Inverse Problems and Imaging

doi:10.3934/ipi.2023049

VISION GRAPH U-NET: GEOMETRIC LEARNING ENHANCED ENCODER FOR MEDICAL IMAGE SEGMENTATION AND RESTORATION

Yuanhong Jiang[1], Qiaoqiao Ding[2], Yu Guang Wang[1,2,3], Pietro Liò[4] and Xiaoqun Zhang*[1,2,3]

[1]School of Mathematical Sciences, MOE-LSC, Shanghai Jiao Tong University, China

[2]Institute of Natural Sciences, Shanghai Jiao Tong University, China

[3]Shanghai Artificial Intelligence Laboratory, China

[4]Department of Computer Science and Technology
University of Cambridge, United Kingdom

(Communicated by Jin Keun Seo)

Abstract. Convolutional neural networks (CNNs) are known for their powerful feature extraction ability, and have achieved great success in a variety of image processing tasks. However, convolution filters only extract local features and neglect long-range self-similarity information, which is the vital prior information commonly existing in image data. To this end, we put forward a new backbone neural network: vision graph U-Net (VGU-Net), which is the first model to construct multi-scale graph structures through the hierarchical down-sampling layers of the U-Net architecture. The graph structure is constructed by the self-attention mechanism. By replacing CNNs in the bottleneck layer and skip connection layers with the graph convolution networks (GCNs), the multi-scale graph structure visualization allows an interpretation of long-range interactions. We extend the VGU-Net backbone model for the widely considered compressed sensing MR image reconstruction task and propose a knowledge-driven deep unrolling scheme based on the half-quadratic splitting algorithm, which combines the interpretability of knowledge-driven model with the versatility of data-driven deep learning method to achieve remarkable reconstruction results. Moreover, we verify the segmentation ability of the VGU-Net backbone model on the multi-modality brain tumor segmentation dataset and white blood cell image segmentation dataset, and both achieve state-of-the-art performance. The code is publicly available at https://github.com/jyh6681/VGU-Net.

2020 *Mathematics Subject Classification.* Primary: 68T07; Secondary: 92C55.

Key words and phrases. Vision graph U-Net, geometric deep learning, self-attention, multi-modality medical image segmentation, knowledge-driven deep unrolling scheme.

This work was supported by Shanghai Municipal Science and Technology Major Project (2021SHZDZX0102), Frontier research on the fundamentals of artificial intelligence mathematics (P22KN00524) and NSFC (No. 12090024 and 12201402). We thank the Student Innovation Center at Shanghai Jiao Tong University for providing us the computing services.

*Corresponding author: Xiaoqun Zhang.

1. **Introduction.** Recently, deep learning approaches - particularly convolutional neural networks - have achieved great success in various image processing and computer vision tasks. A large number of neural network (NN) architectures and techniques have been developed to deal with the existing limitations of traditional methods. An example of such a technique refers to the U-Net [35], comprising a symmetrical encoder-decoder convolutional network with skip connections. This network has been widely employed in the field of medical image processing, particularly for datasets with limited samples. The effectiveness of this neural network architecture has been extensively investigated, with some researchers demonstrating its ability to extract multi-scale image features. For example, learning of a Haar wavelet basis representation can be achieved through the down-sampling process of the U-Net architecture [11]. Additionally, there have been efforts to design variants of U-Net that satisfy the frame condition [13, 44]. Dense versions, including U-Net++ [50] and U-Net3+ [17], were proposed to further enhance its performance. For example, U-Net++ is a popular extension that introduces nested and dense skip connections to address the semantic gap and employs deep supervision learning techniques to improve segmentation performance. Another approach, known as attention U-Net (Att U-Net) [32], incorporates attention blocks before the skip connections to determine the relevance of different features. This attention mechanism assigns weights to the feature maps at each up-sampling stage, thereby improving segmentation accuracy and incorporating the significance of the region of interest (ROI).

Meanwhile, the ResNet [15] and DnCNNs [47] take the concept of residual learning to enhance the network's learning ability, which has been demonstrated to be useful in image restoration tasks. Except for the widely used convolutional neural networks, the self-attention mechanism [39] is introduced and the proposed Transformer architecture, such as Vision Transformer (ViT) [10] and Swin Transformer [25], is widely applied in computer vision tasks. The ViT partitions images into non-overlapping patches and employs the self-attention mechanism to extract features from images, while Swin Transformer applies a shifted window to make ViT more computationally efficient. Though the original ViT model exhibits significantly better performance on large objects, the performance on small objects is not satisfactory[7]. This limitation might arise from the fixed patches size adopted by transformer-based methods. A possible solution to enhance small object detection (SOD) ability is to explore refined patch sizes. By adapting the patch size to better accommodate small objects, it is plausible to improve the overall performance of the detection network[34]. Transformers have also been introduced into the U-Net architecture via the U-Net Transformer [33] and the Swin-Unet [6], both achieving state-of-the-art (SOTA) performance in various medical image segmentation tasks. Specifically, Swin-Unet utilizes a U-Net architecture with a pure Swin Transformer serving as the encoder and decoder. In order to cater specifically to the medical image segmentation task, Swin U-Net employs a smaller patch size of 4×4, which differs from the conventional setting in ViT with a patch size of 7×7. This adaptation allows Swin U-Net to better capture intricate details and nuances present in medical images, leading to improved segmentation performance.

With the development of graph neural networks [5, 8, 22, 40, 43], researchers have developed graph convolutional networks, which update node features by aggregating information from neighbouring nodes. These GCNs have currently been applied not only in graph data naturally formed from social networks, chemical compound graphs, and protein-protein graphs, but also in text classification and image denoising.

For image data possessing no natural geometric structure, the key problem is how to properly encode Euclidean space data as graph data and the inverse process that decodes graph data back to the image domain. Additionally, the Graph-FCN [26] applies a fully convolutional network (FCN) to extract image features and the graph structure is constructed based on the k nearest neighbour methods where the weight adjacent matrix is generated with the Gaussian kernel function. In [48], the dual graph convolution network (DGCNet) constructs the graph structure not only on the spatial domain but also on the feature domain. In the semantic segmentation task, the bilinear interpolation upsampling operation is performed on the down-sampled output of the DGCNet to recover the same image size as the label. In image restoration tasks, such as image denoising, the application of long-range interaction is also extensively employed. In [45, 31], deep neural networks generate multiple feature maps, and a patch-based strategy is taken to generate graph data, where each patch is treated as a vertex. Then, the graph signal is smoothed by the graph Laplacian or GCNs. Next, graph data are re-projected to image data based on the position of the patches (vertices).

In the field of medical image processing, magnetic resonance imaging (MRI) reconstruction can be treated as an image restoration task. The magnetic resonance imaging (MRI) is one of the most vital diagnostic techniques in the clinical applications. It is time-consuming to sample adequate data during the scanning process. Therefore, the under-sampled K-space data for accelerated MRI reconstruction is widely considered. For MRI reconstruction, deep learning approaches exhibit great potentials for both acceleration and high-quality preservation. Methods including [29, 36] use MRI reconstruction networks to enhance data consistency, while [1, 2] employ inversion layers and a learned proximal operator as denoising or anti-artifact layers. The authors of [23] separate the optimization process into several isolated sub-problems of which the regularization term is addressed by neural networks. ADMM-Net [38] and the learned variational network [12] train an end-to-end deep neural network with the so-called unrolling dynamic, which combines the advantages of model based method and deep learning based method. Other approaches such as [18] reconstruct the data directly in the Fourier domain, with remarkable performance.

In this study, we propose a general neural network model integrating the advantages of GCN and CNN for computer vision tasks, namely Vision Graph U-Net (VGU-Net), which is the first model to construct multi-scale graph structures through the hierarchical down-sampling layers of the U-Net architecture. In addition, the graph convolution is applied to enable message passing between nodes in the form of a weighted sum. The resulting graph features are projected back into the image domain through a reverse mapping process involving symmetric up-sampling. The proposed VGU-Net considers both local features extracted by CNN and long distance connections encoded by graph. Moreover, its effectiveness is demonstrated through two widely considered medical image tasks: medical image segmentation and MRI reconstruction, with high accuracy and efficiency in relative to other state-of-the-art methods.

2. The proposed VGU-Net.

The self-similarity is an important prior in images. In numerous convolution-based deep learning methods, the convolution kernel locally extracts image features, while the long-range information interaction is limited. The intuition is to learn a graph structure that reflects the similarities among the images and use powerful GCNs to perform interactions between pixels.

Different from the graph data with natural geometric structure, how to properly transform the regular Euclidean data into the graph data and re-project the graph data back to the original domain is vital. Therefore, the geometric information contained in the image data can be learned to the network more efficiently. The crucial element of such transformation lies in encoding the graph vertex features and their connection behavior, followed by their decoding with the re-projection process.

Therefore, we propose the U-shape neural network (VGU-Net) to effectively realize such encoder-decoder architecture with graphs. Besides, we describe the detailed network architecture as follows.

2.1. Main architecture.

By inheriting the basic encoder-decoder architecture of U-Net, the overall architecture of VGU-Net is illustrated in Figure 2. It consists of three paths of graph encoders and three paths of graph decoders, performing down-sampling from image domain to the graph domain and up-sampling from the graph back to the image domain, respectively.

The graph encoder paths hierarchically down-sample the image data and extract adequate node features for the use in the graph. Afterwards, feature graph convolution is used to establish the connections between the nodes. Graph convolution layers are inserted between the same-level graph encoder-decoder paths.

Mostly similar to the structure of U-Net, the feature extractor employs CNNs of two 3×3 convolutions, each followed by a batch normalization [19] layer and rectified linear unit (ReLU) [14] layer. In terms of the down-sampling and up-sampling process, a 2×2 convolution/transpose convolution layer with stride 2 is applied. In order to perform the message passing between nodes, two layers of graph convolution are used.

For the image restoration task including image denoising and MRI reconstruction, as illustrated by the dash line in Figure 2, there is a skip connection for performing residual learning, rendering the model more efficient and stable [15].

2.2. Graph encoder on feature maps.

Given the input image $x \in \mathbb{R}^{m \times n}$, the convolutional encoder networks generate M channels subsampled feature maps $X = \{f_i\}_{i=1}^{M}$ with $f_i \in \mathbb{R}^{\frac{m}{d} \times \frac{n}{d}}$. With $N = \frac{m}{d} \times \frac{n}{d}$, we obtain the vectorized $\hat{f}_i \in \mathbb{R}^{N \times 1}$ and obtain feature maps of size $\hat{\mathcal{X}} \in \mathbb{R}^{N \times M}$ in order to generate the graph $\mathcal{G}$ with N vertices. As shown in Figure 1, each node corresponds to a size $d \times d$ patch in the image domain and is encoded by the CNNs with its powerful locality feature. As the number of channels M of the feature maps is the same as the size of the feature of each vertex, the graph node naturally encode the CNN features of each pixel in the image.

We employ self-attention [39] to estimate the similarity between two nodes, and take two learnable linear transformations (δ, ψ) on the reshaped feature maps $\hat{\mathcal{X}}$ to produce the adjacency matrix

$$\mathcal{A} = \text{softmax}\left(\frac{\delta(\hat{\mathcal{X}}) \cdot \psi(\hat{\mathcal{X}})^T}{\sqrt{M}}\right),$$

where

$$\operatorname{softmax}(Z_{i,j}) = \frac{\exp(Z_{i,j})}{\sum_{j=1}^{C} \exp(Z_{i,j})}$$

for input $Z \in \mathbb{R}^{N\times C}$ and $i = 1, 2, ..., N$, $j = 1, 2, ..., C$. Moreover, this helps us optimally learn the graph structure for different tasks, dynamically adjusting the connectivity accordingly.

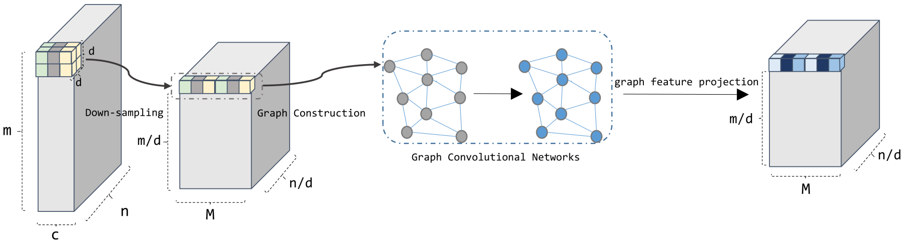

FIGURE 1. The graph construction process involves down-sampling operations. Each $d \times d$ patch with c channels is down-sampled to form a single pixel in the subsequent feature maps, representing a graph node. Following the graph convolutional network embedding and feature projection process, the graph data can be mapped back to image domain.

2.3. **Graph convolution neural networks.** Given a graph $\mathcal{G} = (\mathcal{V}, \mathcal{A})$, where $\mathcal{V}$ is the set of N nodes and each node is represented by a feature vector of size M, forming a feature matrix $h^{(0)} \in \mathbb{R}^{N\times M}$. Besides, $\mathcal{A}$ is the adjacency matrix representing the connectivity between nodes, with $\mathcal{A} \in \mathbb{R}^{N\times N}$.

As described in [22], given the hidden representation $h^{(l)}$ in the $l-$th layer of GCNs, we can compute the one-step forward propagation using the following equation,

$$h^{(l+1)} = \sigma(\tilde{\mathcal{D}}^{-\frac{1}{2}}\tilde{\mathcal{A}}\tilde{\mathcal{D}}^{-\frac{1}{2}}h^{(l)}W^{(l)}), \tag{1}$$

where $\tilde{A} = A + I_N$ is the normalized adjancecy matrix adding self-loop to enhance self-connectivity and lower numerical instabilities, $\tilde{D}_{ii} = \sum_j \tilde{A}_{ij}$ is the corresponding degree matrix, $\sigma(\cdot)$ is the non-linear activation function, and $W^{(l)}$ represents the parameter to be learned. The GCNs efficiently propagate the message of a node by aggregating its neighboring nodes' features, which has been demonstrated by various real-world graph learning tasks.

In the proposed VGU-Net, after obtaining the graph from the graph encoder, we perform two layers of GCNs to encode the geometric information of the image based on the following equation:

$$\begin{aligned} \tilde{\mathcal{X}^{(1)}} &= \sigma(\tilde{\mathcal{D}}^{-\frac{1}{2}}\hat{\mathcal{A}}\tilde{\mathcal{D}}^{-\frac{1}{2}}\hat{\mathcal{X}}\mathcal{W}^{(1)}), \\ \tilde{\mathcal{X}^{(2)}} &= \sigma(\tilde{\mathcal{D}}^{-\frac{1}{2}}\hat{\mathcal{A}}\tilde{\mathcal{D}}^{-\frac{1}{2}}\hat{\mathcal{X}^{(1)}}\mathcal{W}^{(2)}). \end{aligned} \tag{2}$$

2.4. **Graph feature re-projection.** Following the graph convolution layers, the output graph representation $\tilde{\mathcal{X}} \in \mathbb{R}^{N\times M}$ contains the node features that will be converted back to the image domain. We reshape $\tilde{\mathcal{X}}$ into $\tilde{X} = \{\tilde{f}_i\}_{i=1}^{M}, \tilde{f}_i \in \mathbb{R}^{\frac{m}{d}\times\frac{n}{d}}$, mapping each node to its corresponding position in the image space.

The resulting feature maps $\tilde{X}$ is encoded with long-range interaction by two-layer GCNs. Then, $\tilde{X}$ is concantenated with the feature maps from the decoder path. Next, the concantenated feature maps are fed to the up-sampling convolutional neural ntowrks.

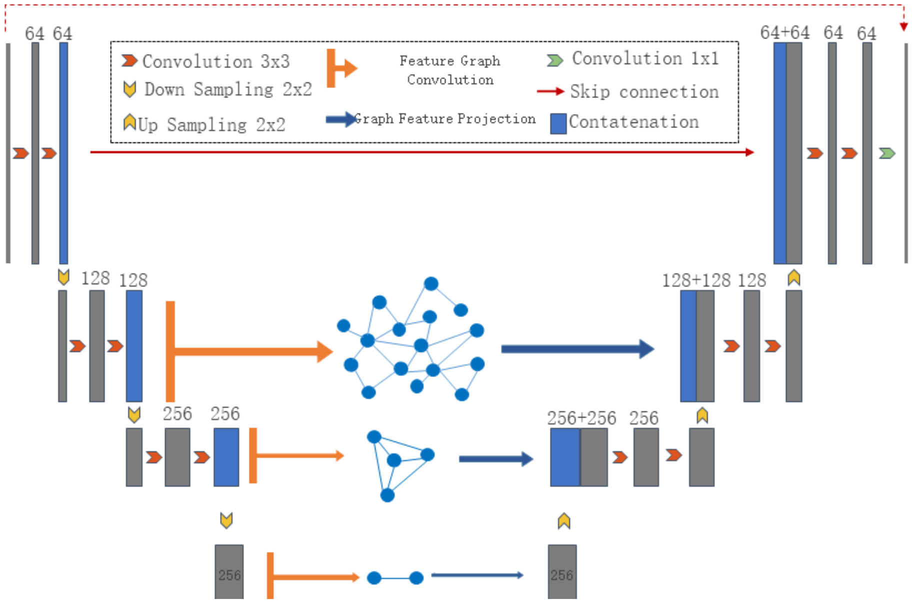

FIGURE 2. The designed diagram of VGU-Net. The encoder-decoder path remains the same as architecture of U-Net. The CNNs in the bottle-neck layer and two skip connections are substituted with two layers of GCNs, performing graph representation learning to take long-range interaction between nodes. The residual connection represented by the dash line is specifically designed for image restoration tasks.

3. **VGU-Net for segmentation and MRI reconstruction.** The medical image usually has a limited number of samples while the processing tasks have unremitting pursuit for precision, safety and imaging speed. Featured with the GNN enhanced encoder and small number of learnable parameters, the proposed back-bone neural network model VGU-Net can be used for various medical image processing tasks, including image segmentation, image denoising, anomaly detection and image reconstruction. In this study, we experiment the VGU-Net model in two medical image segmentation datasets and propose an optimization guided unrolling MR image reconstruction methods with the VGU-Net as the back-bone neural networks.

3.1. **Medical image segmentation.** Medical image segmentation is extremely useful in diagnosis and surgery operation. Image segmentation aims to divide image into several non-overlapped regions. Considering one 2D image signal f, (x, y) belongs to an closed and bounded image domain $\Omega \subset \mathbb{R}^2$. The segmentation problem consists of finding a decomposition of the region $\Omega = \left(\bigcup_{i=1,\dots,K} \Omega_i\right)$, where Ω_i

are disjoint closed sets representing different ROIs. The VGU-Net can be implemented in image segmentation tasks by appending a softmax segmentation head in the end of the decoder. The end-to-end training is generally implemented by feeding the image to the back-bone segmentation neural networks and using the dice loss function [30]. The dice loss function is the negative dice similarity coefficient, providing a widely used metric on image segmentation. To be more specific, given a set G, we define its characteristic/label function by $\iota_G(i) = \begin{cases} 1, & i \in G \\ 0, & o.w. \end{cases}$. The loss function of two sets G and $\hat{G}$ is defined as

$$\ell(G, \hat{G}) = -\frac{2\sum_{i\in\Omega} \iota_G(i) \cdot \iota_{\hat{G}}(i)}{\sum_{i\in\Omega} \left(\iota_G(i) + \iota_{\hat{G}}(i)\right)}, \tag{3}$$

where Ω indicates the domain containing the two sets.

3.2. **Compressed sensing MRI reconstruction.** The residual learning can deal with the gradient vanishing and degradation problem in the deep neural networks [15], enabling the VGU-Net to be used in image restoration. Particularly, we consider the compressed sensing MRI (CS-MRI) reconstruction problem, which can significantly lower the sampling time compared to traditional MRI reconstruction. The problem is formulated as

$$y = D \odot \mathcal{F}x, \tag{4}$$

where $y \in \mathbb{C}^{m\times n}$ represents the measured projection data, $x \in \mathbb{C}^{m\times n}$ is the image to be reconstructed, $D \in \mathbb{R}^{m\times n}$ is the down-sampling matrix that consists of entry value $\{0, 1\}$, $\odot$ indicates the Hadamard product, and $\mathcal{F} \in \mathbb{C}^{m\times m}$ is the Fourier transform matrix.

In terms of CS-MRI reconstruction problem, we take the Half Quadratic Splitting algorithm [16] to iteratively solve the data-consistent sub-problem and refine the solution using the proposed VGU-Net.

3.2.1. *CS-MRI reconstruction model.* The general CS-MRI reconstruction optimization model is formulated as following:

$$\min_x \frac{1}{2}\|D \odot \mathcal{F}x - y\|_2^2 + \lambda R(x), \tag{5}$$

where $R(x)$ suggests the regularization term reflecting the prior knowledge of the image to reconstruct, and λ represents a weight to balance the measurement and regularization term. Then, we will introduce the knowledge-driven deep unrolling scheme for MRI reconstruction integrating knowledge-based solutions and deep neural network priors. Knowledge-based methods offer initial solutions that help deep neural networks converge faster, while the unrolling scheme builds a bridge between knowledge-based methods and deep learning methods.

3.2.2. *Half quadratic splitting algorithm.* By introducing the auxiliary variables $u \in \mathbb{C}^{m\times n}$, we can convert the unconstrained model (5) into the following constrained model

$$\min_{u,x} \frac{1}{2}\|D \odot \mathcal{F}x - y\|_2^2 + \lambda\mathcal{R}(u) \quad s.t. \quad u = x. \tag{6}$$

Minimizing (6) by the penalty function method with hyper-parameter β, we can obtain:

$$\min_{u,x} \frac{1}{2} \|D \odot \mathcal{F}x - y\|_2^2 + \lambda \mathcal{R}(u) + \frac{\beta}{2} \|x - u\|_2^2 . \tag{7}$$

To minimize (7), we alternatively optimize u and x by solving the following two sub-problems:

$$\begin{cases} x^{(k+1)} = \arg\min\limits_{x} \frac{1}{2} \|D \odot \mathcal{F}x - y\|_2^2 + \frac{\beta}{2} \left\|x - u^{(k)}\right\|_2^2 \\ u^{(k+1)} = \arg\min\limits_{u} \frac{\beta}{2} \left\|u - x^{(k+1)}\right\|_2^2 + \lambda \mathcal{R}(u), \end{cases} \tag{8}$$

where $k \in \{0, 1, \ldots, n_{it}\}$ denotes the $k-$th iteration.

Given u^{k-1}, it is easy to obtain the closed-form solution

$$x^{(k)} = \mathcal{F}^{-1} \left(\frac{D \odot y + \beta \mathcal{F} u^{(k-1)}}{D + \beta} \right), \tag{9}$$

for the first sub-problem. For the second sub-problem, we consider the deep learning approach using VGU-Net to obtain $u^{(k)}$. The overall deep unrolling MRI reconstruction workflow is displayed in Fig. 3.

Supposing that $x^{(k)}$ is the input to VGU-Net in the $k-$th stage, and $\mathcal{O}^{(k)}$ is the output. Then, using the residual learning, we obtain:

$$u^{(k)} = x^{(k)} - \mathcal{O}^{(k)}.$$

After n_{it} unrolling stages, the whole networks are trained by minimizing the mean square error loss function

$$\mathcal{L}(\Theta) = \frac{1}{\mathcal{B}} \sum_{b=1}^{\mathcal{B}} \left\| u_b^{(n_{it})} - x_b^{\mathrm{gt}} \right\|_2^2 , \tag{10}$$

where $\Theta = \bigcup\limits_{k=1}^{n_{it}} \theta_k$ and θ_k are the parameters of VGU-Net in the k-th stage, $\mathcal{B}$ is the number of batch, and x^{gt} indicates the ground truth of the reconstructed MR image.

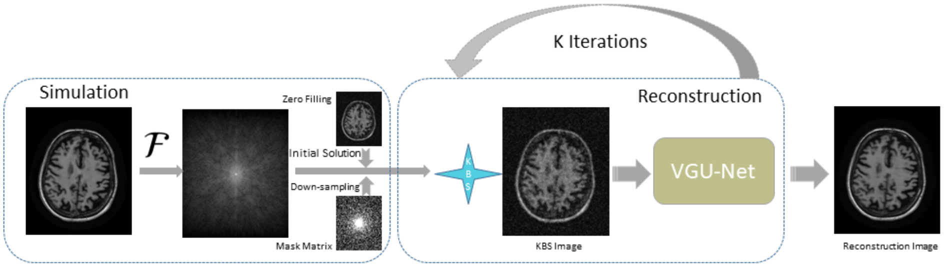

FIGURE 3. The knowledge-driven deep unrolling scheme for MRI reconstrtion. We use zero filling method to obtain initial solution $u^{(0)}$. The KBS module is the knowledge-based solution according to (9).

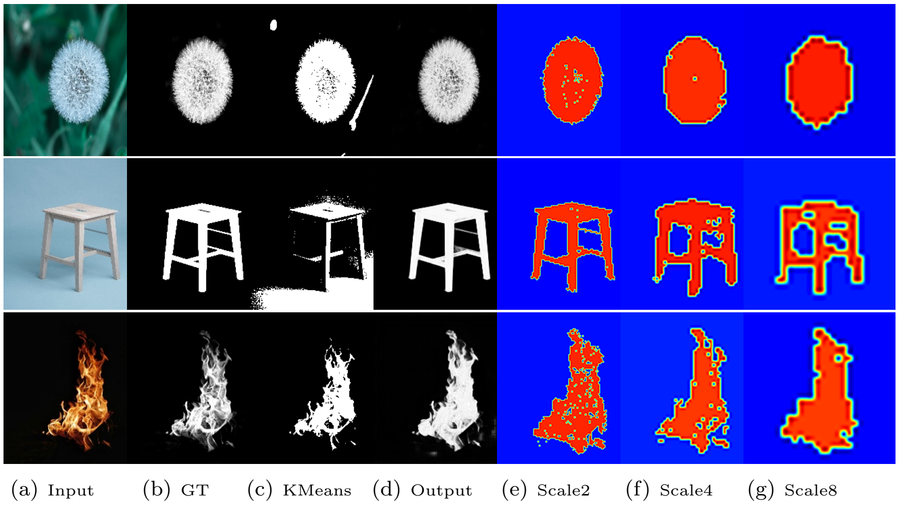

FIGURE 4. Visualization of the learned graph structure of different scales using normalized graph-cut methods. As the scale becomes larger, the details of the clustering result will gradually fade. The isolated regions have the same class as the distant regions, indicating that the long-range similarity has been learned by the dynamic graph.

4. Numerical experiments.

4.1. Geometric structure learned by graph neural networks. The graph structure of VGU-Net model is generated dynamically based on the feature maps, and the edge weights between nodes are computed using an attention mechanism that leverages the features of the nodes. Different from other approaches that lack any learning ability when computing edges, the graph structure is customized to match the image task we are addressing.

We hypothesize that nodes with similar features should be more tightly connected to each other. In order to validate this hypothesis, we conducted an experiment on the Adobe Image Matting dataset, comprising of 500 natural images. Specifically, we trained the VGU-Net model using 470 images and tested on 30 images with the mean square error as the loss function.

As shown in Figure 4, we clustered the graph nodes into two classes using normalized graph-cut methods [37, 28] at three different scales. The results show that the connections of graph nodes reveal the structure of foreground and background, through integrating long-range features in images at different scales.

In Figure 5, we further explore the impact of the graph convolution layers and graph connection from the learned adjacency matrix. The feature maps obtained after applying the graph convolution layers and projecting them back to the image space exhibit a piece-wise constant and smoother effect. Additionally, the graph connections reveal the incorporation of long-range information captured within the graph structure.

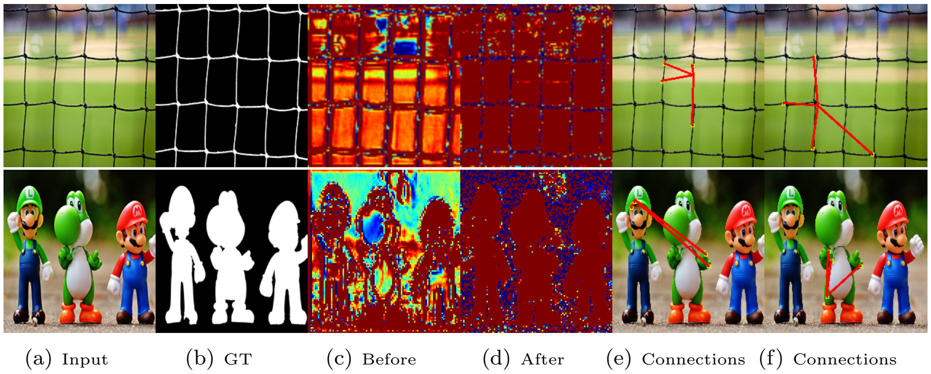

FIGURE 5. We show the feature maps before (c) and after (d) the first feature graph convolution layers. In column (e) and (f), the regions of query patches in the center are labeled with yellow boxes and we present top 4 highly correlated neighbors identified from the learned graph.

4.2. **Medical image segmentation.** The image segmentation experiments are performed by training on publicly available BraTS2018 Challenge datasets and testing on BraTS2019 Challenge datasets, consisting of brain MR images and labels of the tumor area. In addition to the MR image segmentation, we also implement the VGU-Net model on White Blood Cells (WBCs) image dataset [49] used in RU-Net [20], which contains one hundred 300 × 300 color images. The aim is to simultaneously segment the nuclei and cytoplasm, which remains a challenging task mainly because of different cell types, staining techniques and illumination conditions.

4.2.1. *BraTS datasets.* In the BraTS dataset, we extract the 2D slice image data with the size of 160×160 and normalize the four modalities: Flair, T1, T1 contrast enhance (T1ce), and T2 as the inputs of the proposed model. The target is to segment three regions including whole tumor (WT), enhance tumor (ET) and tumor core (TC). The number of training/validation/testing data is 15138/3785/3219.

We take four modalities 2D slice images as the input of the VGU-Net model. No pre-trained model is employed to initialize the model parameters. During the training process, we set the batch size as 16 and the dice loss as the loss function in this segmentation task [30]. The popular Adam [21] optimizer is applied to optimize the model for back-propagation. All the models are well-trained using the grid search method in order to find the best hyper-parameters. All the experiments are conducted with python3.8 and PyTorch 1.7.0 on NVIDIA ® Tesla A100 GPU with 6,912 CUDA cores and 80GB HBM2 mounted on an HPC cluster.

We use dice coefficient, Hausdorff distance [3], and positive predictive value (PPV) as the metric to distinguish the best model. The dice coefficient mainly measures the ratio of overlapped region between two objects: the Hausdorff distance which measures the accuracy of the boundary; the PPV, which is the ratio of true positive samples among all samples categorized as positive, and can evaluate the segmentation accuracy more fairly when addressing the class imbalance situation.

The comparison of the proposed VGU-Net with previous state-of-the-art (SOTA) segmentation models on the BraTS dataset is displayed in Table 1. Experimental results show that the VGU-Net model generally achieves state-of-the-art (SOTA) performance across various segmentation metrics. Regarding the number of trainable parameters, the VGU-Net using less than 5 million parameters makes better performance than most of the baseline models. The reason that the proposed method has less parameters compared to the original U-Net is that we substituted the two CNN layers in the bottleneck layer of U-Net by skip connection layer. Originally, the two CNN layers, featured with a large number of channels, substantially increases the number of additional learnable parameters. The comparison indicates that the performance improvement is not caused by the modification in the number of parameters.

The segmentation results of different methods on the BraTS2019 dataset are presented in Figure 6. The CNN-based methods like DeepLab V3, U-Net, U-Net++ and Attention U-Net tend to have over-segmentation problems, which may be resulted from the locality of the convolution operation. In the $10-$th row, though the graph structure alleviates the over-segmentation problem, the graph-based method DGCNet which takes the bilinear interpolation to up-sample the model's output lowers the accuracy of the segmentation result. In the $9-$th row, the original Swin-Unet taking pure Transformer as a feature extractor can not preserve the local feature. We made modifications to the original Swin-Unet model by employing a smaller patch size in the Transformer layers. Instead of the original 4×4 patch size, a smaller 2×2 patch size is adopted to better capture small objects. Additionally, the window size is ajusted from 7×7 to 5×5 to reduce memory usage. The resulting modified Swin-Unet model, referred as Swin-Unet (M), demonstrates enhanced segmentation capability and improved preservation of local feature extraction ability. As displayed in the Table 1, generally speaking, the VGU-Net integrating both local features from CNNs and long-range information from the GCNs exhibits better segmentation results.

Table 1. BraTS segmentation results. VGU-Net has the least number of trainable parameters and achieves almost the best performance in terms of the following three metrics. We compare the VGU-Net model with multiple SOTA deep learning segmentation models. The Swin-Unet (M) is the modified Swin-Unet model with refined patch partition in the Transformer layers.

Model	Parameters↓	Dice↑			Hausdorff ↓			PPV↑		
		ET	WT	TC	ET	WT	TC	ET	WT	TC
Deeplab V3	58.16M	0.6416$_{\pm 0.0621}$	0.7656$_{\pm 0.06514}$	0.7628$_{\pm 0.1005}$	3.3963$_{\pm 0.6366}$	3.2143$_{\pm 0.5203}$	2.0372 $_{\pm 2.2260}$	0.6434$_{\pm 0.0636}$	0.7662$_{\pm 0.0497}$	0.7939$_{\pm 0.0877}$
DGCNet	48.65M	0.7179$_{\pm 0.0536}$	0.8194$_{\pm 0.0485}$	0.7665$_{\pm 0.1089}$	3.1313$_{\pm 0.6335}$	2.8940$_{\pm 0.4484}$	2.0400$_{\pm 2.1665}$	0.7293$_{\pm 0.0538}$	0.8291$_{\pm 0.0396}$	0.7860$_{\pm 0.107}$
U-Net	5.43M	0.7597$_{\pm 0.0672}$	0.8279$_{\pm 0.0643}$	0.8218$_{\pm 0.0912}$	2.8339$_{\pm 0.6430}$	2.6536$_{\pm 0.4631}$	1.7502$_{\pm 1.8217}$	0.8021$_{\pm 0.0471}$	0.8782$_{\pm 0.0306}$	0.8599$_{\pm 0.0752}$
U-Net++	47.18M	0.7686$_{\pm 0.0647}$	0.8360$_{\pm 0.0596}$	0.8431$_{\pm 0.0766}$	2.8080$_{\pm 0.6279}$	2.6348$_{\pm 0.4581}$	1.7084$_{\pm 1.8466}$	**0.8178**$_{\pm 0.0444}$	**0.8918**$_{\pm 0.0268}$	0.8756$_{\pm 0.0624}$
Att-Unet	34.88M	0.7662$_{\pm 0.0622}$	0.8407$_{\pm 0.0542}$	0.8443$_{\pm 0.0783}$	2.8490$_{\pm 0.6603}$	2.6356$_{\pm 0.4728}$	1.6876$_{\pm 1.9254}$	0.7786$_{\pm 0.0547}$	0.8580$_{\pm 0.0387}$	0.8790$_{\pm 0.0625}$
Swin-Unet	73.46M	0.7612$_{\pm 0.0628}$	0.8367$_{\pm 0.0558}$	0.8368$_{\pm 0.0787}$	2.8396$_{\pm 0.7034}$	2.6705$_{\pm 0.4661}$	1.7581$_{\pm 1.9501}$	0.7891$_{\pm 0.0534}$	0.8689$_{\pm 0.0353}$	0.8730$_{\pm 0.0620}$
Swin-Unet (M)	34.26M	**0.7817**$_{\pm 0.0580}$	**0.8460**$_{\pm 0.0508}$	0.8486$_{\pm 0.0739}$	2.7787$_{\pm 0.5834}$	2.6395$_{\pm 0.4693}$	1.6943$_{\pm 1.7217}$	0.7738$_{\pm 0.0607}$	0.8370$_{\pm 0.0487}$	0.8621$_{\pm 0.0622}$
VGU-Net	**4.99M**	0.7781$_{\pm 0.0628}$	0.8420$_{\pm 0.0554}$	**0.8615**$_{\pm 0.068}$	**2.7587**$_{\pm 0.5923}$	**2.6111**$_{\pm 0.4871}$	**1.5882**$_{\pm 1.6455}$	0.8069$_{\pm 0.0526}$	0.8807$_{\pm 0.0346}$	**0.9132**$_{\pm 0.0421}$

4.2.2. *WBCs dataset.* In the WBCs dataset, we input three channels of RGB images into neural networks to perform end-to-end training. The output of models consists of three parts: the nuclei on white label, the cytoplasm on grey label and the background on black label, as shown in Figure 7. The number of training/validation/testing data is 80/10/10 and the dice loss function is applied.

The dataset comprises different cell types, staining techniques, and illumination conditions which causes large variations in the sample distribution. The main challenge in segmenting the WBCs dataset is to distinguish between cytoplasm and nuclei. In some images, nuclei are incompletely stained, and the color inside the

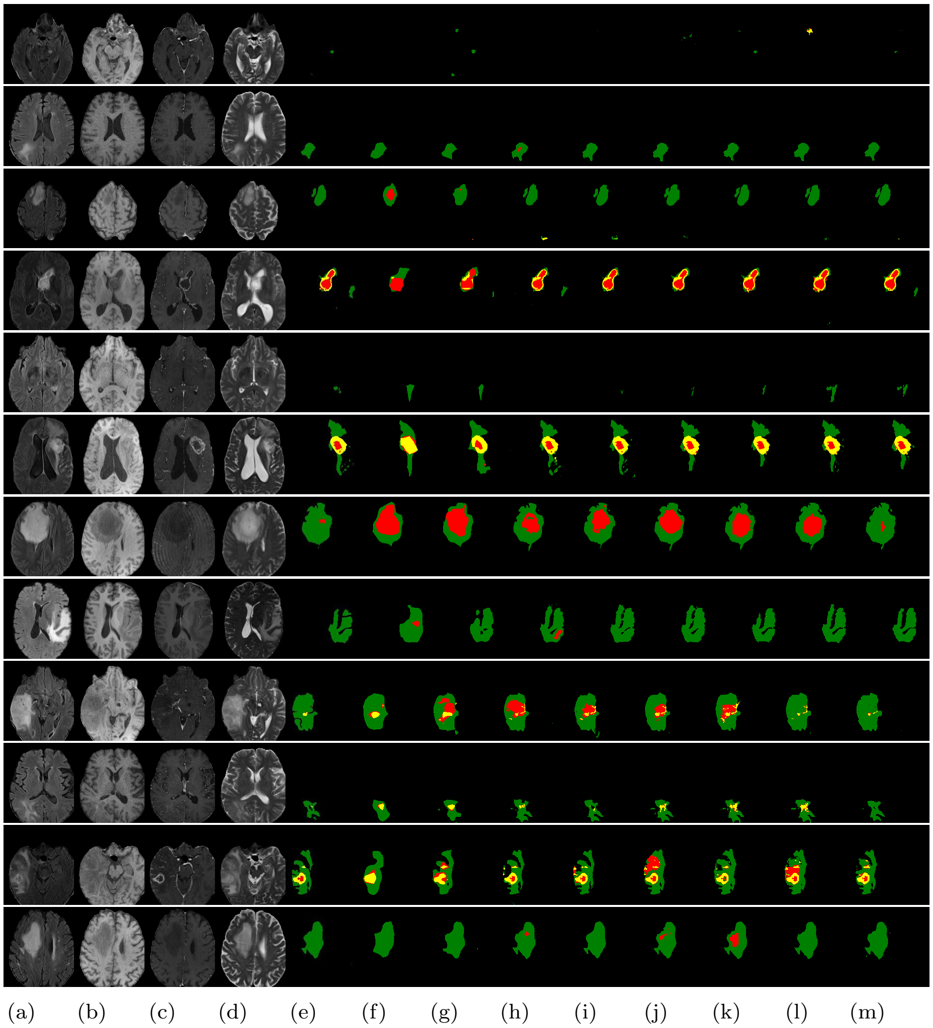

Figure 6. The indices from (a) to (m) represent the following names: Flair, T1, T1ce, T2, Annotation, Deeplab, DGCNet, U-Net, U-Net++, AttU-Net, Swin-Unet, Swin-Unet (M), and VGU-Net, respectively. The yellow region denotes the enhancing tumor (ET), red suggests the nonenhancing tumor (NET) and green refers to the peritumoral edema (ED), which show the following correlation with the segmentation objects, WT = ED + ET + NET, TC = ET+NET. We compare the VGU-Net model with multiple SOTA deep learning segmentation models. The Swin-Unet (M) in column (I) is the modified Swin-Unet model with refined patch partition in Transformer layers. The visualization results demonstrate that VGU-Net delivers promising segmentation performance for both local and distant isolated regions.

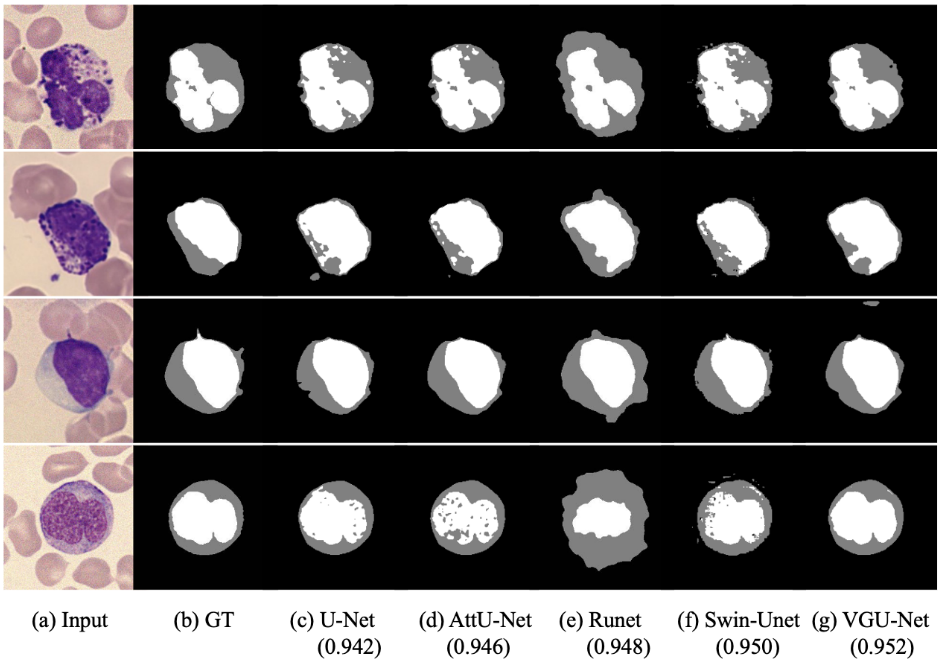

FIGURE 7. The segmentation results on the WBCs dataset, near the name of models the average dice coefficient is attached. The segmentation results of VGU-Net can well preserve the local connectivity and get good performance.

nucleus may resemble that of the cytoplasm. This similarity usually causes canonical CNN-based segmentation models to predict segmentation with holes. However, for accurate segmentation, the ground truth of nucleus must be a connected region without holes. The proposed VGU-Net, which employs multi-scale features, can preserve the region of nucleus, based on the results presented in Figure 7. In the third row of Figure 7, even though the long-range information makes VGU-Net predict some regions on the background as cytoplasm, brought about by the white blood cell outside of the image, the VGU-Net model generally exhibits SOTA performance on this challenging task.

4.3. **MRI reconstruction.** In the current section, we conduct experiments on MRI reconstruction from down-sampled measurements, and compare the proposed methods with several existing approaches. Additionally, we perform an ablation study to confirm the effectiveness of the proposed VGU-Net.

In the experiments, we use the MRI image dataset from ADNI (Alzheimer's Disease Neuroimaging Initiative) of which 300 slices of size 192×160 are used for training and 21 slices are used for training and inferring. There are three different sampling patterns, namely, 1D Gaussian mask, 2D Gaussian mask and radial mask. Three sampling rates 1/5, 1/4, and 1/3 are used for simulating the measurements, respectively. These three different types of masks with a 1/3 sampling ratio are displayed in Figure 8. In addition to the task of noise-free MRI reconstruction, we also consider the corrupted reconstruction task, where the measurement data

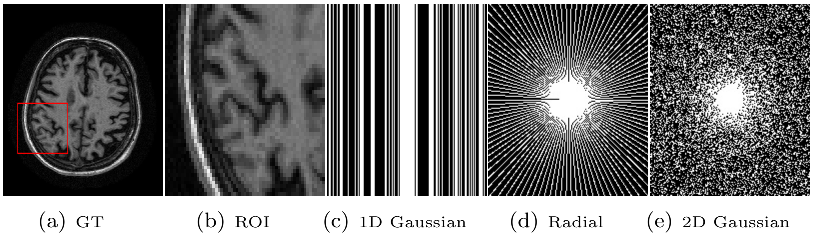

Figure 8. (a) True Image. (b) ROI. (c)(d)(e) Three different types of sampling masks of sample ratio 1/3. (c) 1D Gaussian. (d) radial lines. (e) 2D Gaussian.

is generated by $y = D \odot \mathcal{F}(x + \epsilon)$, with $\epsilon \sim N(0, \sigma^2)$ representing white Gaussian noise. We specify the noise level $\sigma = 10\%$ in the experiments. The corrupted reconstruction task confirms the robustness of the proposed reconstruction model. No additional modifications are made to account for this scenario.

As shown in the experiments, we use 3 unrolling stages which can trade off the reconstruction performance and computational cost of the VGU-Net model. The network is trained with Adam [21], with the batch size of 16, the learning rate of 10^{-4} and the weight decay of 10^{-2}. In addition, both peak signal-to-noise ratio (PSNR) and structural similarity index measure (SSIM) [42] are adopted for the quantitative assessment of image quality.

We compare the VGU-Net method with two non-learning methods, simple zero-filling method (ZF) [4], and TV-regularization-based method [27]. In addition, the deep learning-based methods, ADMM-Net [38], and the plug-&-play methods in [23] with three different networks: SCAE, SNLAE, and GAN, ISTA-Net [46] and Fourier interpolation based approach (FI) [9] are also compared. Table 2 presents the quantitative comparison in terms of PSNR and SSIM. Moreover, the proposed unrolling method combining the VGU-Net as a deep learning model makes the best performance in most cases.

Table 3 presents the results of the ablation study conducted on backbone neural networks. In the unrolling scheme, we treated network as a denoiser and compared the VGU-Net model with popular denoising models including DnCNNs and U-Net with residual learning. The VGU-Net outperforms the other two models in all ablation experiments, also indicating that the long-range interaction enhances the denoising ability of U-Net.

5. **Discussion and conclusion.** In the present study, we introduce a novel general-purpose VGU-Net combining the locality of convolutional neural networks (CNNs) with the long-range interaction property of graph convolutional networks, aiming to extract features from the patch graph of an image. The VGU-Net can learn the graph structure dynamically and the mid GCN modules extract and pass the long-distance pixel features to the decoder layers. We show the effectiveness of our approach by achieving state-of-the-art performance on two challenging image segmentation tasks and a compressed sensing MRI reconstruction task.

Although, GNN in general cannot go deep due to the oversmoothing issue of graph convolution network when the diffusion process makes the node feature indistinguishable. This constraint hinders the possibility of scaling up to larger models

TABLE 2. MRI reconstruction quality (PSNR/SSIM). the best results are marked in bold. 10% additive white noise are added for each case.

Method	Noise	Rate	ZF	TV	ADMM-Net	SCAE	SNLAE	GAN	ISTA-Net	FI	VGU-Net
radial	0%	1/5	24.36/0.47	30.73/0.86	32.31/0.92	32.00/0.92	30.47/0.83	30.13/0.84	21.01/0.64	**33.15**/0.94	32.84/**0.96**
		1/4	25.45/0.51	32.32/0.90	33.67/0.93	33.94/0.94	32.53/0.88	32.26/0.90	23.70/0.74	34.01/0.93	**34.89/0.98**
		1/3	27.25/0.56	34.60/0.94	35.27/0.94	36.37/0.96	35.15/0.92	34.49/0.94	28.05/0.86	34.67/0.91	**36.98/0.98**
	10%	1/5	22.18/0.35	24.69/0.49	25.44/0.59	25.52/0.73	25.98/0.68	25.02/0.73	19.17/0.47	27.43/0.84	**28.03/0.91**
		1/4	22.38/0.36	25.16/0.49	25.96/0.61	26.13/0.70	26.38/0.66	25.53/0.74	20.19/0.57	28.71/0.87	**28.96/0.93**
		1/3	22.37/0.37	25.28/0.49	26.50/0.60	26.64/0.74	26.70/0.65	26.71/0.75	22.91/0.69	29.37/0.86	**29.58/0.94**
2D random	0%	1/5	24.91/0.49	31.69/0.89	33.81/0.93	34.24/0.94	31.95/0.86	31.79/0.89	27.00/0.80	32.89/0.92	**34.47/0.97**
		1/4	25.30/0.50	32.79/0.90	34.97/0.94	35.61/0.95	32.85/0.86	32.94/0.91	29.77/0.86	35.31/0.94	**35.52/0.98**
		1/3	26.32/0.53	34.93/0.93	36.31/0.95	**37.71**/0.96	35.33/0.91	35.10/0.94	33.71/0.92	35.89/0.95	**37.39/0.98**
	10%	1/5	22.37/0.37	24.97/0.51	25.42/0.61	25.90/0.73	25.97/0.67	25.78/0.75	21.67/0.57	**28.61/0.92**	28.58/**0.92**
		1/4	22.38/0.36	24.92/0.49	25.84/0.60	26.06/0.74	26.15/0.67	26.31/0.75	23.78/0.66	**28.83/0.92**	28.73/**0.92**
		1/3	22.37/0.37	24.91/0.47	26.14/0.56	26.38/0.72	26.41/0.62	26.48/0.76	24.87/0.71	29.21/**0.93**	**29.33/0.93**
1D random	0%	1/5	22.78/0.61	25.22/0.75	28.53/0.85	28.79/0.87	28.73/0.86	27.21/0.81	28.65/0.85	30.77/0.92	**31.54/0.95**
		1/4	23.06/0.62	25.77/0.76	28.99/0.87	29.37/0.88	29.06/0.86	27.47/0.82	31.72/0.92	**32.18**/0.90	32.08/**0.96**
		1/3	23.86/0.65	27.34/0.81	32.18/0.91	31.25/0.91	30.98/0.89	30.09/0.86	32.73/0.95	**33.77**/0.95	33.65/**0.97**
	10%	1/5	20.72/0.27	22.38/0.39	22.59/0.40	22.22/0.61	24.52/0.60	22.76/0.67	23.17/0.67	26.37/0.82	**27.10/0.89**
		1/4	20.37/0.26	22.25/0.37	22.98/0.44	22.72/0.63	24.39/0.56	23.32/0.69	25.15/0.75	26.58/0.83	**27.18/0.90**
		1/3	20.37/0.28	22.59/0.37	23.96/0.47	23.75/0.62	24.98/0.58	23.93/0.70	26.58/0.79	**28.14**/0.85	27.87/**0.91**

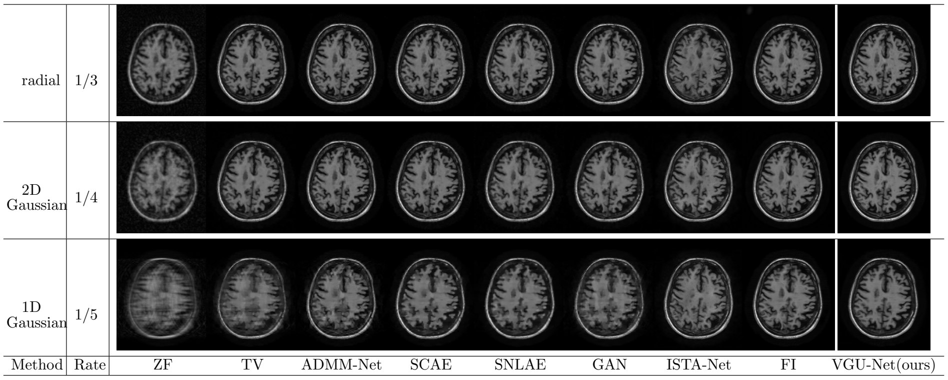

FIGURE 9. MRI reconstruction results from noiseless data with radial, 2D Gaussian, 1D Gaussian mask of sampling ratio 1/3, 1/4 and 1/5 respectively.

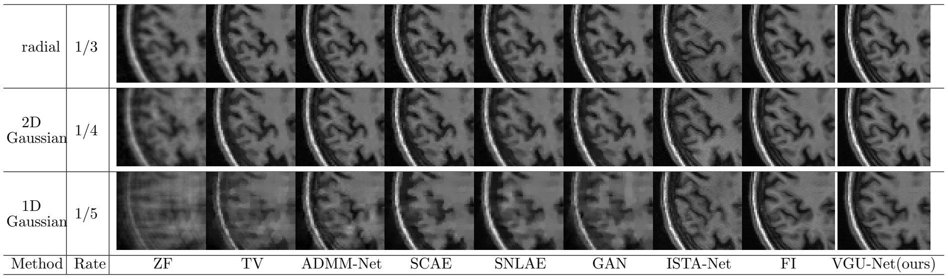

FIGURE 10. Zoom-in results of Figure 9. The last two columns exhibit the best reconstruction quality based on visual assessment.

with a greater degree of generalization prowess. Such problem might be solved by specially designed GNN models, such as neural message passing based on particle system theory or harmonic analysis [24, 41]. These innovative approaches may offer solutions to overcome the limitations and achieve more powerful generalization abilities for graph-based models.

TABLE 3. Ablation experiments (PSNR/SSIM). the best results are in bold.

Method	Noise	Rate	DnCNNs	UNet	VGU-Net(ours)
radial	0%	1/5	32.64/0.96	32.27/**0.97**	**32.84**/0.96
		1/4	34.59/0.97	34.33/0.98	**34.89/0.98**
		1/3	36.83/0.98	36.62/0.98	**36.98/0.98**
	10%	1/5	27.85/0.90	27.22/0.90	**28.03/0.91**
		1/4	27.93/0.90	27.73/0.91	**28.96/0.93**
		1/3	28.12/0.90	28.22/0.94	**29.58/0.94**
2D random	0%	1/5	34.26/0.9	34.03/0.97	**34.47/0.97**
		1/4	35.44/0.980	33.92/0.97	**35.52/0.98**
		1/3	37.27/0.98	35.71/0.98	**37.39/0.98**
	10%	1/5	27.60/0.90	27.51/0.91	**28.58/0.92**
		1/4	27.56/0.90	27.57/0.91	**28.73/0.92**
		1/3	27.89/0.91	28.22/0.92	**29.33/0.93**
1D random	0%	1/5	30.97/0.95	31.45/0.95	**31.54/0.95**
		1/4	31.37/0.95	31.92/0.96	**32.08/0.96**
		1/3	33.40/0.97	33.28/0.97	**33.65/0.97**
	10%	1/5	25.73/0.88	26.89/0.89	**27.10/0.89**
		1/4	25.82/0.88	26.67/0.89	**27.18/0.90**
		1/3	26.32/0.89	26.95/0.89	**27.87/0.91**

REFERENCES

[1] J. Adler and O. Öktem, Solving ill-posed inverse problems using iterative deep neural networks, *Inverse Problems*, **33** (2017), 124007, 24 pp.

[2] J. Adler and O. Öktem, Learned primal-dual reconstruction, *IEEE Transactions on Medical Imaging*, **37** (2018), 1322-1332.

[3] M. Beauchemin, K. P. B. Thomson and G. Edwards, On the hausdorff distance used for the evaluation of segmentation results, *Canadian Journal of Remote Sensing*, **24** (1998), 3-8.

[4] M. A. Bernstein, S. B. Fain and S. J. Riederer, Effect of windowing and zero-filled reconstruction of MRI data on spatial resolution and acquisition strategy, *Journal of Magnetic Resonance Imaging*, **14** (2001), 270-280.

[5] J. Bruna, W. Zaremba, A. Szlam and Y. LeCun, Spectral networks and locally connected networks on graphs, arXiv preprint, (2013), arXiv:1312.6203.

[6] H. Cao, Y. Wang, J. Chen, D. Jiang, X. Zhang, Q. Tian and M. Wang, Swin-unet: Unet-like pure transformer for medical image segmentation, *ECCV 2022: Computer Vision-ECCV 2022 Workshops*, (2023), 205-218.

[7] N. Carion, F. Massa, G. Synnaeve, N. Usunier, A. Kirillov and S. Zagoruyko, End-to-end object detection with transformers, *European Conference on Computer Vision*,(2020), 213-229.

[8] M. Defferrard, X. Bresson and P. Vandergheynst, Convolutional neural networks on graphs with fast localized spectral filtering, *Advances in Neural Information Processing Systems*, **29** (2016).

[9] Q. Ding and X. Zhang, MRI reconstruction by completing under-sampled K-space data with learnable fourier interpolation, *International Conference on Medical Image Computing and Computer-Assisted Intervention*, (2022), 676-685.

[10] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, M. Deghghani, M. Minderer, G. Heigold and S. Gelly, et al., An image is worth 16x16 words: Transformers for image recognition at scale, *International Conference on Learning Representations*, (2021).

[11] F. Falck, C. Williams, D. Danks, G. Deligiannidis, C. Yau, C. C. Holmes, A. Doucet and M. Willetts, A multi-resolution framework for U-Nets with applications to hierarchical VAEs, *Advances in Neural Information Processing Systems*, **35** (2022), 15529-15544.

[12] K. Hammernik, T. Klatzer, E. Kobler, M. P. Recht, D. K. Sodickson, T. Pock and F. Knoll, Learning a variational network for reconstruction of accelerated MRI data, *Magnetic Resonance in Medicine*, **79** (2018), 3055-3071.

[13] Y. Han and J. C. Ye, Framing U-Net via deep convolutional framelets: Application to sparse-view ct, *IEEE Transactions on Medical Imaging*, **37** (2018), 1418-1429.

[14] K. Hara, D. Saito and H. Shouno, Analysis of function of rectified linear unit used in deep learning, *2015 International Joint Conference on Neural Networks (IJCNN)*, (2015), 1-8.

[15] K. He, X. Zhang, S. Ren and J. Sun, Deep residual learning for image recognition, *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, (2016), 770-778.

[16] R. He, W. Zheng, T. Tan and Z. Sun, Half-quadratic-based iterative minimization for robust sparse representation, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **36** (2013), 261-275.

[17] H. Huang, L. Lin, R. Tong, H. Hu, Q. Zhang, Y. Iwamoto, X. Han, Ye. Chen and J. Wu, Unet 3+: A full-scale connected unet for medical image segmentation, *ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, (2020), 1055-1059.

[18] C. M. Hyun, H. P. Kim, S. M. Lee, S. Lee and J. K. Seo, Deep learning for undersampled MRI reconstruction, *Physics in Medicine & Biology*, **63** (2018), 135007.

[19] S. Ioffe and C. Szegedy, Batch normalization: Accelerating deep network training by reducing internal covariate shift. *International Conference on Machine Learning*, (2015), 448-456.

[20] F. Jia, J. Liu and X. Tai, A regularized convolutional neural network for semantic image segmentation, *Analysis and Applications*, **19** (2021), 147-165.

[21] D. P Kingma and J. Ba, Adam: A method for stochastic optimization, *ICLR (Poster)*, (2015).

[22] T. N. Kipf and M. Welling, Semi-supervised classification with graph convolutional networks, *International Conference on Learning Representations*, (2017).

[23] J. Liu, T. Kuang and X. Zhang, Image reconstruction by splitting deep learning regularization from iterative inversion, *International Conference on Medical Image Computing and Computer-Assisted Intervention*, (2018), 224-231.

[24] X. Liu, B. Zhou, C. Zhang and Y. G. Wang, Framelet message passing, arXiv preprint, (2023), `arXiv:2302.14806`.

[25] Z. Liu, Y. Lin, Y. Cao, H. Hu, Y. Wei, Z. Zhang, S. Lin and B. Guo, Swin transformer: Hierarchical vision transformer using shifted windows, *Proceedings of the IEEE/CVF International Conference on Computer Vision*, (2021), 10012-10022.

[26] Y. Lu, Y. Chen, D. Zhao and J. Chen, *Graph-FCN for Image Semantic Segmentation*, Advances in Neural Networks-ISNN 2019, 2019.

[27] M. Lustig, D. Donoho and J. M. Pauly, Sparse MRI: The application of compressed sensing for rapid MR imaging, *Magnetic Resonance in Medicine: An Official Journal of the International Society for Magnetic Resonance in Medicine*, **58** (2007), 1182-1195.

[28] U. von Luxburg. A tutorial on spectral clustering, *Stat. Comput.*, **17** (2007), 395–416.

[29] M. Mardani, E. Gong, J. Y. Cheng, S. S. Vasanawala, G. Zaharchuk, L. Xing and J. M. Pauly, Deep generative adversarial neural networks for compressive sensing MRI, *IEEE Transactions on Medical Imaging*, **38** (2018), 167-179.

[30] F. Milletari, N. Navab and S.-A. Ahmadi, V-net: Fully convolutional neural networks for volumetric medical image segmentation, *2016 Fourth International Conference on 3D vision (3DV)*, (2016), 565-571.

[31] C. Mou, J. Zhang and Z. Wu, Dynamic attentive graph learning for image restoration, *Proceedings of the IEEE/CVF International Conference on Computer Vision*, (2021), 4328-4337.

[32] O. Oktay, J. Schlemper, L. L. Folgoc, M. Lee, M. Heinrich, K. Misawa, K. Mori, S. McDonagh, N. Y. Hammerla and B. Kainz, et al., Attention u-net: Learning where to look for the pancreas, arXiv preprint, (2018), `arXiv:1804.03999`.

[33] O. Petit, N. Thome, C. Rambour, L. Themyr, T. Collins and L. Soler, U-net transformer: Self and cross attention for medical image segmentation, *International Workshop on Machine Learning in Medical Imaging*, (2021), 267-276.

[34] A. M. Rekavandi, S. Rashidi, F. Boussaid, S. Hoefs and E. Akbas, et al., Transformers in small object detection: A benchmark and survey of state-of-the-art, arXiv preprint, (2023), `arXiv:2309.04902`.

[35] O. Ronneberger, P. Fischer and T. Broxm U-net: Convolutional networks for biomedical image segmentationm *International Conference on Medical Image Computing and Computer-Assisted Intervention*, (2015), 234-241.

[36] J. Schlemper, J. Caballero, J. V. Hajnal, A. N. Price and D. Rueckert, A deep cascade of convolutional neural networks for dynamic MR image reconstruction, *IEEE transactions on Medical Imaging*, **37** (2017), 491-503.

[37] J. Shi and J. Malik, Normalized cuts and image segmentation, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **22** (2000), 888-905.

[38] J. Sun, H. Li and Z. Xu, et al., Deep ADMM-Net for compressive sensing MRI, *Advances in Neural Information Processing Systems*, **29** (2016).

[39] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser and I. Polosukhin, Attention is all you need, *Advances in Neural Information Processing Systems*, **30** (2017).

[40] P. Veličković, G. Cucurull, A. Casanova, A. Romero, P. Liò and Y. Bengio, Graph attention networks, *International Conference on Learning Representations*, (2018).

[41] Y. Wang, K. Yi, X. Liu, Y. G. Wang and S. Jin, ACMP: Allen-Cahn message passing with attractive and repulsive forces for graph neural networks, *The Eleventh International Conference on Learning Representations*, (2022).

[42] Z. Wang, A. C. Bovik, H. R. Sheikh and E. P. Simoncelli, Image quality assessment: From error visibility to structural similarity, *IEEE Transactions on Image Processing*, **13** (2004), 600-612.

[43] K. Xu, W. Hu, J. Leskovec and S. Jegelka, How powerful are graph neural networks? *International Conference on Learning Representations*, (2018).

[44] J. C. Ye, Y. Han and E. Cha, Deep convolutional framelets: A general deep learning framework for inverse problems, *SIAM Journal on Imaging Sciences*, **11** (2018), 991-1048.

[45] J. Zeng, J. Pang, W. Sun and G. Cheung, Deep graph laplacian regularization for robust denoising of real images, *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, (2019).

[46] J. Zhang and B. Ghanem, ISTA-net: Interpretable optimization-inspired deep network for image compressive sensing, *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, (2017).

[47] K. Zhang, W. Zuo, Y. Chen, D. Meng and L. Zhang, Beyond a Gaussian denoiser: Residual learning of deep CNN for image denoising, *IEEE Transactions on Image Processing*, **26** (2017), 3142-3155.

[48] L. Zhang, X. Li, A. Arnab, K. Yang, Y. Tong and P. H. Torr, Dual graph convolutional network for semantic segmentation, *30th British Machine Vision Conference 2019, BMVC 2019, Cardiff, UK*, (2019).

[49] X. Zheng, Y. