Edge-Guided U-Net with Frequency Domain Refinement for Efficient Skin Lesion Segmentation

Aryan Kumar Singh[1], Md Shohan Mia[1], Md Rakibul Islam Midul[2], Priyanshu Kumar Gupta[2], Md Kayser Ahmed Hridoy[2], Chowdhury Hasibur Rahaman[2], and Rabul Hussain Laskar[2]

[1] Department of Computer Science and Engineering, National Institute of Technology Silchar
[2] Department of Electronics and Communication Engineering, National Institute of Technology Silchar

Corresponding author(s):rhlaskar@ece.nits.ac.in **Contributing author(s):**aryan21_ug@cse.nits.ac.in md.shohan21_ug@cse.nits.ac.in rakibuli_ug_22@ece.nits.ac.in priyanshug_ug_22@ece.nits.ac.in mdkayser21_ug@ece.nits.ac.in chr.siam7@gmail.com

Abstract. Precise segmentation of skin lesions is crucial to support the diagnosis of skin cancer within computer-aided diagnosis (CAD) systems. Although deep learning-based models like U-Net along with its variants have demonstrated strong segmentation performance, they often require substantial computational and energy resources. In this study, a lightweight and computationally efficient framework for skin lesion segmentation is proposed, leveraging information across the spatial and frequency domains. The proposed model, Edge-Guided U-Net (EGU-Net), integrates edge information directly into the input and incorporates a frequency-domain refinement step based on the discrete cosine transform (DCT). Through this integration, EGU-Net enables improved delineation of lesion boundaries while suppressing high-frequency noise in the predicted masks. The model underwent training and evaluation on the ISIC 2018 dataset, resulting in a Dice score of 87.85%, recall of 90.62%, precision of 89.02%, an intersection over union (IoU) of 80.35% and an accuracy of 95.47%. These results demonstrate that EGU-Net can approximate the performance of more complex architectures while maintaining significantly lower computational complexity. Consequently, EGU-Net establishes a strong baseline for interpretable and efficient pixel-wise skin lesion segmentation under constrained computational settings.

Keywords: Skin Lesion Segmentation · Edge-Guided U-Net · Frequency Domain Filtering · Discrete Cosine Transform (DCT) · Deep Learning · ISIC 2018 Dataset · Lightweight Architecture

1 Introduction

Skin cancer is one of the most common and deadly forms of cancer worldwide, and melanoma is the leading cause of skin cancer-related deaths, making it one

of the most prevalent and dangerous cancers globally. Accurate and automatic segmentation of skin lesions from dermoscopic pictures is critical for early detection and intervention in computer-aided diagnosis (CAD). Deep learning has greatly improved medical picture segmentation, with U-Net[8] and its derivatives leading the way. The encoder-decoder structure and skip connections effectively preserve spatial features, resulting in accurate segmentation[1–5].

However, despite their effectiveness, standard U-Net[8] architectures often struggle with challenges such as imprecise lesion boundaries, noise sensitivity, and suboptimal performance in cases where the lesions are mixed with the surrounding skin. To address these problems, recent studies have focused on improving segmentation networks with edge guidance mechanisms[2–4], attention modules[1] and hybrid spatial frequency techniques. Edge data plays a particularly significant role in medical images, where the definition of a boundary influences direct diagnostic precision[2, 5].

Furthermore, the attention-based and transformed architecture has achieved competitive performance in segmentation tasks[1]. However, practical deployment is hindered by the need for large datasets, high training costs and substantial memory requirements. It makes less potential for use in their resource-constrained clinical settings or on edge devices where the speed of assumptions, simplicity and model transparency are critical.

In our research, we offer a unique segmentation framework termed Edge-Guided U-Net with Frequency Domain Refinement, which is aimed to be both effective and reliable. Our technique enhances a lightweight U-Net model via the combination of two complementary techniques:

i) Edge Guidance: We create a Sobel edge map [6] for each image and concatenate it with the RGB channels to shape a 4-channel input. The edge map is normalized to match the RGB range, helping the network focus on lesion boundaries early in training and improving edge localization [2, 4]. Sobel is selected for efficiency and stability, though alternatives like Canny or learned edge detectors could be explored in future work.

ii) Frequency Domain Refinement: After obtaining the initial segmentation mask, we apply a Discrete Cosine Transform (DCT) [7], retain only the low-frequency coefficients to suppress high-frequency noise and reconstruct the mask using inverse DCT (IDCT). This smooths lesion contours while maintaining key structures and reducing false positives [5, 8]. A fixed cutoff is used for simplicity, with adaptive selection as a potential improvement [7].

