FreqBlueprint: Frequency-Bank Guided Dynamic Graph Convolution for Medical Image Segmentation

Anonymous submission

Abstract

In medical images, lesions are often obscured by complex background textures and ambiguous anatomical boundaries. To derive accurate segmentation boundaries from such challenging visual patterns, we move beyond the one-size-fits-all design of fixed convolutional kernels in conventional convolutional neural networks and propose a frequency-bank-based dual-domain graph convolutional architecture, termed FD-GCN. Inspired by the diagnostic workflow of pathologists, who first examine the overall tissue organization at low magnification and then inspect fine cellular morphology at high magnification, our model establishes an instruction-guided modular assembly mechanism. Specifically, we design a Dynamic Graph Frequency Bank (DGFB) that adaptively generates spatial- and frequency-domain weights from global contextual information. These weights serve as guidance signals for Frequency-Guided Scale-Adaptive Grid Graph Convolution (FSGC) to dynamically assemble multi-scale graph neighborhoods, thereby enlarging the effective receptive field and enhancing structural modeling while maintaining moderate computational complexity. In addition, Spatial-Frequency Fusion Attention (SFFA) and Adaptive Cross-stage Feature Fusion (ACFF) are introduced to promote multi-scale dual-domain feature interaction and alleviate the semantic discrepancy between the encoder and decoder. FD-GCN achieves a Dice score of 85.72% on the Synapse multi-organ segmentation dataset and an IoU of 86.26% on the ISIC2016 dataset, outperforming existing state-of-the-art methods.

Introduction

Medical image segmentation is essential for computer-aided diagnosis, yet accurate segmentation remains challenging due to complex anatomical structures, large intra-class variations, and ambiguous boundaries. Existing CNN-based methods(Alzubaidi et al. 2021) are limited by local receptive fields, while Transformer-based approaches suffer from high computational costs and sensitivity to high-frequency noise. Although graph convolutional networks (GCNs)(Li et al. 2019) provide effective structural modeling capabilities, most existing methods rely on fixed graph structures and primarily focus on spatial-domain information(Li et al. 2017), neglecting the complementary role of frequency-domain features in texture enhancement and noise suppression(Sun et al. 2019).

To address these limitations, we propose a Frequency Dynamic Graph Convolutional Network (FD-GCN), which establishes a spatial-frequency collaborative framework for medical image segmentation. Specifically, we design a Dynamic Graph Frequency Bank (DGFB) to generate spatial-frequency dual-domain weights from global contextual information. These weights serve as guidance for Frequency-Guided Spatial Graph Convolution (FSGC), enabling adaptive graph convolution construction with enhanced receptive fields and topology modeling capability. Furthermore, Spatial-Frequency Fusion Attention (SFFA) and Adaptive Cross-stage Feature Fusion (ACFF) are introduced to improve dual-domain feature interaction and alleviate encoder-decoder semantic gaps. FD-GCN provides an effective solution for integrating spatial semantic information and frequency-aware representations, achieving accurate segmentation of complex anatomical structures.

We conduct extensive experiments on the standard Synapse multi-organ abdominal segmentation dataset. Experimental results demonstrate that FD-GCN significantly outperforms existing state-of-the-art methods, particularly in challenging scenarios involving blurred boundaries and small target organs. The main contributions of this work are summarized as follows:

- To address the limitation of fixed feature-processing strategies, we propose a Dynamic Graph Frequency Bank (DGFB) guided FSGC assembly mechanism. DGFB generates adaptive spatial-frequency weights from global context, which serve as assembly instructions to dynamically aggregate multi-scale graph convolutional branches in FSGC, constructing input-adaptive graph representations.
- To jointly model anatomical structures and fine-grained details, we design a dual-domain graph convolution module. The spatial branch captures structural continuity and long-range dependencies, while the frequency branch enhances boundary and texture representations through frequency-aware feature refinement.
- To alleviate semantic discrepancies during dual-domain and multi-level feature fusion, we introduce a collabora-

This is an anonymized submission for review purposes only. Distribution, citation, or public sharing of this manuscript is strictly prohibited. Copyright and publication details will appear in the final version if accepted.

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

tive fusion mechanism consisting of SFFA and ACFF. SFFA enables adaptive spatial-frequency interaction, while ACFF aligns high-level semantic information with low-level detailed features for improved feature reconstruction.

Related Work

Deep Learning for Medical Image Segmentation

Deep learning has significantly advanced medical image segmentation by enabling end-to-end pixel-wise prediction(Nema et al. 2017). Among early approaches, U-Net(Ronneberger, Fischer, and Brox 2015) established a fundamental encoder-decoder paradigm through skip connections, effectively combining semantic information and spatial details, and has become a widely adopted framework for medical image analysis(Han et al. 2021). However, CNN-based methods remain limited by local receptive fields, making it difficult to capture long-range dependencies and global structural context, especially for organs with complex morphology and ambiguous boundaries.

To overcome these limitations, Transformer-based models such as TransUNet(Chen et al. 2021) and Swin-Unet(Cao et al. 2022) have been introduced to enhance global representation through self-attention mechanisms. Nevertheless, their high computational cost and insufficient local inductive bias often restrict fine-grained boundary segmentation. Hybrid architectures, such as MaxViT(Tu et al. 2022), attempt to combine local convolution and global attention, but achieving an effective balance between semantic consistency and detail preservation remains challenging, particularly in abdominal organ segmentation with large intra-class variations and complex topology(Wang, Sun, and Wang 2024).

Motivated by these limitations, we propose a Dynamic Graph Frequency Bank and a dual-domain graph convolution mechanism to establish an adaptive spatial-frequency representation framework for medical image segmentation.

