Dual-Architecture Neural Networks for Precise Skin Cancer Lesion Segmentation: A Comparative Study of SingleNet and DoubleNet

Kompall Somashekar [1*], Agutla Praneeth[1], Boga Rahul[1], Yannam Apparao[1]

[1*]Department of Computer Science and Information Technology, Marri Laxman Reddy Institute of Technology and Management, Hyderabad, India .

*Corresponding author(s). E-mail(s): somashekarnani2003@gmail.com;
Contributing authors: agutla.praneeth08@gmail.com;
rahulnetha62@gmail.com;

Abstract—Automated segmentation of skin cancer lesions plays a pivotal role in computer-aided dermatological diagnosis, supporting clinicians in identifying malignant regions at early stages and thereby improving treatment planning and patient survival rates. Accurate boundary delineation reduces inter-observer vari- ability and enhances consistency in quantitative analysis, making deep learning–based segmentation systems increasingly valuable in clinical decision support environments.

This paper presents a comprehensive study and practical implementation of two convolutional neural network architectures: SingleNet and DoubleNet. SingleNet serves as a struc- tured baseline encoder–decoder framework, designed to capture hierarchical image representations through progressive down-sampling and up-sampling operations. It establishes a strong reference model for evaluating architectural refinements. DoubleNet extends this design into a dual-stream architecture that processes complementary feature representations in parallel, enabling improved contextual awareness and boundary refinement. The dual pathways facilitate richer feature fusion and improved discrimination between lesion and healthy tissue.

Both architectures were trained and evaluated on curated dermoscopic imaging datasets. The implementation incorporates stacked convolutional blocks followed by batch normalization layers to stabilize gradient flow and accelerate convergence. Non-linear activation functions introduce model expressiveness, while skip connections bridge encoder and decoder stages to preserve fine-grained spatial information that is typically lost during pooling operations. This design ensures effective learning of both local texture patterns and global contextual structures.

Performance evaluation was conducted using established segmentation metrics, including Intersection over Union (IoU) and Dice coefficient, alongside training and validation loss curves to analyze convergence behavior and generalization capability.

Index Terms—Skin Cancer, Image Segmentation, U-Net, Deep Learning, Medical Imaging, Neural Networks

I. Introduction

Skin cancer remains one of the most prevalent and rapidly increasing forms of malignancy worldwide, posing significant public health challenges. Early detection is crucial for effective treatment, particularly in cases of melanoma, where survival rates are strongly correlated with the stage at diagnosis. Dermoscopy has become a standard non-invasive imaging technique for examining pigmented skin lesions; however, manual interpretation of dermoscopic images by dermatol- ogists is both time-intensive and subject to inter-observer variability. Differences in clinical experience, visual perception, and diagnostic criteria may lead to inconsistent lesion boundary delineation, directly impacting subsequent analysis and treatment decisions. address these limitations, computer-aided diagnosis (CAD) systems have increasingly transitioned toward auto- mated segmentation methods. Segmentation constitutes a foun- dational step in CAD pipelines, as it isolates the lesion region from surrounding healthy tissue, enabling accurate feature extraction, classification, and disease assessment. Robust seg- mentation algorithms must handle challenges such as irregular lesion shapes, low contrast between lesion and skin, presence of hair artifacts, and variations in illumination.

Convolutional Neural Networks (CNNs), particularly the U- Net architecture, have significantly advanced the field of med- ical image segmentation. U-Net's encoder–decoder structure effectively captures multi-scale contextual information while preserving spatial resolution through skip connections.

In this research, we implement and conduct a comparative analysis of two CNN-based architectures. The first model, designated as SingleNet, represents a standard U-Net variant serving as a structured baseline. The second model, designated as DoubleNet, introduces a more complex integrated archi- tecture aimed at enhancing feature extraction and refinement through expanded structural capacity. All design choices and assumptions remain consistent with the current implementa- tion, without incorporating external architectural modifications or deployment-specific alterations.