We tested our technique on the ISIC 2018 Skin lesion dataset [18], achieving a Dice score of 87.85%, an IoU of 80.35%, and an accuracy of 95.47%. We used a low-resolution input size (128×128), few training epochs, and avoided significant tweaking or ensemble techniques. Combining standard computer vision priors with deep learning is effective, especially in resource-constrained applications[1].

While recent transformer-based models have shown promise in medical image segmentation[1], they are often computationally expensive and require large training datasets. This work proposes an efficient alternative that achieves strong

performance using lightweight modules based on edge information and frequency-domain refinement.[2, 3, 5]

The following is a summary of this paper's primary contributions:

- A novel 4-channel input formulation combining RGB images and Sobel edge maps to enhance spatial boundary information.
- A lightweight frequency-domain post-processing step using DCT-based low-pass filtering to refine the predicted segmentation masks.
- A compact and interpretable segmentation framework (EGU-Net) demonstrating strong performance on the ISIC 2018 Skin lesion dataset [18] under limited computational settings.
- According to empirical validation, the suggested method works comparably with more intricate deep structures, which qualifies it for embedded or real-time deployment.

This is how the rest of the paper is structured. The literature on edge-aware deep learning models for skin lesion segmentation is reviewed in 2. The architecture and design of the suggested EGU-Net framework are described in Section 3. The dataset, implementation specifics, and assessment findings, Section 4. A thorough analysis of the findings and a comparison with current state-of-the-art methods are given in are presented in Section 5 . In section 6 summarizes the findings and highlights the contributions of this work, followed by future work, which outlines potential direction and extension of this research.

2 Related Work

2.1 Attention Guided Approaches:

Ronneberger et al. [8] have introduced U-Net, an encoder-decoder network with elastic deformation augmentation, which achieves 92% and 77.5% IoU on two datasets. However, it remains largely reliant on proper data augmentation and may fail to generalize to problems with vastly diverse data distributions. Schlemper et al. [9] have proposed an Attention U-Net with soft attention gates, which improves pancreatic segmentation dice scores from 81.4% to 84.0%. Still, the method increases complexity and may struggle to generalize across highly variable unseen data. Chen et al. [10] have proposed Channel-UNet with spatial channel-wise convolution, attaining 98.4% and 94.0% dice scores; however, it is dependent on dataset volume and quality, limiting its broader application.

2.2 Multi-Scale Context Approaches:

Alom et al. [11] have proposed R2U-Net, which incorporates recurrent and residual convolutional blocks into U-Net to improve nuclei segmentation, resulting in a 92.15% dice score. Jha et al. [12] have developed ResUNet++, combining residual, squeeze-and-excitation, ASPP and attention blocks to improve polyp segmentation, achieving dice scores of 81.3% and 79.6%. While effective, the added

complexity increases training time. Azad et al. [13] have developed BCDU-Net, improving U-Net with Bi-Directional ConvLSTM and dense convolutions to increase segmentation accuracy, attaining dice scores of up to 82.2% for retinal vessels and 85.1% for skin lesions.

2.3 Transformer based Approaches:

Chen et al. [1] have proposed TransAttUnet, a Transformer-based U-Net with multi-level attention and multi-scale skip connections for superior segmentation. The model's reliance on global self-attention raises memory and compute costs, restricting its applicability in resource-constrained environments. Ji, Yuanfeng, et al. [14] have introduced a Multi-Compound Transformer with TSA and TCA modules in a UNet-based architecture, achieving best DSC scores over UNet variants. Though effective, it adds moderate computational cost and requires careful tuning. Zhang et al. [15] have proposed TransFuse, a hybrid Transformer–CNN with a BiFusion module, achieving top segmentation accuracy (e.g., 0.942 mDice on polyp data) while using fewer parameters and faster inference. However, it relies on large-scale pretraining and manual hyperparameter tuning. Xie, Enze, et al. [16] have developed SegFormer, combining a hierarchical Transformer encoder without positional encodings and an MLP decoder, achieving 84.0% mIoU on Cityscapes and 51.8% on ADE20K with strong speed–accuracy trade-offs. Despite this, even the smallest models remain relatively heavy for edge deployment. Cao et al. [17] have proposed Swin-Unet, a pure Transformer U-Net using Swin blocks and patch-based skip connections, achieving 79.1% DSC on Synapse and 90.0% on ACDC; though accuracy benefits from global attention, it relies on ImageNet pretraining and is limited to 2D tasks.

