—Accurate skin lesion segmentation is crucial for early melanoma detection and reliable medical diagnosis. Recently, such generative diffusion-based segmentation models have achieved state-of-the-art performance in this task. However, their large number of parameters hinders the deployment in real-world clinical and mobile settings. A key challenge is how to compress these models while preserving their reliability in diagnosis. To this end, we propose an edge-focused knowledge distillation (EFKD) strategy for the generative diffusion-based skin lesion segmentation, prioritizing contour information transfer from teacher to student. Accordingly, we introduce LiteDermoSegDiff, a lightweight diffusion-based lesion segmentation model that integrates knowledge distillation with boundary-aware supervision. This design ensures that the student network maintains sharp lesion boundaries while significantly reducing model parameters. Extensive experiments on the common ISIC2018 and HAM10000 dermoscopic benchmarks demonstrate that our method achieves up to 88.9% parameter reduction while maintaining competitive segmentation accuracy. These results show that diffusion-based segmentation models can be substantially compressed without compromising boundary awareness, paving the way for efficient and clinically deployable solutions in resource-constrained environments. Code can be found at <https://github.com/hoangdv05/EFKD>.

Index Terms—Medical image segmentation, Skin lesion segmentation, Diffusion-based model, Knowledge distillation

I. INTRODUCTION

Early detection of skin lesions is crucial for improving patient outcomes, as timely treatment significantly increases survival rates, particularly for aggressive cases such as melanoma [1]. However, delayed or inaccurate diagnosis often leads to reduced effectiveness of therapies. Manual examination and delineation performed by dermatologists is labor-intensive, time-consuming, and subject to inter-observer variability. With the prevalence of skin conditions steadily rising worldwide, there is an urgent demand for automated, reliable, and scalable diagnostic tools that can support clinical decision-making.

Automated skin lesion segmentation plays a central role in this direction, as accurate boundary delineation is essential for downstream diagnostic tasks, such as measuring asymmetry, border irregularity, and lesion size [2]. Traditional CNN-based architectures like U-Net [3]

and DeepLabv3+ [4], as well as vision transformer-based models [5], [6], have demonstrated strong performance on dermoscopic datasets [7], [8]. Nevertheless, these models often struggle to capture subtle lesion boundaries, leading to reduced diagnostic reliability [9], [10]. Furthermore, their high computational cost hinders deployment in resource-constrained healthcare scenarios such as mobile or edge devices [11].

Recently, Denoising Diffusion Probabilistic Models (DDPMs) [12] have emerged as powerful generative models [13], [14]. By progressively refining noisy inputs through iterative denoising, DDPMs can model complex distributions and produce high-quality, diverse outputs. In medical imaging, their uncertainty estimation capability is especially valuable where ambiguous or blurred lesion boundaries exist, even among expert annotators [15]. Diffusion-based segmentation methods such as MedSegDiff [16] and its variants [17], [18] have demonstrated impressive fidelity but remain computationally heavy, making real-time clinical application challenging [19].

To address efficiency concerns, Knowledge Distillation (KD) [20], [21] has been explored as a model compression technique. KD enables a compact student model to approximate the performance of a larger teacher by transferring representational knowledge, often yielding significant reductions in parameters. In medical imaging, KD has been applied to domains such as brain tumor segmentation [22] and retinal vessel analysis [23]. However, most existing approaches primarily optimize global accuracy metrics without explicitly preserving boundary fidelity. This limitation is problematic for lesion analysis, as precise edge delineation is a cornerstone of trustworthy diagnostic outcomes [24], [25].

Lately, DermoSegDiff [26] introduced a dual-branch diffusion architecture for skin lesion segmentation, achieving state-of-the-art results by leveraging a boundary-aware loss. While highly effective in localizing lesion boundaries, its large parameter count (81M) severely limits practicality in mobile and telemedicine applications [27], [28]. This raises the challenge of how to compress diffusion-based models while re-

taining their generative strengths and boundary sensitivity—two factors critical for clinically deployable solutions. Notably, DermoSegDiff [26] is not only a segmentation network but also a generative model, where its diffusion-based formulation enables richer data representation and synthesis capabilities beyond conventional discriminative segmentation frameworks.