Graph Neural Networks in Vision

Graph neural networks (GNNs)(Corso et al. 2024) have attracted increasing attention in computer vision(Voulodimos et al. 2018) due to their capability of modeling non-Euclidean structures. In medical image segmentation, GNNs can capture complex spatial relationships and anatomical dependencies beyond conventional convolution operations(Peng et al. 2020). However, existing GCN-based methods typically rely on predefined graph structures or fixed receptive fields, limiting their adaptability to content variations. Moreover, spatial-domain-only graph reasoning may suffer from over-smoothing and insufficient preservation of high-frequency details, such as small structures and ambiguous boundaries(Huang et al. 2017).

Recent dynamic graph and graph attention(Velickovic et al. 2017) methods improve adaptability by learning input-dependent relationships(Meng et al. 2021), but they often introduce additional computational costs and still neglect the potential of frequency-domain guidance. To address these limitations, FD-GCN introduces a Dynamic Graph Frequency Bank (DGFB) and Frequency-Guided Spatial Graph Convolution (FSGC), where dual-domain weights are used to dynamically guide multi-scale graph convolution. This design enables adaptive feature aggregation while preserving both global structural information and fine-grained boundary details.

Frequency Domain Learning and Dual-Domain Fusion

Frequency-domain analysis has been increasingly explored in computer vision to enhance texture representation and exploit global image patterns. Transformations such as DWT(Alessio 2015) and FFT(Heckbert 1995) decompose images into low-frequency(van Haarlem et al. 2013) structural information and high-frequency(Engel Jr and da Silva 2012) details, providing complementary representations for feature learning. Existing frequency-based methods, such as Wavelet-Unet(Agnes, Solomon, and Karthick 2024) and FrequencyNet(Binder, Krauss, and Wiesner 2020), demonstrate the effectiveness of frequency information in preserving details and suppressing noise. However, frequency features are typically treated as fixed preprocessing results or auxiliary branches, while dynamic spatial-frequency interaction remains insufficiently explored.

To address this limitation, FD-GCN introduces a Dynamic Graph Frequency Bank (DGFB), which adaptively generates spatial-frequency weights according to input features. These weights act as an "instruction set" to guide FSGC in dynamically aggregating multi-scale convolutional branches, enabling adaptive feature representation and improved boundary preservation. Furthermore, SFFA facilitates complementary fusion between spatial and frequency features, allowing the model to better balance structural information and texture details.

Methodology

Overall Architecture and Frequency-Guided Decoding Flow

As illustrated in Figure 1, FD-GCN adopts an encoder-decoder framework with MaxViT as the backbone. Unlike conventional decoders, it introduces a DGFB-guided dual-domain graph convolution module to jointly model spatial structures and frequency details.

Specifically, ACFF first aligns multi-level features, after which DGFB generates spatial-frequency weights to guide FSGC for adaptive graph convolution. The enhanced dual-domain features are then fused by SFFA for final prediction. Through DGFB-driven feature scheduling, FD-GCN effectively improves segmentation of blurred boundaries, small targets, and complex textures.

Dynamic Graph Frequency Bank (DGFB)

The architecture of the Dynamic Graph Frequency Bank (DGFB) is illustrated in Figure 1. As the strategy generation unit of GCN, DGFB distinguishes FD-GCN from conventional multi-scale convolution and spatial-frequency fusion methods. It generates adaptive weights from global context

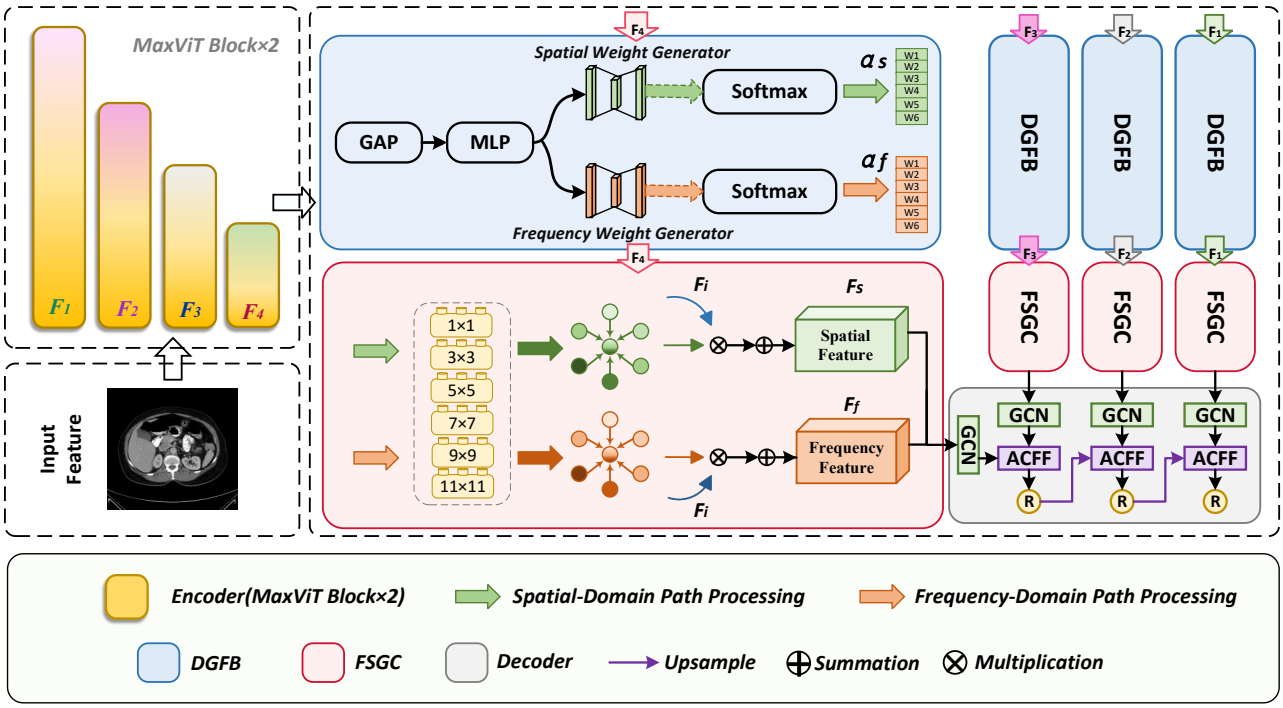

Figure 1: Overall framework of the proposed FD-GCN.

to guide feature processing, serving as a "convolutional assembly instruction" for FSGC and enabling input-adaptive graph convolution.

