Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Adaptive Frequency Routing and Uncertainty-Aware Refinement for Skin Lesion Segmentation

Qun Gu

lzgq66@163.com

Lanzhou University of Technology

Qun Niu

Lanzhou University of Technology

Research Article

Keywords:

Posted Date: November 25th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11616816/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Adaptive Frequency Routing and Uncertainty-Aware Refinement for Skin Lesion Segmentation

Qun Gu[1,*] **Qun Niu**[1]

[1] Lanzhou University of Technology, Lanzhou, China.

* Corresponding author: lzgq66@163.com

Abstract

Skin lesion boundaries are difficult to delineate when pigmentation varies, contours are irregular, or contrast with surrounding skin is weak. Frequency-aware features can help characterize these patterns, but their contributions are not always tied to the structures being segmented. We investigate task-guided frequency routing and uncertainty-based refinement in FreqConvMamba. The Adaptive Multi-band Frequency Fusion Module (AMFFM) combines spectral and contextual features using input-dependent weights. The Task-related Frequency Routing Constraint (TFRC) guides frequency selection during training using boundary spectra derived from lesion annotations. The Frequency-conditioned Uncertainty-aware Residual Refinement (FCURR) module then uses routing descriptors to calibrate decoder features and uncertainty to guide local correction of the output probability map. Evaluations cover ISIC2017, ISIC2018, and MMOTU. On 808 ISIC2018 test images, the model achieves a Dice similarity coefficient (DSC) of 89.62%, a mean intersection over union (mIoU) of 81.20%, and a 95th-percentile Hausdorff distance (HD95) of 8.5683 pixels. Compared with the recorded FreqConvMamba baseline, DSC and mIoU increase by 1.17 and 1.89 percentage points, respectively. DSC reaches 88.76% on ISIC2017 and 88.68% on MMOTU. Eight ablation configurations examine the effects of individual modules and their combinations.

Keywords Skin lesion segmentation · FreqConvMamba · Adaptive frequency routing · Frequency-aware feature fusion · Uncertainty-guided refinement

1 Introduction

Skin lesion boundaries can be difficult to identify when pigmentation varies across a lesion or its edges fade into the surrounding skin. Irregular contours, hair, and imaging artifacts add to this difficulty [1]. The ISIC2018 challenge brought these variations into focus by evaluating segmentation methods on diverse dermoscopic images [2].

U-Net established an encoder–decoder architecture with skip connections to recover spatial detail [3]. TransUNet and Swin-Unet later incorporated Transformer-based contextual modeling [4, 5], while Mamba-UNet and VM-UNet explored state space models for capturing long-range dependencies [6, 7]. These approaches improve feature representation, but ambiguous transitions between lesion and normal skin remain difficult to delineate.

Frequency-domain processing offers another view of lesion structure. Rajab combined Fourier low-pass filtering with clustering for dermoscopic segmentation [8]. Fast Fourier Convolution showed how spectral operations can support non-local feature processing [9]. More recently, FreqConvMamba integrated frequency-guided operations with convolutional and state space modeling, including learnable weighting of frequency components [10]. Building on this work, we examine whether the routing of different frequency bands can be guided more directly by the structures present in annotated lesion boundaries.

Uncertainty provides a complementary signal. Prediction uncertainty offers a way to identify regions where the model is less confident [11]. UANeT and UDBRNet have used uncertainty information for feature calibration and boundary refinement [12, 13]. Here, we investigate a related but distinct connection: whether frequency-routing descriptors can help calibrate decoder features before uncertainty-guided correction is applied to the prediction.

We extend FreqConvMamba with three components, each addressing a different part of this process:

1. **AMFFM** fuses multi-band frequency features with Mamba-based contextual representations using input-dependent weights.
2. **TFRC** uses the frequency distribution of annotated lesion boundaries to guide the learned routing weights during training, without requiring annotations at inference.
3. **FCURR** calibrates decoder channels using the routing descriptors and applies uncertainty-weighted local correction to the segmentation output. The modules are evaluated through eight ablation configurations and experiments on ISIC2017, ISIC2018, and MMOTU.

2 Related Work

2.1 Segmentation Architectures

Encoder–decoder networks remain widely used in medical image segmentation. Building on U-Net [3], UNet++ introduced nested skip pathways [14], while UNet 3+ aggregated features across scales [15]. Attention U-Net added attention gates to select relevant features [16]. UNeXt explored tokenized multilayer perceptron (MLP) operations for efficient segmentation [17]; nnU-Net instead focused on configuring preprocessing and training for individual datasets [18]. These methods address different aspects of feature extraction, recovery, and model configuration.

For long-range context, Transformer architectures use self-attention [19], and SegFormer combines a hierarchical Transformer encoder with lightweight decoding [20]. Mamba offers a different route through selective state space modeling [21]. Such developments improve contextual representations, but frequency information provides another way to characterize the texture and structural variations encountered near lesion boundaries.

2.2 Frequency-aware Segmentation

Frequency information has been used both in feature representation and in training objectives. FcaNet extends channel attention beyond global average pooling by incorporating multiple spectral components [22]. Focal Frequency Loss assigns greater weight to frequencies that are difficult to reconstruct, although it was developed for image reconstruction and synthesis rather than segmentation [23]. The former modifies feature descriptors; the latter modifies the optimization objective.