To overcome these challenges, we make the following contributions:

- 1) **Edge-Focused Knowledge Distillation (EFKD):** we propose a novel distillation strategy for diffusion-based models that prioritizes lesion boundary pixels during teacher–student knowledge transfer, allowing the student to produce well-defined and accurate edges;
- 2) **LiteDermoSegDiff:** As a result, we introduce LiteDermoSegDiff, a compact variant of DermoSegDiff. By inheriting generative strengths from the teacher while leveraging edge-focused distillation, the student achieves sharp lesion delineation with far fewer parameters;
- 3) **Comprehensive evaluation:** experiments on ISIC2018 [29] and HAM10000 [30] confirm that LiteDermoSegDiff achieves comparable segmentation accuracy to DermoSegDiff with up to 88.9% fewer parameters.

The rest of the paper is organized as follows. Section II reviews related work on diffusion models and knowledge distillation for medical imaging. Section III presents the proposed EFKD framework and LiteDermoSegDiff. Section IV details the experimental setup and results, followed by ablation studies in Section V. Finally, Section VI concludes the paper.

II. RELATED WORK

Skin Lesion Segmentation. Research in skin lesion segmentation has made significant progress with the introduction of U-Net [3], which became the standard for medical image segmentation owing to its encoder–decoder design with skip connections. Subsequent extensions, such as U-Net++ [31] and Attention U-Net [32], further enhanced feature representation and boundary delineation. Recent advances have introduced transformer-based approaches for skin lesion analysis. Vision Transformers (ViTs) [5] have been adapted for medical segmentation through architectures like TransUNet [6] and Swin-UNet [9], which leverage self-attention mechanisms to capture global context

Diffusion Models for Medical Imaging. Denoising Diffusion Probabilistic Models (DDPMs), introduced by Ho et al. [12], have revolutionized generative modeling by learning to reverse a noise addition process. These models achieve superior image quality compared to GANs while providing stable training dynamics. The

adaptation of diffusion models to medical image segmentation represents a recent paradigm shift, treating segmentation as a conditional generation problem, overcoming the problem of inductive bias in the discriminative problem.

MedSegDiff [16] pioneered the application of diffusion models to medical segmentation, demonstrating competitive performance with traditional CNN approaches while providing enhanced boundary quality. MedSegDiff-V2 [17] further extends this approach by employing a conditional U-Net, enabling better interaction between noise and semantic features. DermoSegDiff [26], the foundation of our work, specifically addresses skin lesion segmentation by incorporating boundary-aware training mechanisms and achieving state-of-the-art results on dermatological datasets. **Knowledge Distillation.** Originally proposed by Hinton et al. [20], this technique enables the transfer of knowledge from large teacher models to compact student networks through soft target guidance. The original framework was developed for classification tasks, using temperature-scaled softmax outputs to transfer “dark knowledge” from teacher to student.

Subsequent research has extended knowledge distillation to semantic segmentation. Structured knowledge distillation [33] treats segmentation as a structured prediction problem, employing feature alignment and adversarial learning. Channel-wise distillation [34] emphasizes transferring information through key feature channels, while attention transfer [35] leverages spatial attention maps. More recently, Boundary-Privileged Knowledge Distillation (BPKD) [36] explicitly separates edge and body knowledge for more effective transfer. In medical imaging, KD has been applied to tasks such as brain tumor segmentation [22] and retinal vessel analysis [23]. However, these works largely adapt general KD frameworks rather than designing approaches tailored to the unique challenges of medical segmentation, such as boundary ambiguity and clinical reliability.

Despite the rapid progress of diffusion models, knowledge distillation for diffusion-based lesion segmentation remains largely unexplored. Progressive distillation [37] has been proposed to reduce sampling steps in diffusion models, but it primarily focuses on inference acceleration rather than model compression. To the best of our knowledge, no prior work has systematically investigated knowledge distillation for diffusion-based medical image segmentation, especially considering the region-specific importance of boundaries in clinical applications.