Global Context Awareness Given an input feature map F, DGFB first applies global average pooling to obtain a channel-wise descriptor z, which captures the global context of the feature map. Then, z is fed into a lightweight MLP for feature compression and nonlinear projection, producing the latent representation q:

$$z = \mathrm{GAP}(F) \tag{1}$$

$$q = \mathrm{MLP}(z) \tag{2}$$

Through this process, DGFB can perceive the structural complexity, semantic distribution, and potential scale requirements of the current features from a global perspective, thereby providing a basis for the subsequent generation of dual-domain weights.

Spatial-Frequency Processing Weight Generation Based on the global contextual representation q, DGFB employs two independent branches to generate spatial and frequency weights, respectively:

$$\alpha_s = \mathrm{Softmax}(g_s(q)) \tag{3}$$

$$\alpha_f = \mathrm{Softmax}(g_f(q)) \tag{4}$$

Here, α_s and α_f denote the spatial and frequency weights, respectively. Generated by independent branches from the same input feature, they enable different scale preferences for spatial modeling and frequency enhancement, providing flexible dual-domain adaptation.

Frequency-Guided Scale-Adaptive Grid Graph Convolution (FSGC)

After obtaining the spatial- and frequency-domain weights, DGFB guides FSGC to adaptively assemble convolutional operators. Given an input feature map F, each spatial position is regarded as a node on a regular grid, while the region covered by a convolutional kernel defines its local neighborhood. Therefore, kernels of different sizes correspond to grid-graph operators with different connection ranges.

FSGC contains multiple fixed-scale branches, including $1\times1, 3\times3, 5\times5, 7\times7, 9\times9$, and 11×11 kernels, providing candidate neighborhoods from local details to broader context:

$$\mathcal{K} = \{1, 3, 5, 7, 9, 11\}. \tag{5}$$

$$\mathbf{B}_k = \mathrm{Conv}_{k\times k}(F), \qquad k \in \mathcal{K}, \tag{6}$$

where $\mathbf{B}_k$ denotes the response of the k-scale grid neighborhood. The weights generated by DGFB serve as input-dependent instructions to aggregate these candidate responses:

$$\mathrm{FSGC}(F, \boldsymbol{\alpha}) = \sum_{k\in\mathcal{K}} \alpha_k \mathbf{B}_k. \tag{7}$$

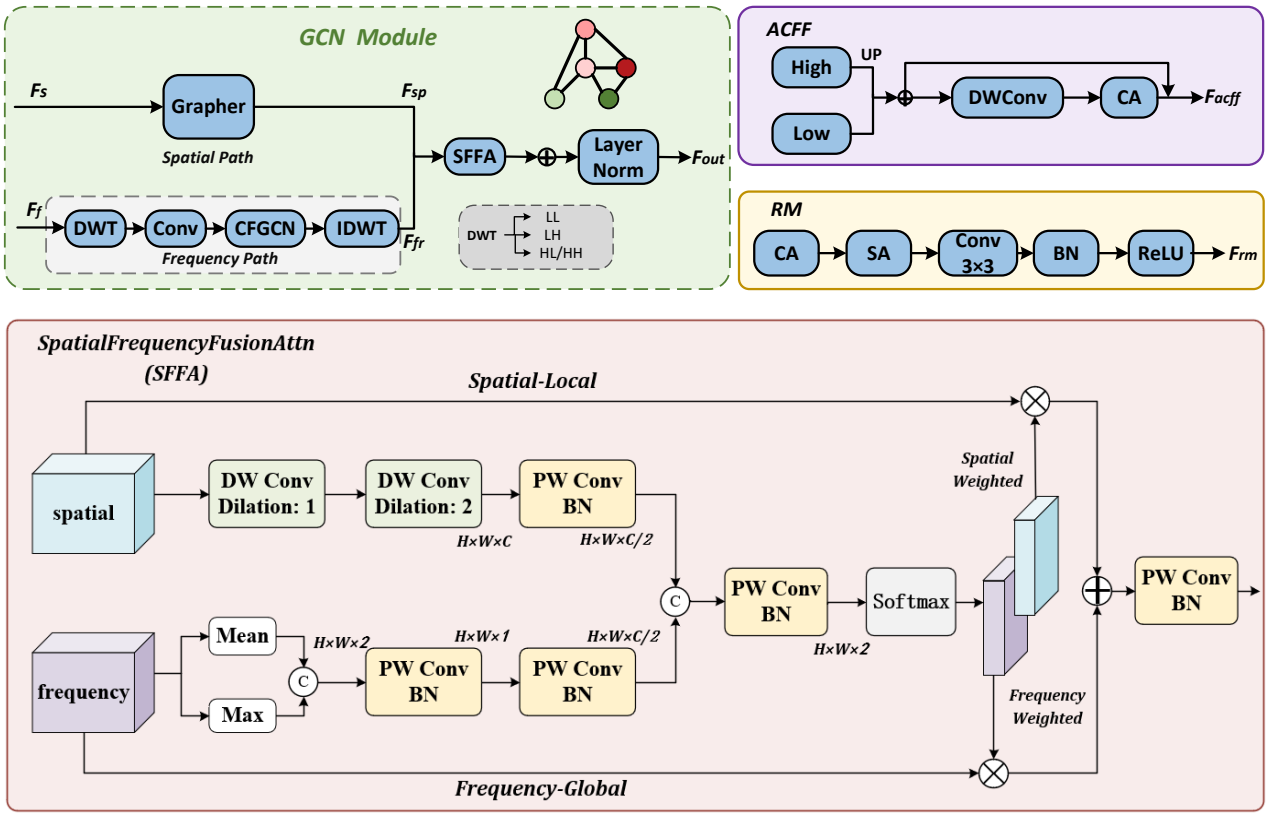

Figure 2: Internal architectures of the GCN, ACFF, and SFFA modules.