FreqConvMamba integrates frequency-guided operations with convolutional and state space modeling for medical image segmentation [10]. Its frequency weighting is already learnable. In our framework, AMFFM combines band-specific features with Mamba-based contextual representations, while TFRC introduces supervision from the frequency distribution of annotated lesion boundaries. Feature fusion and task-related routing supervision thus serve distinct purposes.

2.3 Uncertainty-aware Refinement

Low-contrast transitions between a lesion and surrounding skin often produce uncertain predictions. Tang et al. used uncertainty-guided learning and feature refinement in an end-to-end segmentation framework [24]. Kervadec et al. addressed a different issue through boundary loss for highly unbalanced segmentation [25]; their method changes the training objective rather than estimating prediction uncertainty.

UANeT combines uncertainty-aware feature calibration with boundary refinement [12]. UDBR-Net identifies uncertain regions from disagreements among multiple decoder outputs and uses them to refine organ boundaries [13]. FCURR uses the information differently: frequency-routing descriptors calibrate decoder channels, and uncertainty estimated from the probability map weights a subsequent local correction. The two operations act at separate stages without an additional trained boundary-refinement network.

3 Methodology

3.1 Overall Architecture

The architecture follows the hierarchical encoder–decoder design of FreqConvMamba (Fig. 1). Three additions act at different stages: the Adaptive Multi-band Frequency Fusion Module (AMFFM) combines frequency and contextual features; the Task-related Frequency Routing Constraint (TFRC) supervises band selection; and

Frequency-conditioned Uncertainty-aware Residual Refinement (FCURR) calibrates decoder features and adjusts uncertain predictions.

For an image $\mathbf{I} \in \mathbb{R}^{H\times W\times 3}$, the encoder extracts hierarchical representations using convolutional and state space operations. AMFFM processes these features in parallel frequency and Mamba branches. During training, TFRC compares the learned routing distribution with a target derived from annotated lesion boundaries. The decoder recovers spatial detail through skip connections, while FCURR uses routing descriptors for channel calibration and prediction uncertainty for local correction. At inference, the annotation-dependent constraint is omitted; feature fusion, channel calibration, and prediction correction remain active.

3.2 Adaptive Frequency Fusion

Each AMFFM block receives a feature map $\mathbf{X} \in \mathbb{R}^{B\times H\times W\times C}$. A normalization layer and linear projection prepare the input for the frequency and Mamba branches:

$$\begin{aligned} \mathbf{Z} &= \mathrm{Linear}(\mathrm{Norm}(\mathbf{X}))\,, \\ [\mathbf{Z}_f, \mathbf{Z}_m] &= \mathrm{Split}(\mathbf{Z}). \end{aligned} \tag{1}$$

Here B denotes batch size, H and W denote spatial dimensions, and C is the channel count. The first branch transforms the features into the Fourier domain, processes low-, middle-, and high-frequency bands, and reconstructs spatial features by inverse transformation. In parallel, the second branch uses two-dimensional selective state space modeling (SS2D):

$$\begin{aligned} (\mathbf{F}_f, \mathbf{w}, \mathbf{e}) &= \mathcal{F}_{\mathrm{band}}(\mathbf{Z}_f), \\ \mathbf{F}_m &= \mathrm{SS2D}(\mathrm{Norm}_m(\mathbf{Z}_m))\,. \end{aligned} \tag{2}$$

The branch outputs are $\mathbf{F}_f$ and $\mathbf{F}_m$. The frequency branch also returns band weights $\mathbf{w} \in \mathbb{R}^{B\times 3}$ and energy descriptors $\mathbf{e} \in \mathbb{R}^{B\times 3}$. An MLP uses a spatially pooled feature descriptor to determine the relative contribution of the two branches:

$$\boldsymbol{\alpha} = \mathrm{Softmax}(\mathrm{MLP}(\mathrm{GAP}(\mathbf{Z})))\,, \tag{3}$$

Here $\boldsymbol{\alpha} = [\alpha_f, \alpha_m]$ with $\alpha_f + \alpha_m = 1$. The branch features are weighted, concatenated, and projected before a residual update:

$$\begin{aligned} \mathbf{F}_c &= \mathrm{Concat}(\alpha_f \mathbf{F}_f, \alpha_m \mathbf{F}_m), \\ \mathbf{Y} &= \mathbf{X} + \mathrm{DropPath}[\gamma_a \odot \mathrm{Linear}(\mathbf{F}_c)]\,. \end{aligned} \tag{4}$$

The learnable channel scale γ_a regulates the update. Unlike a fixed sum of branches, this fusion responds to the input feature distribution. The routing weights and band-energy descriptors are also available to TFRC and FCURR.

3.3 Task-related Frequency Routing

Pixel-wise segmentation supervision does not explicitly constrain the frequency router. TFRC adds an auxiliary target derived from the boundary of the ground-truth lesion mask.

3.3.1 Task-related Frequency Targets

For each binary mask $\mathbf{M}_b$, a 3×3 morphological boundary operator produces $\mathbf{B}_b$. We compute its centered Fourier power spectrum as

$$\mathbf{S}_b = \left|\mathrm{FFTShift}(\mathrm{FFT2}(\mathbf{B}_b))\right|^2. \tag{5}$$

We partition the spectrum into low-, middle-, and high-frequency regions Ω_k. The energy in each region is normalized to form a three-band target distribution:

$$\begin{aligned} E_{b,k} &= \sum_{(u,v)\in\Omega_k} \mathbf{S}_b(u,v), \\ t_{b,k} &= \frac{E_{b,k} + \epsilon}{\sum_{j=1}^{3}(E_{b,j} + \epsilon)}. \end{aligned} \tag{6}$$

The constant ϵ prevents division by zero. The resulting vector $\mathbf{t}_b$ represents the relative spectral energy of the annotated boundary in the three bands.