3 Methodology

In this section, a proposed Edge-Guided U-Net (EGU-Net) framework for skin lesion segmentation is detailed. The proposed method incorporates edge guidance at the input stage and frequency-domain refinement to improve the segmentation output. The pipeline is composed of three steps: edge-based input encoding, a modified U-Net [8] backbone and frequency-domain post-processing. The overview of proposed approach is presented in Fig.1

3.1 Data Preprocessing

All skin lesion samples and corresponding binary masks from the ISIC 2018 Skin lesion dataset [18] were rescaled to 128×128 pixels to reduce computational overhead. Pixel intensities were scaled to the $[0, 1]$ range.

To enhance the network's awareness of lesion boundaries, structural information was extracted by first converting each RGB image to its grayscale equivalent. Edge maps were then computed using the Sobel [6] operator, which highlights gradient changes corresponding to object contours. The Sobel edge maps were

normalized separately to the $[0, 1]$ range prior to concatenation for consistency with the RGB channels. No further smoothing was applied since Sobel filtering produces robust gradient responses without introducing artifacts. We selected Sobel over alternatives like Canny or learned edge detectors because it is computationally lightweight, parameter-free, and deterministic, making it appropriate for use in resource-limited clinical settings. The resulting edge map was concatenated with the original RGB image to form a 4-channel input tensor, comprising the three RGB channels and an additional edge channel. This enriched representation enabled the model to better learn spatially discriminative features, particularly around the lesion boundaries.

3.2 EGU-Net Architecture

The proposed segmentation framework is a modified U-Net [8] architecture designed to process a 4-channel input. It follows an encoder–decoder design with skip connections to keep spatial information intact across the network.

The encoder is structured with four stages, with each stage containing two consecutive 3×3 convolution operations, each succeeded by batch normalization and a ReLU nonlinearity, with downsampling accomplished via 2×2 max-pooling operations. The decoder mirrors the encoder structure, incorporating transposed convolutional layers for upsampling and skip connections from the corresponding encoder stages to recover spatial context. The architecture ends with a 1×1 convolution, which is followed by a sigmoid activation function, producing a one-channel output probability map.

The 4-channel input tensor $X \in \mathbb{R}^{4 \times H \times W}$ is passed through the encoder to extract hierarchical feature representations, which are then restored in the decoder path to yield the predicted segmentation mask $\hat{Y}$.

3.3 Frequency-Domain Refinement

To enhance the structural consistency and smoothness of the predicted masks, a post-processing step is applied in the frequency domain. The predicted mask $\hat{Y}$ is transformed using the two-dimensional Discrete Cosine Transform (DCT) [7]. To reduce high-frequency noise and artifacts, low-pass filtering is performed by retaining only the top-left 30×30 DCT coefficients, which predominantly encode the low-frequency components of the mask.

The 30×30 fixed cutoff was empirically selected as it offered an optimal balance between preserving lesion boundaries and removing noise during preliminary experiments. Moreover, this fixed-size cutoff ensures uniform and computationally inexpensive refinement, which is crucial for deployment in resource-constrained environments.

The filtered frequency representation is then transformed back into the spatial domain using the IDCT [7], yielding an enhanced output mask $\tilde{Y}$. This process sharpens lesion boundaries and suppresses spurious predictions, thereby contributing to the overall robustness of the segmentation.

Although our refinement approach is non-trainable and parameter-free, this design choice was intentional to preserve model simplicity, transparency, and efficiency. In future work, we plan to explore adaptive frequency selection schemes and differentiable frequency-domain layers, which would enable the network to learn optimal filtering strategies automatically without compromising computational efficiency.

3.4 Loss Function and Optimization

The training of the model employs a hybrid loss function that integrates Binary Cross-Entropy (BCE) with Dice loss, which balances pixel-wise accuracy with region overlap. Specifically, the BCE loss as described in Eq. 1 measures pixel-level classification accuracy by penalizing incorrect predictions of lesion versus background.

$$\mathcal{L}_{\text{BCE}} = -\frac{1}{N}\sum_{i=1}^{N}\left[y_i \log(\hat{y}_i) + (1-y_i)\log(1-\hat{y}_i)\right] \tag{1}$$

As expressed in Eq. 2, the Dice loss as shown in Eq. 2 directly optimizes the intersection of predicted and reference regions, which is particularly effective for handling imbalance between classes in medical image segmentation.

$$\mathcal{L}_{\text{Dice}} = 1 - \frac{2\sum_{i=1}^{N} y_i \hat{y}_i}{\sum_{i=1}^{N} y_i + \sum_{i=1}^{N} \hat{y}_i + \epsilon} \tag{2}$$

