DEEP LEARNING FOR SKIN LESION SEGMENTATION: A REVIEW

Duy Hung BUI[1,*] **and Huy Hoang VU**[2,+]

[1]202412987, Hanoi University of Science and Technology, Hanoi, Vietnam
[2]202412982, Hanoi University of Science and Technology, Hanoi, Vietnam

ABSTRACT

Skin lesion segmentation is a prominent research topic in medical image processing, which could facilitate the early diagnosis of skin diseases. Over the past few decades, the field has witnessed a considerable advancements in Deep Learning architectures, beginning with the dominance of Convolutional Neural Networks (CNNs) such as U-Net to hybrid Vision Transformers (ViT), leveraging Attention mechanisms. This paper provides a comprehensive review of the skin lesion analysis pipeline, covering pre-processing, data augmentation, segmentation, and classification. The paper comprehensively compares state-of-the-art techniques for skin lesion segmentation, highlights current limitations and identifies promising directions for developing diagnostic systems.

Introduction

The skin serves as a vital interface between the human body and the external environment, governing essential functions such as temperature regulation and fluid retention. Despite its resilience, the skin is prone to a multitude of pathologies. It is estimated that there are over 3,000 distinct types of dermatological disorders, making skin diseases one of the most prevalent and diagnostically challenging health concerns worldwide. Global Cancer Statistics 2020 states that fatal skin lesions claim thousands of lives annually[1]. More precisely, skin cancer ranks as the third most common human malignancy, with melanoma being its most aggressive and lethal form. Epidemiological data indicates a rapid surge in melanoma incidence over the last three decades. Notably, statistical projections estimated approximately 96,480 new diagnoses in the United States in 2019[2].

Dermoscopy, a non-invasive imaging technique, has improved diagnostic accuracy; however, manual interpretation of dermoscopic images is labor-intensive, subjective, and heavily dependent on the clinician's expertise. Consequently, Computer-Aided Diagnosis (CAD) systems have become indispensable tools in clinical dermatology. Within the CAD pipeline, skin lesion segmentation, the process of accurately delineating the lesion boundary from the surrounding healthy skin, is the most critical prerequisite.

Accurate recognition of melanoma presents significant challenges due to several inherent complexities. Firstly, the low contrast between lesions and the surrounding healthy skin often creates ambiguous boundaries[3,4,5]. Secondly, high variability in patient-specific attributes, ranging from skin pigmentation and texture to lesion morphology, complicates the detection process[6,4,5,7]. Furthermore, image quality is frequently compromised by various artifacts, including body hair, specular reflections, air bubbles, shadows, and inconsistent lighting conditions[3,6,4,7]. Thirdly, the scarcity of high-quality annotated training data poses a severe constraint on the model's generalization capability. Fourthly, the class imbalance problem, where the lesion area is disproportionately smaller than the background, significantly impedes segmentation performance. Notably, these aforementioned occlusions and artifacts are pervasive in standard public dermoscopic datasets. Figure 1 visually exemplifies these impediments, highlighting the complexity involved in precise boundary delineation.

While early Convolutional Neural Networks (CNNs) like Fully Connected Network (FCN) and SegNet demonstrated the feasibility of end-to-end segmentation, they were fundamentally constrained by the trade-off between context and localization. Specifically, the repeated down-sampling operations resulted in the degradation of high-frequency edge information. The emergence of U-Net revolutionized this landscape by introducing a novel mechanism: skip connections. These connections facilitate the direct concatenation of feature maps from the encoder to the decoder, enabling the network to combine semantic context with precise localization, a capability that has cemented U-Net's status as the dominant standard in the domain.

Despite their prevalence, CNN-based architectures utilize a fixed local receptive field, which inherently limits their ability to capture long-range dependencies and global semantic context. Conversely, while Transformers successfully mitigate this limitation through the self-attention mechanism, they introduce a new bottleneck: computational inefficiency. Specifically, the standard self-attention operation scales quadratically with the input resolution ($O(N^2)$), rendering it computationally prohibitive for high-resolution medical images.

This paper provides a comprehensive survey of the nascent Mamba architecture (Selective State Space Models) and its

rapid adoption in medical imaging. Unlike previous reviews that focus primarily on CNNs or Transformers, we critically analyze the paradigm shift towards linear complexity models ($O(N)$). Specifically, we investigate the latest hybrid frameworks emerging in 2024, such as AC-MambaSeg, highlighting their ability to maintain global receptive fields while significantly reducing computational overhead compared to quadratic Transformer-based approaches.

To visualize this landscape, **Fig. 1** presents the taxonomy of the computational techniques reviewed in this study. Consequently, the remainder of this paper is organized to reflect this evolutionary trajectory:

- **Section II:** outlines the essential computational pipeline, discussing pre-processing techniques and the role of GANs in addressing data scarcity.

- **Section III:** reviews the foundational era of CNN architectures, tracing the development from FCN and SegNet to the gold standard U-Net and its attention-based variants.