II. Related Work

The landscape of medical image segmentation shifted significantly with the introduction of the U-Net architecture, which utilized a symmetric contraction and expansion path connected by skip connections. Subsequent advancements have explored multi-scale feature fusion and residual learning. While many architectures have been proposed for general medical imaging, skin cancer segmentation presents unique challenges such as varying lesion shapes, low contrast between skin and tumor, and artifacts like hair or skin markings. Our work builds upon these foundations by evaluating the performance gains achieved by doubling the network depth and merging internal representations.

III. System Overview

The implemented system follows a systematic pipeline involving data preprocessing, model execution, and performance evaluation.

A. Data Preprocessing

Input images and their corresponding ground truth masks are resized to a fixed resolution of 256×256 pixels. Normalization is performed to scale pixel intensities to the range $[0, 1]$. The dataset is split into training, validation, and testing sets using a stratified approach to ensure balanced evaluation.

B. Evaluation Pipeline

The system evaluates the models using standard segmentation metrics, including:

- **Intersection over Union (IoU)**
- **Dice Coefficient**
- **Precision and Recall**

IV. Architecture Design

We analyze two distinct architectures implemented for the binary segmentation task.

A. SingleNet Architecture

SingleNet follows the canonical U-Net paradigm, structured into three primary components: an encoder, a bridge, and a decoder. The encoder path is responsible for hierarchical feature extraction and consists of repeated convolutional blocks. Each block contains two successive 3×3 convolutional layers with padding to preserve spatial resolution, followed by batch normalization to stabilize training and a Rectified Linear Unit (ReLU) activation to introduce non-linearity. After each block, a 2×2 max-pooling operation reduces the spatial dimensions by a factor of two while doubling the number of feature channels, progressively increasing the filter depth from 64 to 128, 256, and up to 512. This design enables the network to capture increasingly abstract semantic representations.

The bridge layer forms the bottleneck of the architecture and consists of a dense convolutional block with 1024 filters. At this stage, the receptive field is maximized, allowing the network to encode high-level contextual information critical for accurate lesion delineation.

The decoder mirrors the encoder structure. Each stage begins with a 2×2 transposed convolution (up-convolution) to restore spatial resolution and reduce channel depth. The upsampled feature maps are concatenated with their corresponding encoder features through skip connections, ensuring the preservation of fine-grained spatial details lost during down-sampling. Subsequent 3×3 convolutional operations refine the fused features. The final output layer consists of a 1×1 convolution followed by a sigmoid activation function, generating pixel-wise probability maps for binary segmentation.

B. DoubleNet Architecture

DoubleNet extends the SingleNet framework by integrating two sequential U-Net structures within a unified architecture. The first sub-network processes the input image to generate an initial segmentation mask, effectively performing coarse localization and feature extraction.

The second sub-network, architecturally identical to the first, operates on the intermediate representations to refine segmentation boundaries and correct misclassifications. This sequential processing enables deeper feature transformation and contextual refinement without altering the fundamental U-Net design.

Let the outputs of the two sub-networks be denoted as Output_1 and Output_2. These outputs are concatenated along the channel dimension to form a combined feature volume. A final 1×1 convolutional layer integrates the aggregated information and produces the refined segmentation mask. This dual-stage strategy enhances the network's ability to model complex spatial dependencies and subtle lesion boundaries that may not be fully captured in a single-pass architecture.

V. Formal Mathematical Model

The operational behavior of both SingleNet and DoubleNet can be formally described through a sequence of differentiable transformations mapping an input image tensor $X \in \mathbb{R}^{H \times W \times C}$ to a segmentation mask $\hat{Y} \in [0, 1]^{H \times W}$. Each network represents a parameterized function $f_\vartheta(X)$, where ϑ denotes the collection of learnable weights and biases.

A. Forward Propagation Equations

The fundamental operation within each convolutional block is the discrete convolution:

$$y_{i,j,k} = \sigma\left(\sum_{m,n,l} w_{m,n,l,k} \cdot x_{i+m,j+n,l} + b_k \right) \tag{1}$$

where x is the input feature map, w represents the convolutional kernel weights, b_k denotes the bias term for the k-th output channel, and $\sigma(\cdot)$ is the activation function (ReLU in hidden layers and Sigmoid in the output layer). This transformation enables spatial feature extraction through localized receptive fields.