The two terms are combined into a total hybrid loss as explained in Eq. 3, where $\alpha \in [0,1]$ is a balancing coefficient that controls the relative contribution of each component. In our experiments, α was set to 0.5 to give equal weight to BCE and Dice losses.

$$\mathcal{L}_{\text{Total}} = \alpha\,\mathcal{L}_{\text{BCE}} + (1-\alpha)\,\mathcal{L}_{\text{Dice}} \tag{3}$$

In this context, $y_i \in \{0,1\}$ denotes the true annotation for pixel i, $\hat{y}_i \in [0,1]$ represents estimated probability corresponding to the same pixel, N N denotes the total pixel count in the image or batch, while ϵ is a tiny positive constant added to avoid division-by-zero errors.

The Adam optimizer with a learning rate of 1×10^{-4} is used for optimization. The model is trained for 50 epochs with a batch size of 16 on a GPU-enabled system.

3.5 Algorithm Description

The overall training and inference process of EGU-Net is outlined in Algorithm 1.

Algorithm 1 Procedure for Training and Inference in EGU-Net

Require: Image set $\mathcal{X}$, mask set $\mathcal{Y}$, total epochs E, batch size B

Resize images and masks to 128×128 and normalize to $[0, 1]$
for each image $x_i \in \mathcal{X}$ **do**
  Convert x_i to grayscale
  Compute edge map e_i using Sobel operator
  Normalize e_i to $[0, 1]$
  Concatenate e_i with RGB channels to form $x_i' \in \mathbb{R}^{4\times128\times128}$
end for
Apply data augmentation (random flips, rotations, scaling, brightness/contrast adjustments)
Initialize model parameters θ
for epoch $= 1$ to E **do**
  for each mini-batch $\{x_j', y_j\}_{j=1}^{B}$ **do**
    Predict mask: $\hat{y}_j = \text{EGU-Net}(x_j'; \theta)$
    Compute hybrid loss:

$$\mathcal{L} = \alpha\, \mathcal{L}_{\text{BCE}}(\hat{y}_j, y_j) + (1 - \alpha)\, \mathcal{L}_{\text{Dice}}(\hat{y}_j, y_j)$$

    Update parameters: $\theta \leftarrow \theta - \eta \cdot \nabla_\theta \mathcal{L}$
  end for
end for
for each test sample x_t' **do**
  Predict mask: $\hat{y}_t = \text{EGU-Net}(x_t')$
  Apply DCT: $D = \text{DCT}(\hat{y}_t)$
  Retain only low-frequency coefficients (top-left 30×30 block of D)
  Apply IDCT: $\tilde{y}_t = \text{IDCT}(D)$
end for

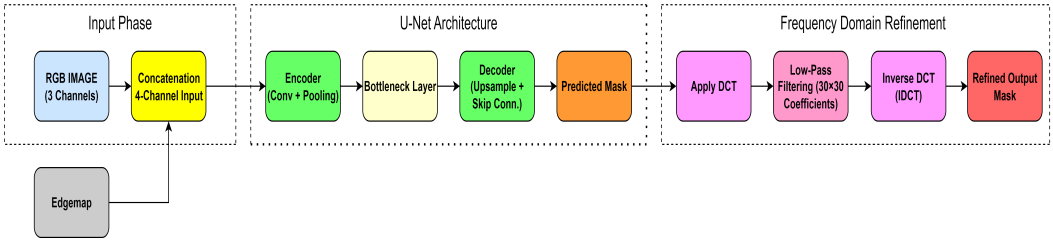

Fig. 1. The overview of the proposed EGU-Net architecture.

4 Experiments

This section demonstrates the evaluation of the Edge-Guided U-Net (EGU-Net) model and how it performs against a number of other existing segmentation

methods. The experimental focus will be on checking the quality of segmentation, the efficiency of the model, as well as the ability to characterize the edges of lesions. Performance is assessed quantitatively and visually to support the model performance.

4.1 Datasets

The evaluation is conducted using the publicly available The evaluation follows the ISIC 2018 Skin Lesion Segmentation Dataset [18] that has been made publicly available for the ISIC 2018 challenge, which is a publicly available dataset of a sizeable dermoscopic images collection and ground truth masks that correspond to its lesions area.