- **Section IV:** analyzes the paradigm shift towards global context modeling, highlighting the strengths and computational limitations of Transformer-based approaches like FAT-Net.

- **Section V:** investigates the emerging era of efficiency, specifically focusing on Mamba and hybrid State Space Models (e.g., AC-MambaSeg) that achieve linear complexity $O(N)$.

- **Section VI:** provides a comparative analysis and discusses the trade-offs between these architectures.

- **Section VII:** concludes the paper and outlines future research directions.

The Computational Pipeline & Data Preparation

Data Pre-processing

Image pre-processing is a crucial stage, entailing several steps to improve, standardize, or prepare images for additional modeling or analysis. Standard preprocessing methods include resizing, normalization, noise reduction color space conversion, and contrast correction. It improves the images' quality, uniformity, and effectiveness, which is crucial for computer vision tasks.

Data Augmentation via Generative Adversarial Networks (GANs)

Deep learning models need to be trained on enormous amounts of data to prevent overfitting and attain superior performance. However, the data is limited for several reasons, including patient privacy and the high cost of collecting medical records. Data augmentation helps to tackle this issue by transforming the available data using some operations or procedures to increase the dataset. Numerous methods including rotation, random cropping, transpose, vertical flip, horizontal flip only produce variations of existing samples and fail to introduce sufficient diversity into the data distribution.

Generative Adversarial Networks (GANs) have emerged as a breakthrough approach for synthetic data generation and robust feature learning.Fundamentally, a GAN architecture consists of two competing neural networks: a Generator and a Discriminator (D). (G) aims to synthesize realistic skin lesion images from random noise distributions, while (D) acts as a binary classifier, attempting to distinguish between real clinical images and the synthetic ones produced by G. Through a min-max adversarial training process, G progressively learns to produce high-fidelity lesion samples that capture the complex texture and color characteristics of real dermoscopic images, thereby enriching the training set and reducing the risk of overfitting. Beyond pure data synthesis, adversarial learning concepts have been successfully adapted to enhance segmentation performance directly.

A notable example is the work by Lei et al[4], who proposed a novel architecture featuring Dual Discriminators. In their framework, the segmentation network functions as the Generator, while two distinct Discriminators are employed to enforce structural consistency and boundary sharpness in the predicted masks. By subjecting the model to this "double scrutiny," the system learns to generate segmentation maps that are not only pixel-wise accurate but also visually consistent with expert annotations. This study demonstrates the versatile potential of GANs: acting as both a tool for data augmentation and a segmentation algorithm.

Methods

Topical subheadings are allowed. Authors must ensure that their Methods section includes adequate experimental and characterization data[3]necessary for others in the field to reproduce their work.

References

1. Sung, H. *et al.* Global cancer statistics 2020: Globocan estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA: A Cancer J. for Clin.* **71**, 209–249, DOI: https://doi.org/10.3322/caac.21660 (2021).

2. Siegel, R. L., Miller, K. D. & Jemal, A. Cancer statistics, 2019. *CA: A Cancer J. for Clin.* **69**, 7–34, DOI: https://doi.org/10.3322/caac.21551 (2019).

3. Adegun, A. & Viriri, S. Deep learning model for skin lesion segmentation: Fully convolutional network. *Image Analysis Recognit.* 232–242, DOI: https://doi.org/10.1007/978-3-030-27272-2_20 (2019).

4. Lei, B. *et al.* Skin lesion segmentation via generative adversarial networks with dual discriminators. *Med. Image Analysis* **64**, 101716, DOI: https://doi.org/10.1016/j.media.2020.101716 (2020).

5. Bagheri, F., Tarokh, M. J. & Ziaratban, M. Skin lesion segmentation from dermoscopic images by using mask r-cnn, retina-deeplab, and graph-based methods. *Biomed. Signal Process. Control.* **67**, 102533, DOI: https://doi.org/10.1016/j.bspc.2021.102533 (2021).

6. Liu, L., Mou, L., Zhu, X. X. & Mandal, M. Skin lesion segmentation based on improved u-net. *2019 IEEE Can. Conf. Electr. Comput. Eng. (CCECE)* 1–4, DOI: https://doi.org/10.1109/CCECE.2019.8861848 (2019).

7. Wu, H. *et al.* Fat-net: Feature adaptive transformers for automated skin lesion segmentation. *Med. Image Analysis* **76**, 102327, DOI: https://doi.org/10.1016/j.media.2021.102327 (2022).

8. Arora, R., Raman, B., Nayyar, K. & Awasthi, R. Automated skin lesion segmentation using attention-based deep convolutional neural network. *Biomed. Signal Process. Control.* **65**, 102358, DOI: https://doi.org/10.1016/j.bspc.2020.102358 (2021).

LaTeX formats citations and references automatically using the bibliography records in your .bib file, which you can edit via the project menu[8]. Use the cite command for an inline citation,.

For data citations of datasets uploaded to e.g. *figshare*, please use the `howpublished` option in the bib entry to specify the platform and the link, as in the `Hao:gidmaps:2014` example in the sample bibliography file.