The spatial- and frequency-domain outputs are generated using independent routing weights:

$$F_s = \mathrm{FSGC}(F, \boldsymbol{\alpha}_s), \qquad F_f = \mathrm{FSGC}(F, \boldsymbol{\alpha}_f). \tag{8}$$

DGFB acts as a routing controller, whereas FSGC provides the multi-scale grid-neighborhood operators. Their combination produces input-adaptive representations with dynamically adjusted effective receptive fields. Unlike conventional multi-scale convolution, whose branches are statically fused, FSGC performs domain-specific and sample-dependent branch aggregation. It does not construct an unrestricted graph; instead, it adaptively selects local grid-graph supports for the subsequent graph reasoning modules.

Dual-Domain Graph Convolution Module (GCN)

As shown in Figure 2, the dual-domain graph convolution module performs complementary reasoning over the spatial and frequency features generated by FSGC. The spatial pathway models structural continuity and semantic dependencies, whereas the frequency pathway enhances boundaries and fine-grained details.

Spatial Graph Modeling Pathway The spatial graph pathway aims to capture the structural continuity and long-range dependencies of target regions. After adaptive neighborhood selection by FSGC, the spatial positions in F_s are treated as graph nodes. Grapher constructs content-dependent relationships among these nodes and propagates information beyond fixed convolutional neighborhoods:

$$F_{\mathrm{sp}} = \mathrm{Grapher}(F_s). \tag{9}$$

FSGC determines an appropriate local neighborhood, while Grapher further models broader feature relationships. Their combination improves morphological consistency and reduces fragmented predictions in irregular or ambiguous-boundary regions.

Frequency-Domain Graph Refinement Pathway The frequency pathway complements spatial modeling by recovering boundary and texture information. Given F_f, a discrete wavelet transform decomposes it into four sub-bands:

$$F_{\mathrm{LL}}, F_{\mathrm{LH}}, F_{\mathrm{HL}}, F_{\mathrm{HH}} = \mathrm{DWT}(F_f). \tag{10}$$

Here, F_{LL} mainly preserves global structures and semantic distributions, while F_{LH}, F_{HL}, and F_{HH} describe directional boundaries, textures, and local variations. Within each sub-band, wavelet coefficients are treated as feature nodes and refined by the corresponding graph operator:

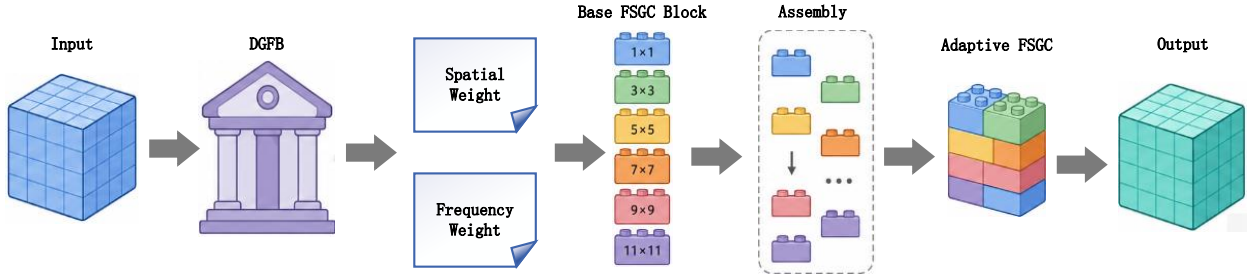

Figure 3: DGFB-guided FSGC Adaptive Dynamic Graph Convolution Illustration.

$$\widehat{F}_i = \mathrm{GCN}_i(F_i), \qquad i \in \{\mathrm{LL}, \mathrm{LH}, \mathrm{HL}, \mathrm{HH}\}. \tag{11}$$

The enhanced sub-bands are then reconstructed through inverse wavelet transform:

$$F_{\mathrm{fr}} = \mathrm{IDWT}\left(\left[\widehat{F}_{\mathrm{LL}}, \widehat{F}_{\mathrm{LH}}, \widehat{F}_{\mathrm{HL}}, \widehat{F}_{\mathrm{HH}}\right]\right). \tag{12}$$

This pathway preserves global structures through the low-frequency component while enhancing directional boundaries and fine-grained details through the high-frequency components, providing complementary features for subsequent spatial-frequency fusion.

SFFA Spatial-Frequency Fusion Attention (SFFA) The spatial and frequency pathways capture complementary information, focusing on structural semantics and boundary details, respectively. Direct fusion may introduce redundancy and noise. Therefore, SFFA is proposed to adaptively integrate dual-domain features. Given F_{sp} and F_{fr}, SFFA jointly models their relationship and generates fusion weights:

$$A = \sigma(\phi([F_{\mathrm{sp}}, F_{\mathrm{fr}}])) \tag{13}$$

$$F_{\mathrm{sf}} = A \odot F_{\mathrm{sp}} + (1 - A) \odot F_{\mathrm{fr}} \tag{14}$$

Where $[\cdot]$ denotes feature concatenation, $\phi([\cdot])$ denotes a lightweight convolutional mapping, and $\sigma(\cdot)$ denotes the Sigmoid activation function. F_{sf} represents the final fused feature.

SFFA dynamically selects informative features rather than favoring a fixed domain. Spatial features provide semantic consistency, while frequency features enhance boundary details, achieving a balance between regional coherence and boundary discrimination.

Adaptive Cross-stage Feature Fusion (ACFF)

In encoder-decoder networks, shallow features provide spatial details, while deep features contain semantic information. Direct skip connections may cause feature inconsistency and noise interference. Therefore, ACFF is introduced to adaptively align and fuse multi-level features.