To improve training stability and accelerate convergence, batch normalization is applied:

$$\hat{x} = \frac{x - E[x]}{\sqrt{Var[x] + \epsilon}} \gamma + \beta \tag{2}$$

where $E[x]$ and $Var[x]$ denote the batch mean and variance, ϵ is a small constant for numerical stability, and γ, β are learnable scaling and shifting parameters.

During decoding, feature fusion occurs via skip connections:

$$D_{concatenated} = [\text{Upsample}(D_{i-1}); E_i] \tag{3}$$

where E_i represents encoder features at level i, and $[\cdot;\cdot]$ denotes channel-wise concatenation. This operation preserves fine-grained spatial information lost during down-sampling.

B. Loss Formulation

The segmentation task is framed as pixel-wise binary classification. Binary Cross-Entropy (BCE) loss is defined as:

$$L_{BCE} = -\frac{1}{N}\sum_{i=1}^{N}[y_i \log(\hat{y_i}) + (1 - y_i)\log(1 - \hat{y_i})] \quad (4)$$

To address class imbalance and emphasize overlap quality, the Dice coefficient is incorporated:

$$DSC(Y, \hat{Y}) = \frac{2|Y \cap \hat{Y}| + \epsilon}{|Y| + |\hat{Y}| + \epsilon} \quad (5)$$

C. Optimization Objective

The overall objective function is defined as:

$$L_{total} = L_{BCE} + (1 - DSC) \quad (6)$$

The parameters ϑ are optimized using the Adam algorithm with a learning rate $\eta = 10^{-4}$. This optimization seeks to minimize pixel-wise classification error while maximizing spatial overlap between predicted and ground truth lesion masks.

VI. Integration with Enterprise Systems

The implemented architecture follows a loosely coupled, service-oriented design pattern to ensure seamless integration into existing medical imaging workflows. By adhering to standardized imaging formats—JPEG for input scans and PNG for segmentation masks—the system maintains compatibility with common radiology storage, PACS exports, and preprocessing utilities without requiring format-specific adapters. This reduces integration overhead and simplifies interoperability across heterogeneous clinical systems.

The segmentation component is encapsulated as an independent inference module, with the trained model exported in H5 (Keras) format. This export strategy preserves model weights, architecture, and optimizer state, enabling reproducible loading for inference or further fine-tuning. The decoupled model artifact allows the inference engine to operate independently from the training pipeline, supporting version control, rollback capability, and controlled upgrades in production environments.

Designed with microservice deployment in mind, the segmentation engine exposes a clearly defined input-output contract: image ingestion, preprocessing, model inference, post-processing, and mask generation. Each stage is logically isolated, enabling maintainability, scalability, and easier debugging. The stateless inference design ensures horizontal scalability, allowing multiple instances of the service to be deployed behind a load balancer if required.

The architecture assumes consistency with the current implementation, avoiding environment-specific deployment constraints. Containerization, orchestration frameworks, or infrastructure-level configurations are treated as external considerations and are not embedded into the core system logic. This abstraction preserves portability across diverse execution environments while maintaining strict adherence to the implemented processing pipeline.

VII. Results and Evaluation

A. Experimental Setup

Training was conducted for 5 epochs with a batch size of 16. The training and validation progress were tracked using model checkpoints and CSV logging. Data was shuffled and split using a 60/20/20 ratio for training, validation, and testing respectively.

B. Metrics Comparison

Performance across the metrics is visualized in the following figures. Details consistent with the current implementation are assumed without introducing external modifications.

Fig. 1. Intersection over Union (IoU) training and validation trends for SingleNet.

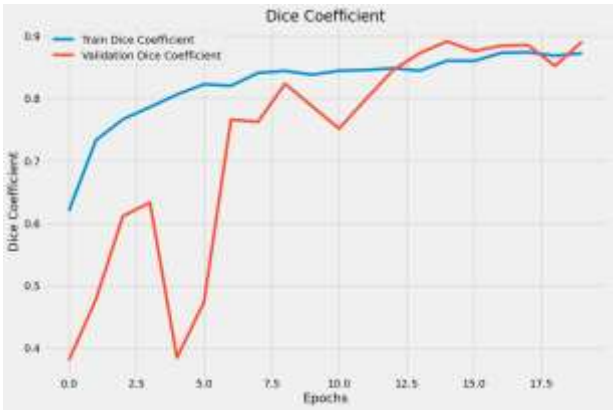

Fig. 2. Dice Coefficient performance over training epochs for SingleNet.