Acknowledgements (not compulsory)

Acknowledgements should be brief, and should not include thanks to anonymous referees and editors, or effusive comments. Grant or contribution numbers may be acknowledged.

Author contributions statement

Must include all authors, identified by initials, for example: A.A. conceived the experiment(s), A.A. and B.A. conducted the experiment(s), C.A. and D.A. analysed the results. All authors reviewed the manuscript.

Additional information

To include, in this order: **Accession codes** (where applicable); **Competing interests** (mandatory statement).

The corresponding author is responsible for submitting a competing interests statement on behalf of all authors of the paper. This statement must be included in the submitted article file.

Condition	n	p
A	5	0.1
B	10	0.01

Table 1. Legend (350 words max). Example legend text.

Figures and tables can be referenced in LaTeX using the ref command, e.g and Table 1.
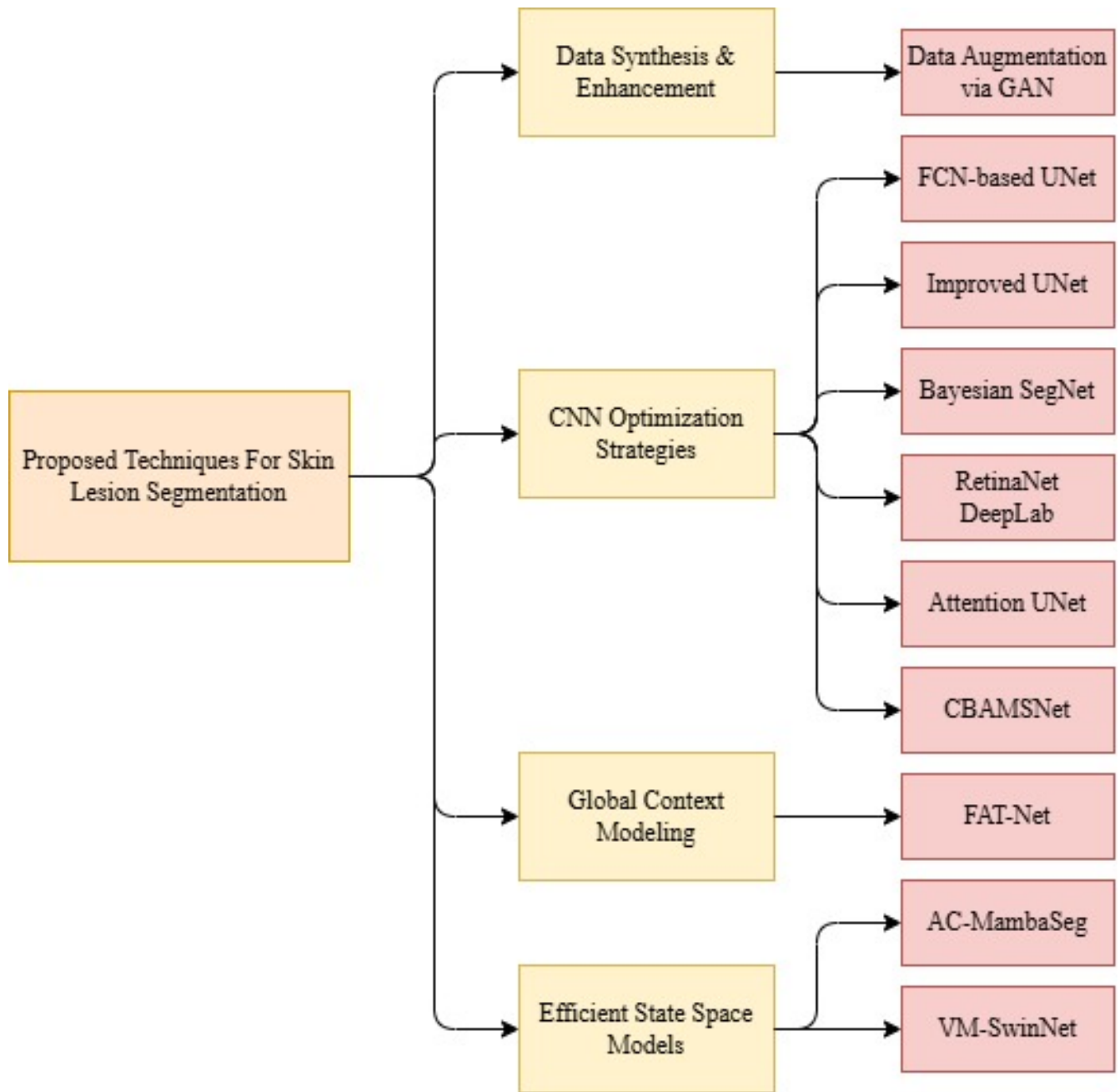
Figure 1. Taxonomy of computational techniques for skin lesion segmentation utilized in this survey, categorizing methods from foundational CNNs to state-of-the-art Hybrid Mamba architectures.