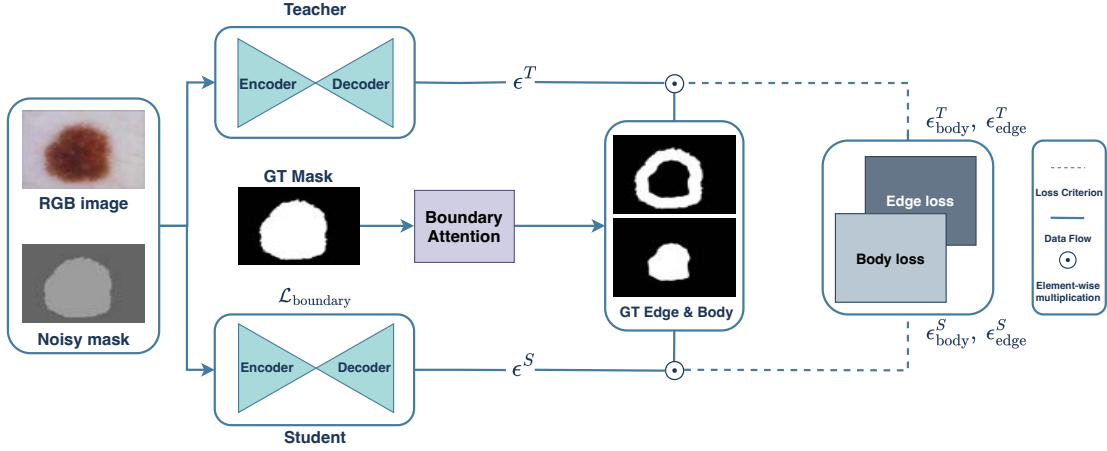


Fig. 1: Overview of the proposed **Edge-Focused Knowledge Distillation (EFKD)** framework. The teacher and student are diffusion-based segmentation networks predicting noise maps ϵ^T and ϵ^S from noisy masks. Ground-truth masks are processed by a boundary attention module to generate edge and body masks, which guide element-wise multiplication (\odot) with teacher and student predictions. The edge and body objective losses are then combined with the boundary-aware loss $\mathcal{L}_{boundary}$ to form the final training criterion.

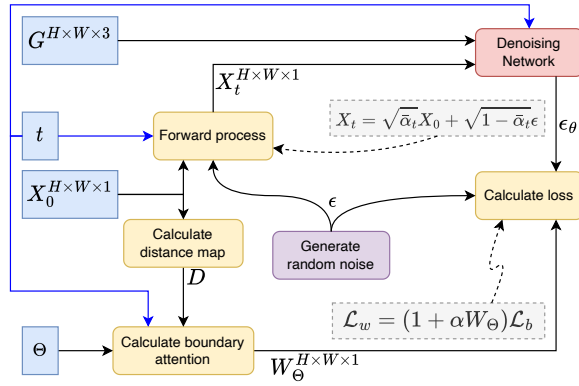


Fig. 2: Overall architecture of DermoSegDiff [26]

III. PROPOSED METHOD

A. Teacher-Student Framework

Our framework follows a teacher-student paradigm where knowledge is transferred from a large, accurate teacher to a lightweight student model (Fig. 1). Both networks are trained to predict the added noise in the forward diffusion process given an input dermoscopic image g and a noisy mask x_t . The teacher provides reliable prediction *signals* that act as learning *guidance* for the student, ensuring that the compact model inherits the structural priors of the full architecture.

B. Teacher: DermoSegDiff

We employ **DermoSegDiff** [26], a diffusion-based generative lesion segmentator with a boundary-aware design, as the teacher model. In this setup, segmentation is posed as a conditional generation problem: starting from a noisy mask, the model progressively denoises it using the dermoscopic image g as conditional input.

A distinctive element of DermoSegDiff [26] is its *boundary-aware mechanism*. Instead of treating all pixels equally, the method introduces a distance-based attention map that places stronger *constraints* on pixels near lesion contours. This boundary-aware loss is given by Eq. 1 below:

$$\mathcal{L}_w = (1 + \alpha W_\Theta) \|\epsilon - \epsilon_\theta(x_t, g, t)\|^2, \quad (1)$$

where W_Θ assigns weights to pixels based on their proximity to lesion borders. This design enhances the model in capturing fine lesion morphology and producing sharper contour predictions.