3.3.2 Frequency-routing Supervision

The predicted band weights $\mathbf{w}_b$ are converted to a distribution by temperature-scaled softmax:

$$\mathbf{p}_b = \mathrm{Softmax}(\mathbf{w}_b / T), \tag{7}$$

We use $T = 5$. The midpoint distribution is

$$\mathbf{m}_b = \frac{\mathbf{p}_b + \mathbf{t}_b}{2}. \tag{8}$$

The TFRC loss is the Jensen–Shannon divergence between the predicted distribution and the boundary-derived target:

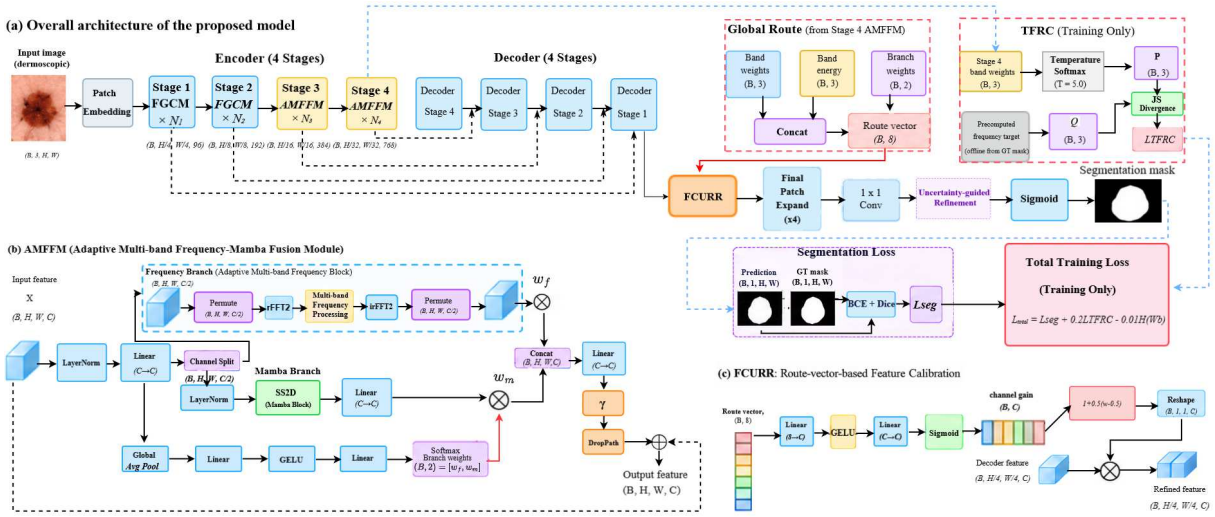

Fig. 1 Overall architecture of the proposed adaptive frequency routing and uncertainty-aware refinement framework.

$$\mathcal{L}_{\mathrm{TFRC}} = \frac{1}{2B}\sum_{b=1}^{B}\big[D_{\mathrm{KL}}(\mathbf{p}_b\|\mathbf{m}_b) + D_{\mathrm{KL}}(\mathbf{t}_b\|\mathbf{m}_b)\big]. \tag{9}$$

Here D_{KL} is the Kullback–Leibler divergence. We also regularize the original band weights through their entropy:

$$\mathcal{H}_r = -\frac{1}{B}\sum_{b=1}^{B}\sum_{k=1}^{3} w_{b,k}\log(w_{b,k}+\epsilon). \tag{10}$$

Because entropy enters the total objective with a negative coefficient, minimizing that objective discourages a routing distribution concentrated on a single band. Both terms are training-only constraints.

3.4 FCURR Refinement

FCURR operates on both decoder features and the predicted probability map. Routing information determines the feature calibration; prediction uncertainty controls the subsequent local correction.

3.4.1 Feature Calibration

Three band weights, three energy descriptors, and two branch coefficients form an eight-dimensional routing descriptor:

$$\mathbf{q} = \mathrm{Concat}(\mathbf{w}, \mathbf{e}, \boldsymbol{\alpha}) \in \mathbb{R}^{B\times 8}. \tag{11}$$

A lightweight MLP with GELU activation and a sigmoid output maps this descriptor to channel weights. These weights modulate the decoder features $\mathbf{F}_d$:

$$\begin{aligned}\mathbf{g} &= \sigma(\mathrm{MLP}(\mathbf{q}))\,,\\ \widetilde{\mathbf{F}}_d &= \mathbf{F}_d \odot [1 + 0.5(\mathbf{g} - 0.5)]\,.\end{aligned} \tag{12}$$

The vector $\mathbf{g} \in \mathbb{R}^{B\times C_d}$ has one coefficient per decoder channel. Broadcasting over the spatial dimensions preserves the feature shape; the scale factor lies between 0.75 and 1.25.