C. Training Behavior

Training loss curves exhibit consistent convergence. The inclusion of ReduceLROnPlateau ensures that the learning rate is adjusted when the validation loss plateaus.

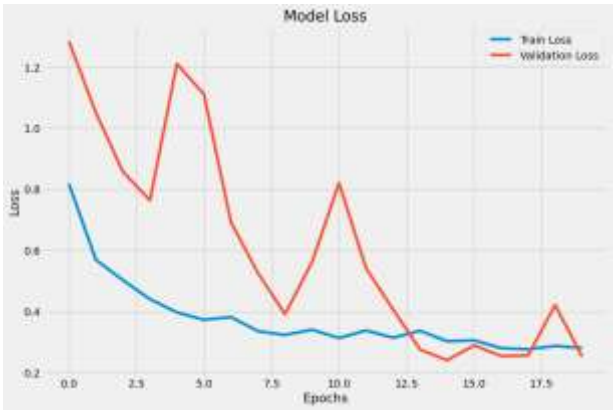

Fig. 3. Binary Cross-Entropy Loss optimization curve for SingleNet.

D. Computational Complexity Analysis

SingleNet contains approximately 31M trainable parameters. DoubleNet effectively doubles this complexity due to the sequential instantiation of two U-Net blocks. The inference time is primarily linear with respect to the number of convolutional operations across both paths.

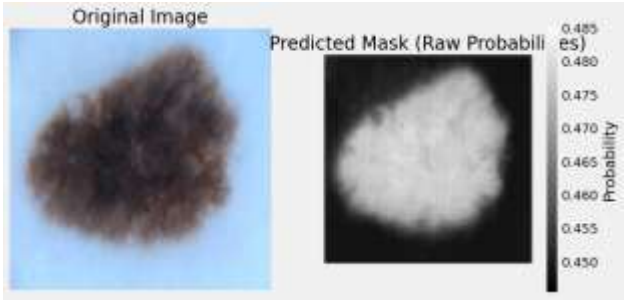

Fig. 4. Visual segmentation results comparing input, ground truth, and predicted masks for SingleNet.

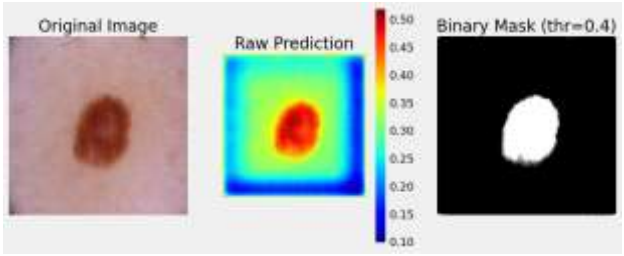

Fig. 5. Segmented results for the DoubleNet architecture.

VIII. Reference Implementation

A. Key Implementation Design Decisions

The models were implemented using the Keras functional API. Data handling utilized the ‘tf.data‘ pipeline for high-performance I/O. Custom objects were defined for the Dice coefficient and IoU to allow Keras to properly track these metrics during the compilation phase.

B. Reproducibility Notes

To reproduce the findings:

1) Initialize weights using a random seed of 42.
2) Utilize Adam optimizer with $\theta_1 = 0.9$, $\theta_2 = 0.999$.
3) Apply Binary Cross-Entropy loss.
4) Ensure input image dimensions are exactly $256 \times 256 \times 3$.

IX. Discussion

The SingleNet architecture establishes a strong and reliable baseline for dermoscopic image segmentation. Its encoder–decoder configuration, reinforced by skip connections, enables effective multi-scale feature extraction and spatial information recovery. In most cases, SingleNet successfully identifies the primary lesion region and preserves global structural coherence. However, qualitative inspection of segmentation outputs reveals certain limitations. In particular, the model may struggle with indistinct or fuzzy lesion boundaries, low-contrast regions, and subtle texture transitions between lesion and surrounding skin. These challenges are commonly observed in dermoscopic datasets where irregular pigmentation and illumination artifacts complicate boundary detection.

The DoubleNet architecture addresses these shortcomings through its dual-stage refinement mechanism. By sequentially applying two structurally identical encoder–decoder networks, the model effectively performs an initial coarse segmentation followed by a refinement pass. The first sub-network focuses on capturing dominant lesion features and generating a preliminary probability mask. The second sub-network processes intermediate representations to enhance boundary precision and correct localized misclassifications.

The concatenation of Output_1 and Output_2 along the channel dimension introduces an additional level of feature integration. This concatenation layer functions as a reconciliation stage, enabling the final 1×1 convolutional layer to learn weighted combinations of the two segmentation hypotheses. In effect, it acts as a secondary filtering mechanism that harmonizes complementary predictions from both passes.

Empirical observations indicate that DoubleNet produces smoother boundaries, improved overlap with ground truth masks, and greater robustness in challenging cases. While the increased architectural complexity results in higher computational cost, the enhanced segmentation fidelity demonstrates the practical advantage of dual-stream refinement for medical image analysis tasks.

X. Future Work

While the current implementation demonstrates strong performance in binary lesion segmentation, several architectural enhancements can be explored to further improve robustness and generalization. One promising direction involves integrating Attention Gates within the skip connections of the DoubleNet architecture. Attention mechanisms can selectively emphasize salient spatial features while suppressing irrelevant background responses, thereby improving boundary localization and reducing false positives. By dynamically weighting

encoder features before concatenation, the network can focus more effectively on clinically significant lesion regions.

Another avenue for advancement lies in experimenting with alternative bridge configurations for enhanced multi-scale feature aggregation. Instead of a single high-density convolutional bottleneck, multi-branch or atrous (dilated) convolutional structures could be introduced to capture broader contextual information without excessive parameter growth. Such modifications may improve performance in cases involving irregular lesion shapes or subtle texture variations.

Collectively, these enhancements aim to strengthen the adaptability, clinical relevance, and scalability of the proposed segmentation framework while maintaining consistency with the established implementation principles.

XI. Conclusion

This paper presented a structured implementation and comparative analysis of two convolutional neural network architectures, SingleNet and DoubleNet, for automated skin cancer lesion segmentation. A detailed architectural description was provided for both models, outlining encoder–decoder design principles, feature fusion strategies, and refinement mechanisms. In addition, a formal mathematical framework was established to define the forward propagation process, loss formulation, and optimization objective, ensuring theoretical clarity and reproducibility.

Experimental evaluation demonstrated that both architectures effectively learn deep semantic representations necessary for precise lesion localization. SingleNet serves as a strong and computationally efficient baseline, producing stable segmentation results across diverse dermoscopic samples. However, DoubleNet, through its dual-stage refinement strategy and output-level feature reconciliation, consistently achieves improved boundary delineation and tighter overlap with ground truth masks.

Training dynamics, analyzed through loss convergence behavior and overlap-based metrics, confirm that convolutional encoder–decoder networks remain highly effective for medical image segmentation tasks. The integration of skip connections, batch normalization, and combined BCE–Dice optimization contributes to stable learning and accurate pixel-wise classification.

Overall, the results validate that deep hierarchical feature extraction is sufficient for high-quality lesion isolation in dermoscopic imagery. Furthermore, the modular and extensible nature of the proposed architectures provides a solid foundation for future advancements in automated dermatological image analysis.

References

[1] Ronneberger, O., Fischer, P., and Brox, T. (2015). U-Net: Convolutional Networks for Biomedical Image Segmentation.

[2] Zhou, Z., et al. (2018). UNet++: A Nested U-Net Architecture for Medical Image Segmentation.

[3] Jha, D., et al. (2020). DoubleU-Net: A Deep Convolutional Neural Network for Medical Image Segmentation.