C. Student Design

The student network \mathcal{S} is obtained by simplifying the U-Net-like architecture of the teacher. Specifically, we reduce the number of stages and restrict the maximum channel width to create a more compact variant. This design is chosen to balance efficiency and representational capacity, making the student lightweight while still capable of capturing essential lesion features.

D. Knowledge Distillation Strategies

1) *General Distillation*: A common approach to knowledge distillation is to enforce a global alignment between teacher and student predictions as in Eq. 2:

$$\mathcal{L}_{KD} = \|\hat{\epsilon}_S - \hat{\epsilon}_T\|^2. \quad (2)$$

This uniform loss distributes the learning progress equally across all pixels. While simple, it often results in blurred contours and a lack of fine-grained detail along lesion boundaries. To mitigate this, prior methods such as BPKD [36] introduce pre- and post-mask filtering that emphasize edges and suppress irrelevant signals. These techniques are effective in multi-class

segmentation, where multiple object categories and their respective boundaries must be modeled. However, in binary lesion segmentation the situation is different: only a single foreground-background contour exists, making class-wise boundary modeling less meaningful. This limitation motivates the need for a boundary-preserving distillation strategy tailored to the binary mask setting.

2) *Edge-Focused Distillation (EFKD)*: To explicitly preserve fine lesion boundaries, we introduce **Edge-Focused Knowledge Distillation (EFKD)**. At each diffusion step, both the teacher \mathcal{T} and the student \mathcal{S} predict the noise, as shown in Eq. 3:

$$\hat{\epsilon}_T = \mathcal{T}(x_t, g, t), \quad \hat{\epsilon}_S = \mathcal{S}(x_t, g, t). \quad (3)$$

Instead of applying the same constraint to the entire lesion area, EFKD separates it into edge and body regions. This separation is done using the distance-aware weight W_Θ from the teacher as in Eq. 1. A threshold r is then applied as the following Eq. 4:

$$M_e(i, j) = \begin{cases} 1, & W_\Theta > r, \\ 0, & \text{otherwise,} \end{cases} \quad M_b(i, j) = 1 - M_e(i, j). \quad (4)$$

Here, M_e marks the edge pixels and M_b marks the body pixels. Smaller thresholds r yield thinner edge bands focused on sharp boundary details, while larger r values expand the edge region to capture more contextual information.

The corresponding objectives are then formulated separately for each region. The edge-focused distillation loss in Eq. 5 is given by:

$$\mathcal{L}_{\text{edge}} = \frac{1}{|M_e|} \sum_{i,j} M_e(i, j) \|\hat{\epsilon}_S^{(i,j)} - \hat{\epsilon}_T^{(i,j)}\|^2, \quad (5)$$

while the body-focused objective in Eq. 6 is:

$$\mathcal{L}_{\text{body}} = \frac{1}{|M_b|} \sum_{i,j} M_b(i, j) \|\hat{\epsilon}_S^{(i,j)} - \hat{\epsilon}_T^{(i,j)}\|^2. \quad (6)$$

Formally, $|M_e|$ and $|M_b|$ denote the number of pixels in the edge and body regions, respectively. Dividing by these factors normalizes the loss, ensuring balanced contributions from both regions.

Finally, the overall EFKD loss combines these two components, as expressed in Eq. 7:

$$\mathcal{L}_{\text{EFKD}} = \lambda_e \mathcal{L}_{\text{edge}} + \lambda_b \mathcal{L}_{\text{body}}, \quad (7)$$

where λ_e and λ_b balance the trade-off between sharp boundary preservation and stable interior learning.

IV. EXPERIMENTAL RESULTS

A. Implementation Details

Our method is implemented in PyTorch (v1.14.0) and trained on a single NVIDIA RTX A4000 GPU.