3.4.2 Uncertainty-guided Refinement

The prediction head produces segmentation logits $\mathbf{P}_c$ from the calibrated decoder features. Applying a sigmoid gives the probability map

$$\mathbf{p}_c = \sigma(\mathbf{P}_c). \tag{13}$$

We estimate prediction uncertainty using normalized binary entropy:

$$\mathbf{U} = -\frac{1}{\log 2}\big[\mathbf{p}_c\log(\mathbf{p}_c+\epsilon) + (1-\mathbf{p}_c)\log(1-\mathbf{p}_c+\epsilon)\big]. \tag{14}$$

Probabilities near 0.5 receive the highest uncertainty scores; values near 0 or 1 receive

the lowest. The uncertainty map weights a local residual correction:

$$\begin{aligned} \Delta \mathbf{p} &= \mathrm{AvgPool}_{3\times 3}(\mathbf{p}_c) - \mathbf{p}_c, \\ \mathbf{p}_f &= \mathbf{p}_c + \eta \mathbf{U} \odot \Delta \mathbf{p}. \end{aligned} \tag{15}$$

We set $\eta = 0.25$, clip the corrected probabilities to $[0, 1]$, and threshold the result to obtain the segmentation. The correction does not require an additional trained prediction head.

3.5 Overall Loss Function

The training objective contains a binary cross-entropy (BCE)–Dice segmentation term and two routing terms. We define

$$\mathcal{L}_{\mathrm{seg}} = \mathcal{L}_{\mathrm{BCE}} + \mathcal{L}_{\mathrm{Dice}}. \tag{16}$$

BCE penalizes pixel-wise errors, while Dice loss measures overlap between the prediction and reference mask. The full objective is

$$\mathcal{L}_{\mathrm{total}} = \mathcal{L}_{\mathrm{seg}} + 0.2\mathcal{L}_{\mathrm{TFRC}} - 0.01\mathcal{H}_r. \tag{17}$$

At inference, no ground-truth boundary spectra are needed. AMFFM uses input features to determine the routing weights, and FCURR retains both channel calibration and uncertainty-guided correction.

4 Experiments and Results

4.1 Experimental Setup

4.1.1 Datasets

We evaluated the framework on three datasets: ISIC2017 and ISIC2018 for dermoscopic lesion segmentation [2, 26], and MMOTU for ovarian tumor ultrasound segmentation [27]. The ultrasound data provide an additional imaging modality for examining the method. HAM10000 is another publicly available dermoscopic resource [28], but it was not included in the present evaluation.

Following FreqConvMamba [10], we used the same ISIC2018 split with 808 test images. Input images were resized to 256 × 256 pixels and normalized, while masks were resized using label-preserving interpolation. Random flips and rotations were applied during training. Validation and testing followed the preprocessing used in the corresponding evaluation runs.

4.1.2 Implementation Details

The implementation uses FreqConvMamba [10] with an initial feature dimension of 96. We trained the model using AdamW, with an initial learning rate of 1×10^{-4}, weight decay of 1×10^{-2}, a batch size of 8, and up to 200 epochs. The coefficients of TFRC and routing entropy were 0.2 and 0.01, respectively. A threshold of 0.5 was used to obtain binary predictions.

The comparison tables contain results from separate recorded evaluation runs. Their values are reported without assuming that every method was tested under an identical configuration.

4.1.3 Evaluation Metrics

We assessed region overlap using the Dice similarity coefficient (DSC) and mean intersection over union (mIoU), and boundary agreement using the 95th-percentile Hausdorff distance (HD95) [29]. Precision and recall were also reported for the multi-method comparison. For binary predictions, the overlap measures are defined as

$$\begin{aligned} \mathrm{DSC} &= \frac{2TP}{2TP + FP + FN}, \\ \mathrm{IoU} &= \frac{TP}{TP + FP + FN}, \end{aligned} \tag{18}$$

where TP, FP, and FN are the true-positive, false-positive, and false-negative pixel counts. Overlap values appear as proportions in Tables 1 and 2, and as percentages in Tables 3 and 4. HD95 is given in pixels. Higher overlap scores and lower HD95 are favorable; Jaccard-based evaluation has also been used in lesion segmentation [30].

4.2 Ablation Study

To examine the effect of each addition, we compared eight configurations on ISIC2018: the backbone, three individual modules, their three pairwise combinations, and the complete framework (Table 1).

4.2.1 Effect of Individual Modules

Each individual module increases DSC relative to the backbone. With AMFFM alone, DSC rises from 0.884590 to 0.887422 and HD95 decreases from 9.8877 to 9.5041 pixels. TFRC-only and FCURR-only reach DSC values of 0.886847 and 0.887210. These modest but consistent DSC gains

Table 1 Ablation results for AMFFM, TFRC, and FCURR on ISIC2018.

Method	AMFFM	TFRC	FCURR	mIoU ↑	DSC ↑	HD95 ↓
Baseline	×	×	×	0.793060	0.884590	9.8877
AMFFM-only	✓	×	×	0.797627	0.887422	9.5041
TFRC-only	×	✓	×	0.796698	0.886847	9.4692
FCURR-only	×	×	✓	0.797290	0.887210	9.5772
AMFFM + TFRC	✓	✓	×	0.808245	0.893955	8.7026
AMFFM + FCURR	✓	×	✓	0.799621	0.888655	9.5976
FCURR + TFRC	×	✓	✓	0.804945	0.891933	9.1965
Ours (Complete Model)	✓	✓	✓	**0.811995**	**0.896244**	**8.568283**

Note: ✓/× indicates whether a module is enabled/disabled. The table consolidates available progressive and supplementary ablation records; the complete-model result is from the confirmed final evaluation.

indicate that the components provide benefits even when used separately.

4.2.2 Effect of Module Combinations

The pairwise combinations reveal a less uniform pattern. AMFFM + TFRC performs best among them, with a DSC of 0.893955 and an HD95 of 8.7026 pixels. FCURR + TFRC reaches a DSC of 0.891933. By contrast, AMFFM + FCURR reaches 0.888655, and its HD95 of 9.5976 pixels is slightly worse than either corresponding single-module result. The combination of modules therefore matters more than their number alone.