As shown in Figure 2, ACFF first aligns the resolution and channels of high- and low-level features, then performs feature fusion and local refinement. ECA attention further generates adaptive weights to balance semantic and detail information. The output of ACFF is formulated as:

$$F_{\mathrm{acff}} = W \odot F_l + (1 - W) \odot F_h \tag{15}$$

Here, F_l and F_h denote aligned low- and high-level features, W represents the attention-based fusion weight,and $\odot$ denotes element-wise multiplication. ACFF adaptively balances multi-level features, reducing the encoder-decoder semantic gap and providing effective inputs for dual-domain graph convolution.

Refinement Module (RM)

As illustrated in Figure 2, RM refines intermediate decoding features. It removes redundant responses and background noise through lightweight feature recalibration, improving feature quality and segmentation consistency.

Given an input feature F , RM first enhances effective responses through a channel-spatial attention module:

$$F_a = \mathrm{CA_SA}(F) \tag{16}$$

Channel attention enhances semantic features, while spatial attention highlights target regions. The refined feature is then updated by convolution, normalization, and activation:

$$F_{\mathrm{rm}} = \mathrm{ReLU}(\mathrm{BN}(\mathrm{Conv}_{3\times3}(F_a))). \tag{17}$$

Although lightweight, RM refines intermediate features, suppresses redundant responses, and provides stable representations for subsequent dual-domain graph convolution.

Experiments

Datasets and Evaluation Metrics

We evaluate FD-GCN on four public medical image segmentation datasets, including ISIC2016, ISIC2017, ISIC2018(Mahbod et al. 2020) for skin lesion segmentation and Synapse for multi-organ segmentation. The ISIC datasets assess lesion segmentation performance, while Synapse evaluates the model's generalization ability on complex anatomical structures. Following common protocols, we adopt standard training/testing splits for all datasets to ensure fair comparison with existing.

Method	Performance Measures														
	ISIC2016					ISIC2017					ISIC2018				
	IoU	D	A	Sp	Sn	IoU	D	A	Sp	Sn	IoU	D	A	Sp	Sn
TransUnet(Chen et al. 2021)	82.69	90.53	94.49	97.38	87.75	77.2	87.14	96.22	98.19	85.09	78.69	88.08	93.61	96.49	86.01
Swin-Unet(Cao et al. 2022)	76.69	86.8	92.66	97.86	80.52	69.97	82.33	95.18	98.79	74.72	71.44	83.34	91.4	96.28	78.47
Transdeeplab(Azad et al. 2022)	73.62	84.81	91.17	95.02	82.18	65.14	78.89	94.04	97.58	74.03	66.97	80.22	90.15	96.7	72.82
MISSFormer(Pan et al. 2023)	80.99	89.5	94.08	98.32	84.15	75.37	85.95	96.03	98.75	80.66	76.87	86.92	93.22	97.39	82.18
DAE-Former(Azad et al. 2023)	79.66	88.68	93.4	96.46	86.23	71.72	83.53	95.54	99.14	75.19	73.76	84.89	92.31	97.39	78.83
D-LKA Net(Azad et al. 2024)	85.69	92.29	95.52	98.17	89.35	79.35	88.48	96.68	98.8	84.69	80.59	89.25	94.17	96.39	88.29
ParaTransCNN(Sun, Xu, and Duan 2024)	83.44	90.97	94.6	96.32	90.61	76.02	86.37	95.86	97.39	87.23	79.39	88.51	93.93	97.17	85.32
AgileFormer(Qiu et al. 2024)	74.12	85.13	91.78	97.44	78.54	67.56	80.64	94.39	97.34	77.69	76.55	86.72	93.08	97.09	82.45
MSA-2Net(Ghorbani Kolahi et al. 2024)	84.72	92.31	94.57	98.51	88.69	79.59	88.63	96.69	98.49	86.36	80.09	88.94	94.15	97.29	85.84
CENet(Bozorgpour et al. 2025)	76.18	86.48	92.64	98.72	78.45	77.79	87.50	96.37	98.49	84.23	78.38	87.88	93.74	97.90	82.74
MRP-UNet(Liu et al. 2025)	82.41	90.18	95.17	90.37	88.21	76.06	88.36	93.51	92.18	85.42	78.84	89.36	94.19	91.04	85.51
Skin-DeepNet(Al-Waisy et al. 2025)	80.55	90.76	93.73	95.39	86.47	74.47	84.09	94.89	98.86	87.15	77.71	86.34	93.88	95.03	85.95
PraNet-V2(Hu et al. 2026)	79.92	90.03	94.26	93.20	86.21	74.99	86.91	95.26	94.11	83.94	77.50	84.36	93.25	95.72	83.53
FD-GCN(Ours)	86.26	93.57	95.18	98.58	87.25	80.67	89.57	97.34	98.11	86.59	81.59	90	95.71	96.55	89.02

Table 1: Comparison of different methods on ISIC2016, ISIC2017, and ISIC2018 datasets.

Methods	Dice	HD95	Spl	RKid	LKid	Gal	Liv	Sto	Aor	Pan
TransUNet (2021)	81.50	18.50	92.71	81.23	83.94	68.24	95.03	79.66	87.43	63.80
Swin-Unet (2022)	76.21	20.59	87.58	74.09	81.78	66.52	93.87	74.49	81.35	49.97
TransDeepLab (2022)	81.99	22.53	92.11	80.11	84.85	71.33	94.18	81.09	87.11	65.11
MISSFormer (2023)	80.65	20.10	92.67	79.11	82.76	70.28	94.54	80.22	85.72	59.88
DAEFormer (2023)	81.29	25.69	90.91	79.25	85.71	68.49	94.88	79.54	86.63	64.93
D-LKA Net (2024)	82.44	17.16	91.60	81.67	87.58	65.32	94.97	83.99	87.84	66.53
MSA-2Net (2024)	84.13	14.26	93.46	85.52	88.64	68.87	95.46	84.04	88.33	68.62
G-CASCADE(2024)	83.06	21.76	91.91	81.49	86.16	69.68	95.73	82.99	87.64	68.69
CENet (2025)	83.73	16.10	92.26	84.67	87.01	71.80	95.04	81.43	88.22	68.38
PraNet-V2 (2026)	83.73	16.10	91.85	81.36	83.96	74.40	95.60	85.74	87.37	68.62
FD-GCN (Ours)	85.72	19.84	95.89	80.96	89.03	72.29	95.91	86.11	88.40	69.05

Table 2: Quantitative comparison with state-of-the-art methods on the Synapse multi-organ segmentation dataset.