We adopt a diffusion process with $T = 250$ steps and linearly increasing variances from $\beta_{\text{start}} = 0.0004$ to $\beta_{\text{end}} = 0.08$. Training runs for 2,000 epochs with batch size 16 using Adam optimizer at an initial learning rate of 2×10^{-4} , reduced by a factor of 0.5 when there is no improvement in the last 10 epochs. The edge, body loss weights are set as $\lambda_e = 5$ and $\lambda_b = 2$, following weighting strategy from BPKD [36]. Edge and body regions are separated with a threshold of $\tau = 0.6$. We augment data via Albumentations, including affine transforms, flipping, coarse dropout and additional intensity/geometric perturbations. At inference, nine segmentation masks are sampled per test image, averaged, and thresholded at 0 to produce the final prediction.

B. Datasets

To evaluate the proposed method, we use two publicly available skin lesion segmentation datasets: ISIC2018 [29] and HAM10000 [30]. **ISIC2018** contains 2,594 dermoscopic images with pixel-level annotations of skin lesions. **HAM10000** is a subset of the ISIC archive with 10,015 dermoscopic images and their segmentation masks, split into 7,200/1,800/1,015 for train/val/test. Each image of all datasets is resized to 128×128 pixels.

C. State-of-the-art Performance Comparison

Quantitative Results. Table I reports the performance of LiteDermoSegDiff on the ISIC2018 [29] and HAM10000 [30] datasets, evaluated with four standard segmentation metrics. Across both benchmarks, LiteDermoSegDiff achieves competitive performance while reducing parameters substantially from 81.6M/40.3M in DermoSegDiff to only 9.3M/8.5M. Notably, it secures the second-best Dice Score on both datasets, closely approaching the top-performing teacher model.

Compared to CNN-based methods [3], [4], LiteDermoSegDiff consistently delivers higher Dice Scores while achieving a more balanced trade-off between sensitivity and specificity. Against Transformer-based and hybrid approaches like TransUNet [6] and Swin-UNet [9], our student model also shows clear advantages, with consistent gains in Dice Score and sensitivity without sacrificing overall accuracy. Comprehensive comparisons of GFLOPs, memory consumption, and inference speed (FPS) for both the teacher and student models are available in our GitHub repository

These results confirm that LiteDermoSegDiff successfully transfers the generative capabilities of its teacher while remaining lightweight, making it a practical choice for real-world skin lesion segmentation where both accuracy and efficiency are essential.

Qualitative Results. Qualitative comparisons in Fig. 3 validate the effectiveness of our method. LiteDermoSegDiff produces sharp and coherent lesion

TABLE I: Quantitative comparison on ISIC2018 [29] and HAM10000 [30] datasets. Best, and second best results are in **blue**, and **red**, respectively.

Methods	ISIC 2018 [29]					HAM10000 [30]				
	Params ↓	DSC ↑	SE ↑	SP ↑	ACC ↑	Params ↓	DSC ↑	SE ↑	SP ↑	ACC ↑
U-Net [3]	31.0M	0.8729	0.8739	0.9515	0.9303	31.0M	0.9317	0.9480	0.9728	0.9480
TransUNet [6]	66.8M	0.8610	0.8668	0.9447	0.9223	66.8M	0.8610	0.8668	0.9447	0.9233
Swin-UNet [9]	41.3M	0.8532	0.8148	0.9641	0.9232	41.3M	0.9283	0.9336	0.9723	0.9656
DeepLabv3+ [4]	41.1M	0.8820	0.8560	0.9770	0.9510	41.1M	0.9268	0.9277	0.9768	0.9651
Att-UNet [32]	34.8M	0.8425	0.8524	0.9734	0.9395	34.8M	0.9285	0.9187	0.9784	0.9663
MissFormer [38]	42.4M	0.8697	0.8256	0.9725	0.9320	42.4M	0.9254	0.9271	0.9760	0.9644
DermoSegDiff [26]	81.6M	0.8926	0.8812	0.9647	0.9419	40.3M	0.9487	0.9351	0.9887	0.9759
LiteDermoSegDiff (Ours)	9.3M	0.8826	0.8715	0.9616	0.9365	8.5M	0.9453	0.9506	0.9810	0.9738