4.2.3 Complete Framework

With all three modules enabled, the model obtains an mIoU of 0.811995, a DSC of 0.896244, and an HD95 of 8.568283 pixels, the best values among the eight configurations. Relative to the backbone, the gains are 0.018935 in mIoU and 0.011654 in DSC, while HD95 falls by 1.319417 pixels. This result is stronger than any partial configuration, although the pairwise experiments show that the contributions are not simply additive.

4.3 Comparison with Existing Methods

Table 2 presents the recorded results of representative CNN-, Transformer-, and Mamba-based methods on ISIC2018. Our method has the highest mIoU (0.8120), DSC (0.8962), and precision (0.9146) in this comparison. Its DSC exceeds that of TransUNet by only 0.0011. TransUNet, however, obtains higher recall (0.8891 versus 0.8786) and lower HD95 (8.3923 versus 8.5683 pixels).

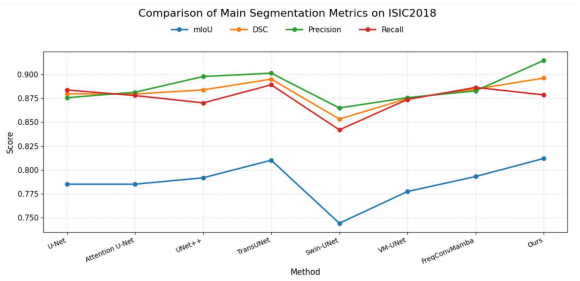

Fig. 2 Comparison of mIoU, DSC, precision, and recall on ISIC2018.

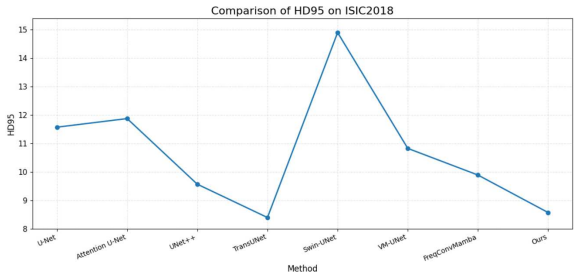

Fig. 3 Comparison of HD95 on ISIC2018; lower values are better.

The results show that stronger region overlap does not necessarily coincide with better boundary-distance performance. Comparisons should also be read in light of the separate evaluation records.

The trends in region-overlap and boundary metrics are also shown in Figs. 2 and 3.

4.4 Qualitative Visualization

Figure 4 shows six ISIC2018 images with different lesion shapes and boundary contrast. Green

Table 2 Recorded quantitative comparison with representative methods on ISIC2018.

Method	mIoU ↑	DSC ↑	Precision ↑	Recall ↑	HD95 ↓
U-Net	0.7851	0.8796	0.8756	0.8837	11.5714
Attention U-Net	0.7851	0.8796	0.8814	0.8779	11.8738
UNet++	0.7918	0.8838	0.8978	0.8702	9.5647
TransUNet	0.8102	0.8951	0.9013	**0.8891**	**8.3923**
Swin-UNet	0.7442	0.8533	0.8649	0.8420	14.9013
VM-UNet	0.7774	0.8747	0.8757	0.8738	10.8251
FreqConvMamba	0.7931	0.8846	0.8828	0.8864	9.8877
Ours	**0.8120**	**0.8962**	**0.9146**	0.8786	8.5683