Method	Architecture	Params (M)	FLOPs (G)
G-CASCADE	CNN + Graph	26.64	4.252
Swin-UNet	Transformer	27.17	6.16
CENet	CNN + Transformer	33.39	12.76
TransUNet	CNN + Transformer	96.07	88.91
2D D-LKA Net	CNN + Deformable Attention	101.64	19.92
MSA2Net	CNN + MaxViT	112.77	15.56
FD-GCN (Ours)	CNN + GCN + Spatial/Frequency	108.84	17.274

Table 3: Comparison of model parameters and computational complexity.

Implementation Details

FD-GCN is implemented with PyTorch and trained on an NVIDIA A800 GPU. We use the AdamW optimizer with a learning rate of 0.001 and weight decay of 0.0001. The input size is set to 224×224, with a batch size of 5 and 250 training epochs. Data augmentation, including random flipping, rotation, and color dithering, is applied to improve model robustness.

Performance Comparisons

Qualitative Results on Skin Lesion Dataset As shown in Table 1, FD-GCN is compared with recent state-of-the-art methods on three public skin lesion segmentation datasets. We select MRP-UNet, Skin-DeepNet, and CE-Net as representative baselines, covering multi-scale attention-based U-Net variants, application-oriented frameworks, and context-aware CNN models. For fair evaluation, all methods are re-trained under the same settings, and each experiment is repeated five times with average results reported. Red and blue indicate the best and second-best scores, respectively. FD-GCN achieves superior performance across the three datasets, demonstrating its effectiveness in handling complex lesion structures, ambiguous boundaries, and texture variations.

Performance Comparison on the Synapse Multi-organ Dataset As shown in Table 2, FD-GCN achieves superior segmentation performance on the Synapse multi-organ dataset, obtaining an average Dice score of 85.72%, outperforming recent methods such as MSA2Net (84.13%), CENet (83.73%), and G-CASCADE (83.06%)(**?**). At the organ level, FD-GCN achieves competitive or leading results on most organs, with notable improvements on challenging cases such as the gallbladder and pancreas, which suffer from blurred boundaries and large morphological variations. These results demonstrate that the proposed DGFB and dual-domain graph convolution effectively integrate structural and frequency information to enhance complex organ representation.

Params and FLOPs Table 3 compares the computational complexity of different medical image segmentation methods. Although FD-GCN has a relatively large parameter capacity, its FLOPs remain within a moderate range and are lower than those of several computationally intensive methods. This suggests that DGFB and FSGC improve feature representation while maintaining a reasonable balance between model capacity and computational cost.

Quantitative Results

Qualitative Results on Skin Lesion Datasets Figure 4 presents visual comparisons on the ISIC skin lesion segmentation task. Due to blurred boundaries, uneven appearance, and background texture interference, existing methods may produce boundary loss or false-positive regions. In contrast,

FD-GCN achieves more complete lesion coverage and better contour continuity, demonstrating improved robustness to complex textures and weak boundary information.

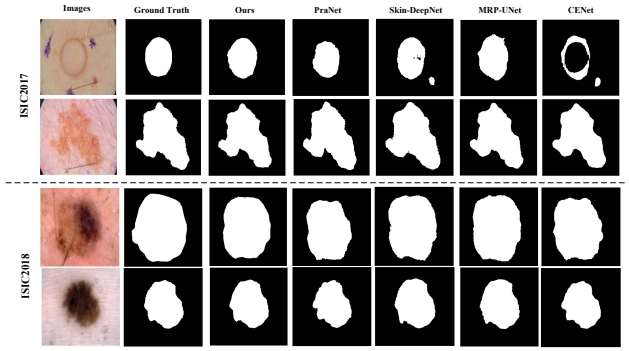

Figure 4: Visual comparison of segmentation results between FD-GCN and other models on the ISIC2017 and ISIC2018 datasets.

Qualitative Results on the Synapse Multi-organ Dataset In the Synapse task, FD-GCN effectively handles organ scale variations and ambiguous boundaries. As shown in the red-box regions, it preserves contours and reduces segmentation errors, demonstrating the effectiveness of dual-domain graph convolution in structural and boundary enhancement.

Overall, the qualitative results on both ISIC and Synapse further validate the effectiveness of the proposed method in skin lesion segmentation and multi-organ segmentation tasks. These results demonstrate that FD-GCN has strong generalization capability for targets with blurred boundaries, complex textures, and significant scale variations.

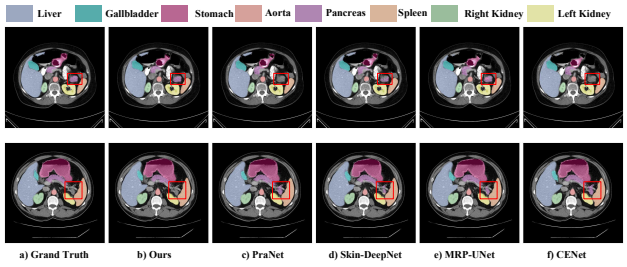

Figure 5: Visual comparison of segmentation results between FD-GCN and other models on the Synapse multi-organ dataset.