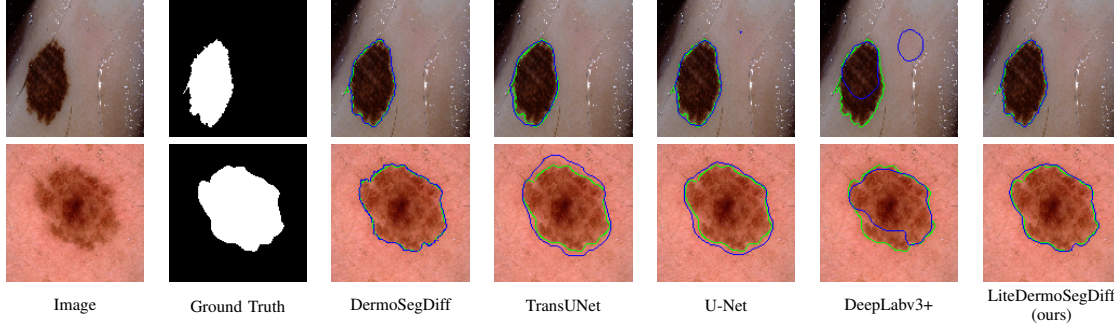


Fig. 3: Qualitative comparison of segmentation results on the ISIC2018 [29] dataset. Ground truth boundaries are shown in **green**, and predicted boundaries in **blue**.

TABLE II: Parameter reduction of DermoSegDiff [26] on ISIC2018 [29] and HAM10000 [30].

Step	Remaining Depth	ISIC2018 Params	HAM10000 Params
Original	6	81.6M	40.3M
Cut Stage 6	5	39.9M	20.3M
Cut Stage 5	4	19.9M	12.6M
Cut Stage 4	3	9.3M	8.5M

boundaries, showing comparable results to its teacher. Despite its compact size, the student model preserves the visual quality of the teacher predictions, confirming its efficiency and practicality for real-world skin lesion segmentation. Compared to other baseline models, LiteDermoSegDiff delineates lesion contours more accurately, yielding clearer and more reliable boundary representations that are essential for clinical diagnosis.

V. ABLATION STUDIES

We conduct ablation experiments to analyze the contributions of parameter reduction and threshold selection in our LiteDermoSegDiff. These studies express how our designs affect both efficiency and accuracy.

Parameter Reduction. We first investigate how cutting the depth of the teacher DermoSegDiff [26] impacts parameter size. As shown in Table II, progressively reducing the depth from six stages to three stages significantly decreases the number of parameters. For ISIC2018 [29], the parameter count drops from 81.6M to 9.3M, while for HAM10000 [30], the reduction goes from 40.3M to 8.5M. The difference in teacher parameter sizes across the two datasets arises from

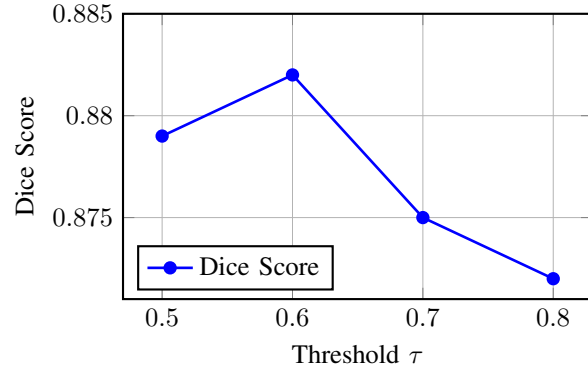


Fig. 4: Ablation study on the boundary-aware threshold values τ on ISIC2018 [29]

the default configuration of DermoSegDiff [26]. Despite this significant decrease, LiteDermoSegDiff retains competitive performance compared to the full teacher model. This demonstrates that our lightweight design maintains essential representational ability while eliminating redundancy, making it suitable for real-world deployment in resource-constrained scenarios.