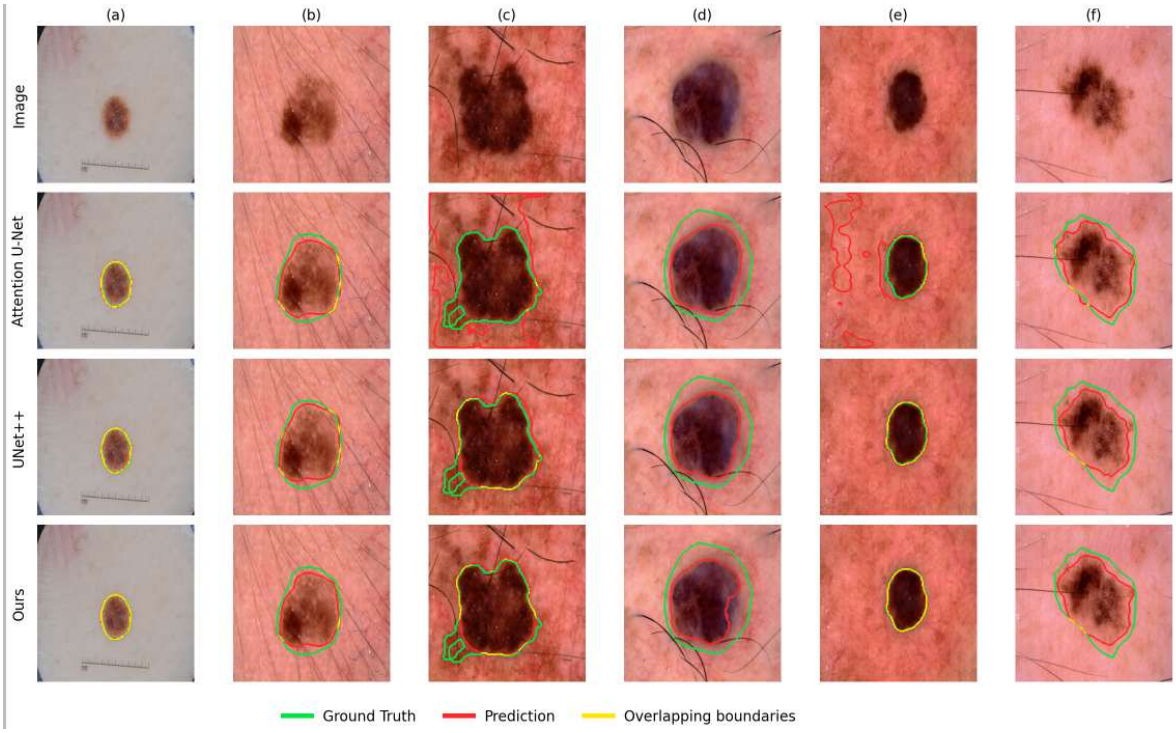

Fig. 4 Boundary comparisons on six ISIC2018 images using Attention U-Net, UNet++, and the proposed method. Green, red, and yellow denote ground-truth-only, prediction-only, and overlapping boundaries, respectively.

Table 3 Recorded FreqConvMamba baseline and proposed-method results on three datasets.

Dataset	Method	mIoU (%) ↑	DSC (%) ↑	HD95 (pixel) ↓
ISIC2017	FreqConvMamba	76.85	86.91	10.7366
	Ours	**79.78**	**88.76**	**8.5687**
ISIC2018	FreqConvMamba	79.31	88.46	9.8877
	Ours	**81.20**	**89.62**	**8.5683**
MMOTU	FreqConvMamba	78.25	87.80	15.4184
	Ours	**79.67**	**88.68**	**14.6692**

contours mark ground-truth-only regions, red contours mark prediction-only regions, and yellow indicates boundary overlap. Differences among the

Table 4 Historical multi-method comparison on ISIC2017 and MMOTU. DSC and mIoU are reported as percentages.

Method	ISIC2017		MMOTU	
	DSC (%) ↑	mIoU (%) ↑	DSC (%) ↑	mIoU (%) ↑
U-Net	82.85	73.10	82.44	73.78
UNet++	82.69	72.99	82.65	74.61
Attention U-Net	83.04	73.44	83.16	75.55
MedT	85.99	77.94	79.88	70.25
TransUNet	83.27	73.54	85.05	77.02
HiFormer	87.42	78.42	86.59	78.61
VM-UNet	87.62	77.97	84.32	72.89
SkinMamba	87.87	78.58	86.91	76.85
FreqConvMamba	88.37	79.01	88.02	79.10
Ours	**88.76**	**79.78**	**88.68**	**79.67**

Table 5 Recorded model complexity and inference throughput.

Method	Params (M)	FLOPs (G)	FPS
FreqConvMamba	16.33	7.61	55.15
Ours	23.19	8.22	38.97

methods are most visible around weak or irregular contours in these selected examples. These images help interpret the boundary errors, but the quantitative comparison remains based on the full test set.

4.5 Multi-dataset Evaluation

The additional evaluations are summarized in two tables. Table 3 compares our framework with FreqConvMamba on ISIC2017, ISIC2018, and MMOTU. Table 4 reports a separate series of comparisons on ISIC2017 and MMOTU, including MedT [31], HiFormer [32], and SkinMamba [33]. These are different evaluation records: notably, the FreqConvMamba entries in the two tables are not identical.

In Table 3, DSC increases by 1.85, 1.16, and 0.88 percentage points on ISIC2017, ISIC2018, and MMOTU, respectively. The mIoU gains are 2.93, 1.89, and 1.42 percentage points, with lower HD95 on all three datasets. These comparisons show improvement relative to their recorded baselines; they do not establish that the same gain would hold under a common evaluation protocol.

For the separate series in Table 4, our method records the highest DSC and mIoU on both datasets. Because the baseline entries differ from those in Table 3, we treat the two tables as separate comparisons rather than combine them into one ranking.

4.6 Computational Complexity

Table 5 shows the cost of the additional modules. Parameters rise from 16.33 to 23.19 million and floating-point operations (FLOPs) from 7.61 to 8.22 G. Recorded throughput falls from 55.15 to 38.97 frames per second (FPS). The improved segmentation scores therefore come with a higher model cost. A controlled speed comparison would require the same hardware, input size, and batch size.

4.7 Discussion

Several observations emerge from these experiments. First, AMFFM and TFRC are more effective together than either is alone, whereas pairing AMFFM with FCURR does not consistently reduce boundary error. The complete model still performs best in the ablation study, suggesting that the three mechanisms work most effectively in the full configuration. This interpretation is limited to the recorded experiments and does not imply that every module interaction is favorable.

Second, the multi-method results separate overlap quality from boundary accuracy: our

model leads in DSC and mIoU, while TransUNet has a slightly lower HD95. The additional datasets broaden the assessment but do not constitute a single matched benchmark. The increase in parameters and the lower recorded FPS also make efficiency a practical direction for further work.

5 Conclusion

We examined whether boundary-derived supervision could make frequency routing more useful for skin lesion segmentation. Starting from FreqConvMamba, AMFFM combines multi-band frequency features with contextual representations, and TFRC guides the routing weights using the spectra of annotated lesion boundaries. FCURR then uses the learned routing descriptors to calibrate decoder features and applies uncertainty-weighted local correction to the predicted probability map.