Ablation Study

Table 4 presents the ablation results of the main FD-GCN components. The improvements after introducing ACFF and RM demonstrate that cross-stage fusion and feature refinement effectively reduce semantic gaps and stabilize decoding. The addition of spatial and frequency pathways further verifies their complementary roles in structure modeling and detail enhancement. Finally, SFFA achieves further gains,

Variant	ACFF	RM	Spatial Path	Frequency Path	SFFA	DGFB +FSGC	ISIC2018 IoU↑
Baseline decoder	×	×	×	×	×	×	77.14
ACFF	✓	×	×	×	×	×	78.42
ACFF+RM	✓	✓	×	×	×	×	79.98
Spatial Path	✓	✓	✓	×	×	×	79.04
Frequency Path	✓	✓	×	✓	×	×	80.21
Spatial+Frequency	✓	✓	✓	✓	×	×	80.48
SFFA	✓	✓	✓	✓	✓	×	80.89
Full FD-GCN	✓	✓	✓	✓	✓	✓	**81.59**

Table 4: Ablation study of the main components of FD-GCN on the ISIC2018 dataset.

Variant	Dynamic Routing	S/F	FSGC	ISIC2018 IoU↑
Fixed 3×3	×	×	×	77.05
Fixed 7×7	×	×	×	77.31
Equal Multi-scale	×	×	✓	79.66
Shared Routing	✓	×	✓	79.84
DGFB + FSGC	✓	✓	✓	**81.59**

Table 5: Ablation study of dynamic routing and FSGC on the ISIC2018 dataset.

confirming the superiority of adaptive spatial-frequency fusion over simple branch combination.

Table 5 evaluates the effectiveness of DGFB and FSGC in dynamic convolutional assembly. A single fixed kernel achieves limited performance, indicating that fixed receptive fields cannot adapt to scale variations. Equal-weight multi-scale convolution improves results by introducing diverse receptive fields, but lacks input-dependent scheduling. With dynamic routing, performance is further improved, demonstrating that DGFB-generated weights effectively guide multi-scale feature aggregation. The complete DGFB+FSGC achieves the best results on both ISIC2018 and Synapse, validating the effectiveness of the "frequency bank generates instructions and convolutional blocks perform assembly" strategy for adaptive feature modeling.

Conclusion

In this paper, we propose FD-GCN, a frequency-bank-guided dual-domain graph convolutional network for medical image segmentation. Unlike methods that treat frequency information as auxiliary features, FD-GCN exploits frequency responses to guide spatial-frequency feature scheduling.

The core of FD-GCN is the collaboration between DGFB and FSGC. DGFB generates spatial-frequency weights as an "instruction set" to guide FSGC in aggregating multi-scale convolutional branches, enabling adaptive graph representations with dynamic receptive fields. Experiments on ISIC and Synapse datasets demonstrate the effectiveness of FD-GCN. The results show that frequency information can serve not only as feature enhancement but also as a dynamic guidance mechanism for adaptive convolutional assembly.

References

Agnes, S. A.; Solomon, A. A.; and Karthick, K. 2024. Wavelet U-Net++ for accurate lung nodule segmentation in CT scans: Improving early detection and diagnosis of lung cancer. *Biomedical Signal Processing and Control*, 87: 105509.

Al-Waisy, A. S.; Al-Fahdawi, S.; Khalaf, M. I.; Mohammed, M. A.; Al-Attar, B.; and Al-Andoli, M. N. 2025. A deep learning framework for automated early diagnosis and classification of skin cancer lesions in dermoscopy images. *Scientific Reports*, 15(1): 31234.

Alessio, S. M. 2015. Discrete wavelet transform (DWT). In *Digital signal processing and spectral analysis for scientists: concepts and applications*, 645–714. Springer.

Alzubaidi, L.; Zhang, J.; Humaidi, A. J.; Al-Dujaili, A.; Duan, Y.; Al-Shamma, O.; Santamaría, J.; Fadhel, M. A.; Al-Amidie, M.; and Farhan, L. 2021. Review of deep learning: concepts, CNN architectures, challenges, applications, future directions. *Journal of big Data*, 8(1).

Azad, R.; Arimond, R.; Aghdam, E. K.; Kazerouni, A.; and Merhof, D. 2023. Dae-former: Dual attention-guided efficient transformer for medical image segmentation. In *International workshop on predictive intelligence in medicine*, 83–95. Springer.

Azad, R.; Heidari, M.; Shariatnia, M.; Aghdam, E. K.; Karimijafarbigloo, S.; Adeli, E.; and Merhof, D. 2022. Transdeeplab: Convolution-free transformer-based deeplab v3+ for medical image segmentation. In *International Workshop on PRedictive Intelligence In MEdicine*, 91–102. Springer.

Azad, R.; Niggemeier, L.; Hüttemann, M.; Kazerouni, A.; Aghdam, E. K.; Velichko, Y.; Bagci, U.; and Merhof, D. 2024. Beyond self-attention: Deformable large kernel attention for medical image segmentation. In *Proceedings of the IEEE/CVF winter conference on applications of computer vision*, 1287–1297.

Binder, K.; Krauss, S.; and Wiesner, P. 2020. A new visualization for probabilistic situations containing two binary events: the frequency net. *Frontiers in psychology*, 11: 750.

Bozorgpour, A.; Kolahi, S. G.; Azad, R.; Hacihaliloglu, I.; and Merhof, D. 2025. CENet: Context Enhancement Network for Medical Image Segmentation. *arXiv preprint arXiv:2505.18423*. Early accepted at MICCAI 2025.

Cao, H.; Wang, Y.; Chen, J.; Jiang, D.; Zhang, X.; Tian, Q.; and Wang, M. 2022. Swin-unet: Unet-like pure transformer for medical image segmentation. In *European conference on computer vision*, 205–218. Springer.