Threshold Sensitivity. The threshold τ acts as a critical factor in distinguishing edge and body regions, ensuring effective knowledge transfer. A low threshold may include ambiguous pixels and reduce precision, while a high threshold can exclude valid lesion areas and lower sensitivity. As shown in Figure 4, performance peaks at $\tau = 0.6$, which provides a balanced trade-off between lesion coverage and boundary preci-

sion. This setting yields stable results across ISIC2018 [29] and HAM10000 [30].

Summary. Overall, these ablation studies confirm that (i) reducing the network depth can significantly decrease parameter size without sacrificing segmentation quality, and (ii) careful threshold selection is essential for maximizing Dice score, with $\tau = 0.6$ providing the optimal balance between lesion coverage and contour sharpness. Together, these insights validate the efficiency and robustness of LiteDermoSegDiff.

VI. CONCLUSION

In this work, we proposed an edge-focused knowledge distillation (EFKD) framework tailored for diffusion-based skin lesion segmentation. Building on EFKD, we developed LiteDermoSegDiff, a lightweight student model that preserves boundary details while reducing parameters by up to 88.9%. Experiments on ISIC2018 and HAM10000 benchmarks showed that our approach achieves competitive or even superior results compared to larger models, while maintaining sharp lesion boundaries. These results demonstrate that diffusion-based generative segmentation model can be effectively compressed without compromising diagnostic quality. In future work, we plan to extend our framework to other medical imaging tasks beyond skin lesions and explore advanced distillation strategies to further improve lightweight segmentation.

VII. ACKNOWLEDGEMENTS

This research is funded by University of Information Technology - Vietnam National University Ho Chi Minh City under grant number D1-2025-32.