For the purpose of the experiment, specifically, 2,596 dermoscopic images and related binary segmentation labels are used. A random split of the dataset assigns 2,076 images, representing 80%, to the training set while the rest, 520 images for testing(20%). The dermoscopic images contain a diverse appearance of lesion sizes, shapes, colors and surfaces, against which segmentation tasks present a difficult benchmark dataset, as they share common characteristics. This variability represents a good opportunity for testing a model's robustness or generalizability, with respect to different lesion appearance.

4.2 Experimental Settings

Baselines: To assess the effectiveness of the EGU-Net system, multiple well-established U-Net-based segmentation architectures are selected as baselines. These include the original U-Net [8], ResUNet [12], Attention U-Net [9] and BCDU-Net [13]. These models serve as representative methods in the domain of segmentation of medical images and allow to enable an unbiased evaluation against the proposed approach.

Implementation Details: All tests were carried out using the dataset [18] of ISIC 2018 skin lesion segmentation. [18]. Each input image and its corresponding mask were resized to 128×128 pixels. This resolution was selected to mimic deployment in environments with limited resources, where low computational and memory costs are essential. While higher resolutions can improve delineation of fine-grained boundaries, our primary aim here is to demonstrate efficiency under constrained conditions. Every model was developed using the PyTorch platform.

To enhance robustness and generalization, the training procedure incorporated data augmentation. The training data underwent augmentation, including flips along both axes, rotations within $\pm 15°$, random scaling, and slight brightness/contrast adjustments. These augmentations simulate real-world variations in dermoscopic images such as changes in orientation, illumination, and lesion size.

The Adam optimizer with a 1×10^{-4} learning rate was used to train all models for 50 epochs and the hybrid BCE+Dice loss function defined in Eq. 3. The batch

size was 16. Predicted segmentation outputs were thresholded at 0.5 to obtain binary masks for evaluation. No pretraining or model ensembling was used; only the proposed frequency-domain refinement was applied as post-processing in EGU-Net.

Evaluation Metrics: To evaluate segmentation performance, five standard metrics are used: Dice coefficient (DICE), Recall (REC), Accuracy (ACC), Intersection over Union (IoU) and Precision (PRE). These metrics are computed from per-pixel tallies of true negatives (TN), false positives (FP), false negatives (FN) and true positives (TP).

The Dice coefficient, as described in Eq. 4, quantifies the similarity between the predicted mask and the actual ground truth, giving more weight to correct positive predictions. As presented in Eq. 5 the Intersection over Union, also evaluates overlap but penalizes false positives more strongly. Accuracy, as given in Eq. 6, evaluates the ratio of accurately predicted pixels across both lesion and background. According to Eq. 7, recall determines how many lesion pixels are properly detected, while Precision, as shown in Eq. 8, it evaluates the fraction of predicted lesion pixels that are accurate that are actually correct. Together, they deliver a complete evaluation of both the segmentation boundaries and the overall quality.

$$\text{Dice} = \frac{2 \times TP}{2 \times TP + FP + FN} \tag{4}$$

$$\text{IoU} = \frac{TP}{TP + FP + FN} \tag{5}$$

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \tag{6}$$

$$\text{Recall} = \frac{TP}{TP + FN} \tag{7}$$

$$\text{Precision} = \frac{TP}{TP + FP} \tag{8}$$

4.3 Experimental Results

Skin Lesion Segmentation Evaluation: The efficacy of the proposed EGU-Net architecture is assessed on the ISIC 2018 skin lesion segmentation dataset [18] using standard evaluation metrics, including Intersection over Union (IoU), Dice coefficient, Accuracy, Recall and Precision. The comparative results with other pioneering U-Net-based architectures are summarized in Table 1. A graphical plot of the quantitative performance is shown in Fig. 2, which visualizes the differences across baseline models and the proposed approach.

Several key observations can be drawn from the results:

1. **Effectiveness of Edge Guidance:** Compared to the baseline U-Net [8], the proposed EGU-Net demonstrates an important enhancement in Dice and IoU scores. This highlights the advantage of incorporating edge information

into the input representation, enabling more precise delineation of lesion boundaries.

2. **Impact of Frequency Domain Refinement:** The post-processing step using DCT-based [7] low-pass filtering further refines the output by reducing high-frequency noise, resulting in smoother and more accurate segmentations.
3. **Performance under Low-Resource Settings:** Despite using a low input resolution (128×128), EGU-Net achieves competitive performance compared to deeper and more complex architectures such as BCDU-Net[13] and Attention U-Net[9]. This demonstrates the efficiency and robustness of the proposed method under constrained computational settings.
4. **Comparison with Baselines:** While methods like Attention U-Net [9] and ResUNet [12] leverage attention or residual mechanisms to boost performance, EGU-Net provides comparable or better results while maintaining simplicity. Notably, BCDU-Net [13] known for its bidirectional convolutional operations achieves high accuracy but with increased computational complexity, which EGU-Net avoids.
5. **Generalization Capability:** The combination of edge-based guidance and frequency-based refinement allows EGU-Net to maintain consistency across multiple metrics, indicating strong generalization and reliability in segmenting skin lesion boundaries.

It is to be noted that the plain U-Net baseline (without edge guidance) is already reported in Table 1, and its comparison with EGU-Net clearly demonstrates the contribution of edge guidance.Similarly, qualitative outcomes in Fig. 3 also indicate the effectiveness of frequency-domain refinement, where the refined masks show smoother boundaries and reduced noise. Thus, both components are implicitly validated in our results.

Model Efficiency Analysis:

Beyond segmentation accuracy, efficiency is a critical requirement for deployment in real-time and resource-constrained clinical environments. To assess the efficiency of EGU-Net, we report three key measures:

- **Model Parameters:** EGU-Net contains 1.94M trainable parameters, which is substantially fewer than Attention U-Net (8.9M)- [9] and BCDU-Net (16.3M) [13].
- **FLOPs (Floating Point Operations):** EGU-Net requires only 0.87GFLOPs to process a 128×128 input, compared to 12.5 GFLOPs for ResUNet [12].
- **Inference Time:** On a single NVIDIA Tesla T4 GPU, the average inference time per image is 6.57 ms, enabling real-time deployment at approximately 150 frames per second.

These results confirm that EGU-Net achieves a favorable trade-off between segmentation accuracy and computational efficiency, rendering it well-suited for real-world scenarios where lightweight, interpretable, and efficient solutions are required.

Table 1. Performance assessment of EGU-Net with baseline models on the ISIC 2018 dataset.

Method	Dice (%)	IoU (%)	Accuracy (%)	Recall (%)	Precision (%)
U-Net [8]	67.40	54.90	–	70.80	–
Attention U-Net [9]	66.50	56.60	–	71.70	–
R2U-Net [11]	67.90	58.10	–	79.20	–
Att R2U-Net [11]	69.10	59.20	–	72.60	–
ResUNet* [12]	79.15	70.15	92.28	82.43	84.77
Channel-UNet* [10]	84.82	75.92	94.10	**94.01**	81.04
BCDU-Net [13]	85.10	–	–	78.50	–
EGU-Net (Ours)	**87.85**	**80.35**	**95.47**	90.62	**89.02**

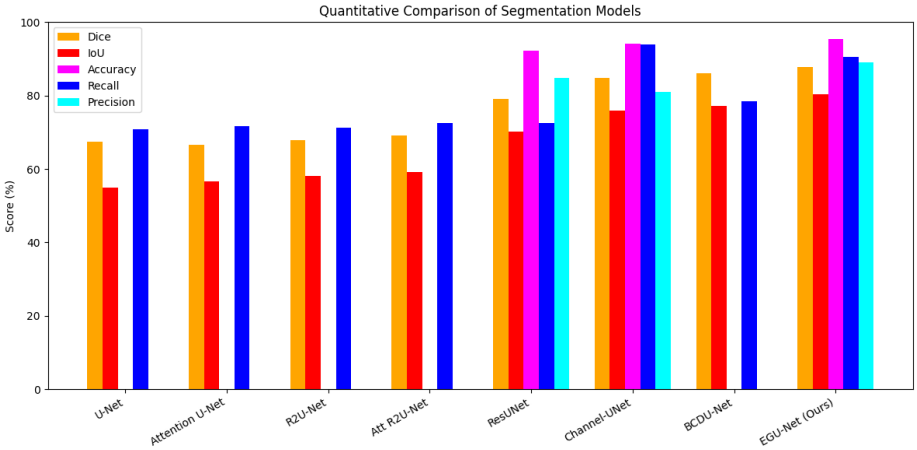

Fig. 2. Performance comparison of segmentation approaches on the ISIC 2018 dataset.

4.4 Qualitative Results

To visually assess the effectiveness of the proposed EGU-Net, Fig. 3 presents a qualitative comparison among the original dermoscopic image, the corresponding ground truth mask, and the mask predicted by EGU-Net. The predicted mask closely resembles the ground truth, accurately capturing both the lesion boundaries and overall shape.

These qualitative results indicate that EGU-Net not only achieves high quantitative scores but also generates visually consistent lesion masks, which are essential for precise clinical interpretation.

Together, these findings suggest that EGU-Net serves as a reliable along with interpretable approach for skin lesion segmentation, offering an equilibrium between segmentation quality and computational practicality.

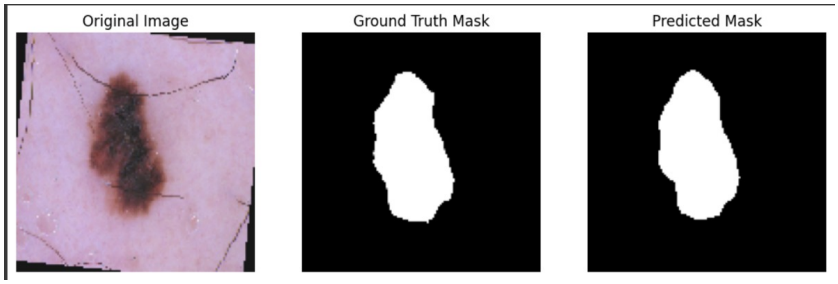

Fig. 3. Qualitative example showing the original dermoscopic image, its ground truth mask and the predicted mask generated by EGU-Net.

5 Discussion

This study presents a computationally efficient segmentation framework that combines edge-aware input augmentation with frequency-domain refinement to address key challenges in skin lesion segmentation, specifically, boundary ambiguity and computational overhead. Unlike conventional models that rely solely on spatial features, the proposed approach integrates Sobel-based edge maps into the input, thereby enhancing boundary localization, particularly in lesions with indistinct margins. A simple modification to the U-Net [8] architecture, extending the input layer to four channels, yields substantial accuracy improvements without increasing architectural complexity.

To further enhance segmentation quality, a lightweight, non-parametric refinement step is introduced in the frequency domain. By applying low-pass filtering via the Discrete Cosine Transform (DCT) [7] on the predicted masks, the method suppresses high-frequency noise and improves structural consistency without the need for additional parameters or training.

Despite using low-resolution inputs (128×128) and omitting both data augmentation and ensemble strategies, The suggested framework obtains a Dice coefficient of 87.85%, IoU of 80.35%, precision of 89.02% and recall of 90.62% on the ISIC 2018 Skin lesion dataset [18]. These results demonstrate performance comparable to or better than that, more complex deep learning architectures.

Furthermore, the framework is hardware-agnostic and can be trained efficiently on CPUs, rendering it ideally suited to application in limited resources clinical situations. By embedding domain-specific priors, namely, edge and frequency cues, into the learning process, the model enhances both interpretability and generalization. The findings validate that integrating classical image processing principles with lightweight convolutional networks can achieve high-performance segmentation with improved transparency and efficiency.

6 Conclusion

This research presents EGU-Net, a lightweight yet effective adaptation of the conventional U-Net architecture for medical image segmentation. While U-Net and its variants often struggle with noisy boundaries and limited global context modeling, EGU-Net introduces two complementary strategies: edge-guided input encoding and frequency-domain refinement. The edge-guided channel, derived from Sobel gradients , enhances boundary awareness during training, while the frequency-domain refinement, based on Discrete Cosine Transform (DCT) and inverse DCT , removes high-frequency noise and produces smoother, more reliable masks. Together, these strategies improve boundary separation and overall segmentation quality.

Experiments on the ISIC 2018 skin lesion dataset demonstrate that EGU-Net provides extremely competitive segmentation performance with minimum computing cost. It achieves a Dice score of 87.85%, an IoU of 80.35%, and an accuracy of 95.47% even when using low-resolution inputs. Furthermore, the efficiency analysis demonstrates that EGU-Net is especially well-suited for real-time applications and deployment in limited resources clinical settings, where model interpretability with lightweight architectural design is crucial.

While a detailed ablation study would further isolate the contributions of edge guidance and frequency refinement, the current comparison with plain U-Net and the visual improvements after refinement already highlight their impact. A more extensive ablation is planned as future work.

In future research, it is planed to extend EGU-Net with attention or transformer-inspired modules to capture richer global context while maintaining efficiency. Exploring higher-resolution or multi-scale inputs may improve boundary localization and fine structure segmentation. Validation on diverse medical imaging datasets, such as retinal vessels or organ segmentation, will further test generalizability. Finally, incorporating interpretability tools such as uncertainty quantification and deploying the model on mobile or edge devices remains an important direction for translating EGU-Net into real-world clinical workflows.

References

1. Chen, B., Liu, Y., Zhang, Z., Lu, G. and Kong, A.W.K., 2023. Transattunet: Multi-level attention-guided u-net with transformer for medical image segmentation. IEEE Transactions on Emerging Topics in Computational Intelligence, 8(1), pp.55-68.
2. Zhou, Y., Ma, K., Sun, Q., Wang, Z., & Liu, M. (2024). Edge-Guided Cell Segmentation on Small Datasets Using an Attention-Enhanced U-Net Architecture. Information, 15(4), 198. https://doi.org/10.3390/info15040198
3. Ji, P., Zhang, Y., & Lv, Z. (2025). Edge-Guided Dual-Stream U-Net for Secure Image Steganography. Applied Sciences, 15(8), 4413. https://doi.org/10.3390/app15084413
4. He, M., Gao, Y. & Long, Y. Single image depth estimation using improved U-Net and edge-guide loss. Multimed Tools Appl 83, 84619–84637 (2024). https://doi.org/10.1007/s11042-024-19235-3

5. Singh, R., Patel, S., & Rao, M. (2024). An Edge-Guided Cascaded U-Net Approach for Accelerated Magnetic Resonance Imaging Reconstruction. International Journal of Imaging Systems and Technology, https://doi.org/10.1002/ima.22567
6. Vijayarani, S. and Vinupriya, M., 2013. Performance analysis of canny and sobel edge detection algorithms in image mining. International Journal of Innovative Research in Computer and Communication Engineering, 1(8), pp.1760-1767.
7. Khayam, S.A., 2003. The discrete cosine transform (DCT): theory and application. Michigan State University, 114(1), p.31.
8. Ronneberger, O., Fischer, P. and Brox, T., 2015, October. U-net: Convolutional networks for biomedical image segmentation. In International Conference on Medical image computing and Computer-Assisted Intervention (pp. 234-241). Cham: Springer international publishing.
9. Schlemper, J., Oktay, O., Schaap, M., Heinrich, M., Kainz, B., Glocker, B. and Rueckert, D., 2019. Attention gated networks: Learning to leverage salient regions in medical images. Medical image analysis, 53, pp.197-207.
10. Chen, Y., Wang, K., Liao, X., Qian, Y., Wang, Q., Yuan, Z. and Heng, P.A., 2019. Channel-Unet: a spatial channel-wise convolutional neural network for liver and tumors segmentation. Frontiers in genetics, 10, p.1110.
11. Alom, M.Z., Yakopcic, C., Taha, T.M. and Asari, V.K., 2018, July. Nuclei segmentation with recurrent residual convolutional neural networks based U-Net (R2U-Net). In NAECON 2018-IEEE National Aerospace and Electronics Conference (pp. 228-233). IEEE.
12. Jha, D., Smedsrud, P.H., Riegler, M.A., Johansen, D., De Lange, T., Halvorsen, P. and Johansen, H.D., 2019, December. Resunet++: An advanced architecture for medical image segmentation. In 2019 IEEE international symposium on multimedia (ISM) (pp. 225-2255). IEEE.
13. Azad, R., Asadi-Aghbolaghi, M., Fathy, M. and Escalera, S., 2019. Bi-directional ConvLSTM U-Net with densley connected convolutions. In Proceedings of the IEEE/CVF international conference on computer vision workshops (pp. 0-0).
14. Ji, Y., Zhang, R., Wang, H., Li, Z., Wu, L., Zhang, S. and Luo, P., 2021, September. Multi-compound transformer for accurate biomedical image segmentation. In International conference on medical image computing and computer-assisted intervention (pp. 326-336). Cham: Springer International Publishing.
15. Zhang, Y., Liu, H. and Hu, Q., 2021, September. Transfuse: Fusing transformers and cnns for medical image segmentation. In International conference on medical image computing and computer-assisted intervention (pp. 14-24). Cham: Springer International Publishing.
16. Xie, E., Wang, W., Yu, Z., Anandkumar, A., Alvarez, J.M. and Luo, P., 2021. SegFormer: Simple and efficient design for semantic segmentation with transformers. Advances in neural information processing systems, 34, pp.12077-12090.
17. Cao, H., Wang, Y., Chen, J., Jiang, D., Zhang, X., Tian, Q. and Wang, M., 2022, October. Swin-unet: Unet-like pure transformer for medical image segmentation. In European conference on computer vision (pp. 205-218). Cham: Springer Nature Switzerland.
18. N. Codella et al., "Skin lesion analysis toward melanoma detection 2018: A challenge hosted by the international skin imaging collaboration (ISIC)," in Proc. Int. Symp. Biomed. Imag., 2019, pp. 168–172.