Chen, J.; Lu, Y.; Yu, Q.; Luo, X.; Adeli, E.; Wang, Y.; Lu, L.; Yuille, A. L.; and Zhou, Y. 2021. Transunet: Transformers make strong encoders for medical image segmentation. *arXiv preprint arXiv:2102.04306*.

Corso, G.; Stark, H.; Jegelka, S.; Jaakkola, T.; and Barzilay, R. 2024. Graph neural networks. *Nature Reviews Methods Primers*, 4(1): 17.

Engel Jr, J.; and da Silva, F. L. 2012. High-frequency oscillations–where we are and where we need to go. *Progress in neurobiology*, 98(3): 316–318.

Ghorbani Kolahi, S.; Kamal Chaharsooghi, S.; Khatibi, T.; Bozorgpour, A.; Azad, R.; Heidari, M.; Hacihaliloglu, I.; and Merhof, D. 2024. MSA2Net: Multi-scale Adaptive Attention-guided Network for Medical Image Segmentation. *arXiv e-prints*, arXiv–2407.

Han, K.; Xiao, A.; Wu, E.; Guo, J.; Xu, C.; and Wang, Y. 2021. Transformer in transformer. *Advances in neural information processing systems*, 34: 15908–15919.

Heckbert, P. 1995. Fourier transforms and the fast Fourier transform (FFT) algorithm. *Computer Graphics*, 2(1995): 15–463.

Hu, B.-C.; Ji, G.-P.; Shao, D.; and Fan, D.-P. 2026. Pranet-v2: Dual-supervised reverse attention for medical image segmentation. *Computational Visual Media*.

Huang, G.; Liu, Z.; Van Der Maaten, L.; and Weinberger, K. Q. 2017. Densely connected convolutional networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 4700–4708.

Li, G.; Muller, M.; Thabet, A.; and Ghanem, B. 2019. DeepGCNs: Can GCNs go as deep as CNNs? In *Proceedings of the IEEE/CVF international conference on computer vision*, 9267–9276.

Li, Y.; Yu, R.; Shahabi, C.; and Liu, Y. 2017. Diffusion convolutional recurrent neural network: Data-driven traffic forecasting. *arXiv preprint arXiv:1707.01926*.

Liu, Z.; Hu, J.; Gong, X.; and Li, F. 2025. Skin lesion segmentation with a multiscale input fusion U-Net incorporating Res2-SE and pyramid dilated convolution. *Scientific Reports*, 15(1): 7975.

Mahbod, A.; Schaefer, G.; Wang, C.; Dorffner, G.; Ecker, R.; and Ellinger, I. 2020. Transfer learning using a multi-scale and multi-network ensemble for skin lesion classification. *Computer methods and programs in biomedicine*, 193: 105475.

Meng, Y.; Zhang, H.; Gao, D.; Zhao, Y.; Yang, X.; Qian, X.; Huang, X.; and Zheng, Y. 2021. BI-GCN: Boundary-aware input-dependent graph convolution network for biomedical image segmentation. *arXiv preprint arXiv:2110.14775*.

Nema, P.; Khapra, M.; Laha, A.; and Ravindran, B. 2017. Diversity driven attention model for query-based abstractive summarization. *arXiv preprint arXiv:1704.08300*.

Pan, Z.; Ding, S.; Sun, G.; Zhang, A.; Jia, X.; and Fu, H. 2023. Multi-scale spectral-spatial dual-transformer network for hyperspectral image classification. *International Journal of Remote Sensing*, 44(7): 2480–2494.

Peng, H.; Xue, C.; Shao, Y.; Chen, K.; Xiong, J.; Xie, Z.; and Zhang, L. 2020. Semantic segmentation of litchi branches using DeepLabV3+ model. *Ieee Access*, 8: 164546–164555.

Qiu, P.; Yang, J.; Kumar, S.; Ghosh, S. S.; and Sotiras, A. 2024. AgileFormer: spatially agile transformer UNet for medical image segmentation. *arXiv preprint arXiv:2404.00122*.

Ronneberger, O.; Fischer, P.; and Brox, T. 2015. U-net: Convolutional networks for biomedical image segmentation. In *Medical image computing and computer-assisted intervention–MICCAI 2015: 18th international conference,*

Munich, Germany, October 5-9, 2015, proceedings, part III 18, 234–241. Springer.

Sun, H.; Xu, J.; and Duan, Y. 2024. Paratranscnn: Parallelized transcnn encoder for medical image segmentation. *arXiv preprint arXiv:2401.15307*.

Sun, K.; Zhao, Y.; Jiang, B.; Cheng, T.; Xiao, B.; Liu, D.; Mu, Y.; Wang, X.; Liu, W.; and Wang, J. 2019. High-resolution representations for labeling pixels and regions. *arXiv preprint arXiv:1904.04514*.

Tu, Z.; Talebi, H.; Zhang, H.; Yang, F.; Milanfar, P.; Bovik, A.; and Li, Y. 2022. Maxvit: Multi-axis vision transformer. In *European conference on computer vision*, 459–479. Springer.

van Haarlem, M. P.; Wise, M. W.; Gunst, A.; Heald, G.; McKean, J. P.; Hessels, J. W.; de Bruyn, A. G.; Nijboer, R.; Swinbank, J.; Fallows, R.; et al. 2013. LOFAR: The low-frequency array. *Astronomy & astrophysics*, 556: A2.

Velickovic, P.; Cucurull, G.; Casanova, A.; Romero, A.; Lio, P.; Bengio, Y.; et al. 2017. Graph attention networks. *stat*, 1050(20): 10–48550.

Voulodimos, A.; Doulamis, N.; Doulamis, A.; and Protopapadakis, E. 2018. Deep learning for computer vision: A brief review. *Computational intelligence and neuroscience*, 2018(1): 7068349.

Wang, W.; Sun, H.; and Wang, X. 2024. Lssnet: A method for colon polyp segmentation based on local feature supplementation and shallow feature supplementation. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, 446–456. Springer.