REFERENCES

- [1] R. L. Siegel, K. D. Miller, N. S. Wagle, and A. Jemal, "Cancer statistics, 2023," *CA: A Cancer Journal for Clinicians*, vol. 73, no. 1, pp. 17–48, 2023.
- [2] C. Barata, M. Ruela, M. Francisco, and et al., "Two systems for the detection of melanomas in dermoscopy images using texture and color features," *IEEE Systems Journal*, vol. 8, no. 3, pp. 965–979, 2014.
- [3] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," 2015.
- [4] L.-C. Chen, Y. Zhu, G. Papandreou, F. Schroff, and H. Adam, "Encoder-decoder with atrous separable convolution for semantic image segmentation," 2018.
- [5] A. Dosovitskiy, L. Beyer, A. Kolesnikov, et al., "An image is worth 16x16 words: Transformers for image recognition at scale," 2021.
- [6] J. Chen, Y. Lu, Q. Yu, X. Luo, E. Adeli, Y. Wang, L. Lu, A. L. Yuille, and Y. Zhou, "Transunet: Transformers make strong encoders for medical image segmentation," 2021.
- [7] H. Xu and T. H. Hwang, "Automatic skin lesion segmentation using deep fully convolutional networks," 2018.
- [8] M. A. Al-Masni, M. A. Al-Antari, J. M. Park, and et al., "Skin lesion segmentation in dermoscopy images via deep full resolution convolutional networks," *Computer Methods and Programs in Biomedicine*, vol. 162, pp. 221–231, 2018.
- [9] H. Cao, Y. Wang, J. Chen, D. Jiang, et al., "Swin-unet: Unet-like pure transformer for medical image segmentation," 2021.
- [10] R. Azad, M. Asadi-Aghbolaghi, M. Fathy, and et al., "Attention deeplabv3+: Multi-level context attention mechanism for skin lesion segmentation," in *ECCV Workshops*, 2020.
- [11] M. Sandler, A. Howard, M. Zhu, et al., "Mobilenetv2: Inverted residuals and linear bottlenecks," 2019.
- [12] J. Ho, A. Jain, and P. Abbeel, "Denoising diffusion probabilistic models," 2020.
- [13] A. Ramesh, P. Dhariwal, A. Nichol, et al., "Hierarchical text-conditional image generation with clip latents," 2022.
- [14] R. Rombach, A. Blattmann, D. Lorenz, P. Esser, and B. Ommer, "High-resolution image synthesis with latent diffusion models," in *IEEE/CVF CVPR*, pp. 10674–10685, 2022.
- [15] A. Jungo, F. Balsiger, and M. Reyes, "Analyzing the quality and challenges of uncertainty estimations for brain tumor segmentation," *Frontiers in Neuroscience*, vol. 14, 2020.
- [16] J. Wu, R. Fu, H. Fang, and et al., "Medsegdiff: Medical image segmentation with diffusion probabilistic model," *arXiv preprint arXiv:2211.00611*, 2022.
- [17] J. Wu, H. Fang, Y. Zhang, and et al., "Medsegdiff-v2: Diffusion based medical image segmentation with transformer," *arXiv preprint arXiv:2301.11798*, 2023.
- [18] A. Kazerooni, E. K. Aghdam, et al., "Diffusion models for medical image analysis: A comprehensive survey," 2023.
- [19] W. Peebles and S. Xie, "Scalable diffusion models with transformers," 2023.
- [20] G. Hinton, O. Vinyals, and J. Dean, "Distilling the knowledge in a neural network," 2015.
- [21] A. Romero, N. Ballas, S. E. Kahou, and et al., "Fitnets: Hints for thin deep nets," in *ICLR*, 2015.
- [22] D. Lachinov, E. Shipunova, and V. Turlapov, "Knowledge distillation for brain tumor segmentation," 2020.
- [23] L. Peng, L. Lin, P. Cheng, H. He, and X. Tang, "Student becomes decathlon master in retinal vessel segmentation via dual-teacher multi-target domain adaptation," 2022.
- [24] A. A. Taha and A. Hanbury, "Metrics for evaluating 3d medical image segmentation: analysis, selection, and tool," *BMC Medical Imaging*, vol. 15, no. 1, pp. 1–28, 2015.
- [25] F. Milletari, N. Navab, and S.-A. Ahmadi, "V-net: Fully convolutional neural networks for volumetric medical image segmentation," 2016.
- [26] A. Bozorgpour, Y. Sadegheih, A. Kazerooni, R. Azad, and D. Merhof, "Dermosegdiff: A boundary-aware segmentation diffusion model for skin lesion delineation," 2023.
- [27] A. Esteva, B. Kuprel, R. A. Novoa, and et al., "Dermatologist-level classification of skin cancer with deep neural networks," *Nature*, vol. 542, pp. 115–118, 2017.
- [28] Y. Liu, A. Jain, C. Eng, et al., "A deep learning system for differential diagnosis of skin diseases," *Nature Medicine*, vol. 26, p. 900–908, May 2020.
- [29] N. Codella, V. Rotemberg, et al., "Skin lesion analysis toward melanoma detection 2018: A challenge hosted by the international skin imaging collaboration (isic)," 2019.
- [30] P. Tschandl et al., "The ham10000 dataset, a large collection of multi-source dermatoscopic images of common pigmented skin lesions," *Scientific Data*, vol. 5, Aug. 2018.
- [31] Z. Zhou, M. M. R. Siddiquee, et al., "Unet++: A nested u-net architecture for medical image segmentation," 2018.
- [32] O. Oktay, J. Schlemper, L. L. Folgoc, et al., "Attention u-net: Learning where to look for the pancreas," in *arXiv preprint arXiv:1804.03999*, 2018.
- [33] Y. Liu, K. Chen, C. Liu, Z. Qin, Z. Luo, and J. Wang, "Structured knowledge distillation for semantic segmentation," in *IEEE/CVF CVPR*, June 2019.
- [34] C. Shu, Y. Liu, J. Gao, Z. Yan, and C. Shen, "Channel-wise knowledge distillation for dense prediction," 2021.
- [35] S. Zagoruyko and N. Komodakis, "Paying more attention to attention: Improving the performance of cnns via attention transfer," in *ICLR*, 2017.
- [36] L. Liu, Z. Wang, M. H. Phan, et al., "Bpkd: Boundary privileged knowledge distillation for semantic segmentation," 2023.
- [37] T. Salimans and J. Ho, "Progressive distillation for fast sampling of diffusion models," in *ICLR*, 2022.
- [38] X. Huang, Z. Deng, D. Li, and X. Yuan, "Missformer: An effective medical image segmentation transformer," 2021.