On the ISIC2018 test set of 808 images, the complete model achieved a DSC of 0.8962, an mIoU of 0.8120, and an HD95 of 8.5683 pixels. The eight ablation configurations showed that gains were not uniformly additive: AMFFM and TFRC formed the strongest two-module variant, while the complete model performed best overall. The recorded DSC values were 88.76% for ISIC2017 and 88.68% for MMOTU.

The performance gains came with more parameters and lower recorded inference throughput. In addition, the available cross-dataset comparisons were collected under differing evaluation settings, limiting strict claims about performance across datasets. Future work should reduce the computational cost and assess the method under matched evaluation protocols.

Statements and Declarations

Author contributions. Qun Niu developed the methodology, implemented the model, conducted the experiments, analyzed the data, and drafted the manuscript. Qun Gu contributed the initial research idea and provided academic guidance. Both authors reviewed and approved the final manuscript.

Funding. No funding was received for conducting this study.

Data availability. The ISIC2017 and ISIC2018 datasets are publicly available through the International Skin Imaging Collaboration (ISIC) Challenge data portal (ISIC2017; ISIC2018). The MMOTU dataset is available through the original authors' public repository.

Competing interests. The authors declare that they have no competing interests.

References

[1] Mirikharaji, Z., Abhishek, K., Bissoto, A., Barata, C., Avila, S., Valle, E., Celebi, M.E., Hamarneh, G.: A survey on deep learning for skin lesion segmentation. Medical Image Analysis **88**, 102863 (2023) https://doi.org/10.1016/j.media.2023.102863

[2] Codella, N.C.F., Rotemberg, V., Tschandl, P., Celebi, M.E., Dusza, S., Gutman, D., Helba, B., Kalloo, A., Liopyris, K., Marchetti, M., Kittler, H., Halpern, A.: Skin Lesion Analysis Toward Melanoma Detection 2018: A Challenge Hosted by the International Skin Imaging Collaboration (ISIC) (2019). https://doi.org/10.48550/arXiv.1902.03368

[3] Ronneberger, O., Fischer, P., Brox, T.: U-net: Convolutional networks for biomedical image segmentation. In: Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015. Lecture Notes in Computer Science, vol. 9351, pp. 234–241. Springer, Cham (2015). https://doi.org/10.1007/978-3-319-24574-4_28

[4] Chen, J., Lu, Y., Yu, Q., Luo, X., Adeli, E., Wang, Y., Lu, L., Yuille, A.L., Zhou, Y.: TransUNet: Transformers Make Strong Encoders for Medical Image Segmentation (2021). https://doi.org/10.48550/arXiv.2102.04306

[5] Cao, H., Wang, Y., Chen, J., Jiang, D., Zhang, X., Tian, Q., Wang, M.: Swin-unet: Unet-like pure transformer for medical image segmentation. In: Computer Vision – ECCV 2022 Workshops. Lecture Notes in Computer Science, vol. 13803, pp. 205–218. Springer, Cham (2023). https://doi.org/10.1007/978-3-031-25066-8_9

[6] Wang, Z., Zheng, J.-Q., Zhang, Y., Cui, G., Li, L.: Mamba-UNet: UNet-Like Pure Visual Mamba for Medical Image Segmentation (2024). https://arxiv.org/abs/2402.05079

[7] Ruan, J., Li, J., Xiang, S.: VM-UNet: Vision Mamba UNet for Medical Image Segmentation (2024). https://arxiv.org/abs/2402.02491

[8] Rajab, M.I.: Segmentation of dermatoscopic images by frequency domain filtering and k-means clustering algorithms. Skin Research and Technology **17**(4), 469–478 (2011) https://doi.org/10.1111/j.1600-0846.2011.00520.x

[9] Chi, L., Jiang, B., Mu, Y.: Fast fourier convolution. In: Advances in Neural Information Processing Systems, vol. 33, pp. 4479–4488 (2020)

[10] Song, Y., Dou, W., Qian, Y., Jia, J., Zhu, Z.: Freqconvmamba: Frequency-guided hierarchical hybrid ssm-cnn for medical image segmentation. Medical Image Analysis **111**, 104019 (2026) https://doi.org/10.1016/j.media.2026.104019

[11] Kendall, A., Gal, Y.: What uncertainties do we need in bayesian deep learning for computer vision? In: Advances in Neural Information Processing Systems, vol. 30, pp. 5574–5584 (2017)
[12] Zhang, Q., Wang, J., Wu, W., Zheng, Y.: Uanet: An uncertainty-aware network for feature calibration and boundary refinement in medical image segmentation. Biomedical Signal Processing and Control **112**, 108413 (2026) https://doi.org/10.1016/j.bspc.2025.108413
[13] Hassan, R., Mondal, M.R.H., Ahamed, S.I.: Udbrnet: A novel uncertainty driven boundary refined network for organ at risk segmentation. PLOS ONE **19**(6), 0304771 (2024) https://doi.org/10.1371/journal.pone.0304771
[14] Zhou, Z., Siddiquee, M.M.R., Tajbakhsh, N., Liang, J.: Unet++: A nested u-net architecture for medical image segmentation. In: Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support vol. 11045, pp. 3–11. Springer, Cham (2018). https://doi.org/10.1007/978-3-030-00889-5_1
[15] Huang, H., Lin, L., Tong, R., Hu, H., Zhang, Q., Iwamoto, Y., Han, X., Chen, Y.-W., Wu, J.: Unet 3+: A full-scale connected unet for medical image segmentation. In: 2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 1055–1059. IEEE, Barcelona, Spain (2020). https://doi.org/10.1109/ICASSP40776.2020.9053405
[16] Oktay, O., Schlemper, J., Le Folgoc, L., Lee, M., Heinrich, M., Misawa, K., Mori, K., McDonagh, S., Hammerla, N.Y., Kainz, B., Glocker, B., Rueckert, D.: Attention U-Net: Learning Where to Look for the Pancreas (2018). https://doi.org/10.48550/arXiv.1804.03999
[17] Valanarasu, J.M.J., Patel, V.M.: Unext: Mlp-based rapid medical image segmentation network. In: Medical Image Computing and Computer-Assisted Intervention – MICCAI 2022, pp. 23–33. Springer, Cham (2022). https://doi.org/10.1007/978-3-031-16443-9_3
[18] Isensee, F., Jaeger, P.F., Kohl, S.A.A., Petersen, J., Maier-Hein, K.H.: nnu-net: A self-configuring method for deep learning-based biomedical image segmentation. Nature Methods **18**(2), 203–211 (2021) https://doi.org/10.1038/s41592-020-01008-z
[19] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, L., Polosukhin, I.: Attention is all you need. In: Advances in Neural Information Processing Systems, vol. 30, pp. 5998–6008 (2017)
[20] Xie, E., Wang, W., Yu, Z., Anandkumar, A., Alvarez, J.M., Luo, P.: Segformer: Simple and efficient design for semantic segmentation with transformers. In: Advances in Neural Information Processing Systems, vol. 34, pp. 12077–12090 (2021)
[21] Gu, A., Dao, T.: Mamba: Linear-Time Sequence Modeling with Selective State Spaces (2023). https://doi.org/10.48550/arXiv.2312.00752
[22] Qin, Z., Zhang, P., Wu, F., Li, X.: Fcanet: Frequency channel attention networks. In: Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), pp. 783–792 (2021). https://doi.org/10.1109/ICCV48922.2021.00082
[23] Jiang, L., Dai, B., Wu, W., Loy, C.C.: Focal frequency loss for image reconstruction and synthesis. In: Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), pp. 13919–13929 (2021). https://doi.org/10.1109/ICCV48922.2021.01366
[24] Tang, P., Yang, P., Nie, D., Wu, X., Zhou, J., Wang, Y.: Unified medical image segmentation by learning from uncertainty in an end-to-end manner. Knowledge-Based Systems **241**, 108215 (2022) https://doi.org/10.1016/j.knosys.2022.108215
[25] Kervadec, H., Bouchtiba, J., Desrosiers, C., Granger, E., Dolz, J., Ben Ayed, I.: Boundary loss for highly unbalanced segmentation. Medical Image Analysis **67**, 101851 (2021) https://doi.org/10.1016/j.media.2020.101851
[26] Codella, N.C.F., Gutman, D.A., Celebi, M.E., Helba, B., Marchetti, M.A., Dusza, S.W., Kalloo, A., Liopyris, K., Mishra, N., Kittler, H., Halpern, A.: Skin lesion analysis toward melanoma detection: A challenge at the 2017 international symposium on biomedical imaging (isbi), hosted by the international skin imaging collaboration (isic). In: 2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI), pp. 168–172. IEEE, Washington, DC, USA (2018). https://doi.org/10.1109/ISBI.2018.8363547
[27] Lyu, S., Zhao, Q., Bai, W., Cai, L., Cheng, G., Cui, G., Yang, M., Chen, L., Zhou, H.: Unsupervised cross-domain semantic segmentation on multimodality ovarian tumor ultrasound data. Pattern Recognition **171**, 112311 (2026) https://doi.org/10.1016/j.patcog.2025.112311
[28] Tschandl, P., Rosendahl, C., Kittler, H.: The ham10000 dataset, a large collection of multi-source dermatoscopic images of common pigmented skin lesions. Scientific Data **5**, 180161 (2018) https://doi.org/10.1038/sdata.2018.161
[29] Taha, A.A., Hanbury, A.: Metrics for evaluating 3d medical image segmentation: Analysis, selection, and tool. BMC Medical Imaging **15**, 29 (2015) https://doi.org/10.1186/s12880-015-0068-x
[30] Yuan, Y., Chao, M., Lo, Y.-C.: Automatic skin lesion segmentation using deep fully convolutional networks with jaccard distance. IEEE Transactions on Medical Imaging **36**(9), 1876–1886 (2017) https://doi.org/10.1109/TMI.2017.2695227
[31] Valanarasu, J.M.J., Oza, P., Hacihaliloglu, I., Patel, V.M.: Medical transformer: Gated axial-attention for medical image segmentation. In: Medical Image Computing and Computer Assisted Intervention – MICCAI 2021. Lecture Notes in Computer Science, vol. 12901, pp. 36–46. Springer, Cham (2021). https://doi.org/10.1007/978-3-030-87193-2_4
[32] Heidari, M., Kazerouni, A., Soltany, M., Azad, R., Aghdam, E.K., Cohen-Adad, J., Merhof, D.: HiFormer: Hierarchical multi-scale representations using transformers for medical image segmentation. In: Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, pp. 6202–6212 (2023)
[33] Zou, S., Zhang, M., Fan, B., Zhou, Z., Zou, X.: SkinMamba: A Precision Skin Lesion Segmentation Architecture with Cross-Scale Global State Modeling and Frequency Boundary Guidance. arXiv preprint arXiv:2409.10890 (2024). https://doi.org/10.48550/arXiv.2409.10890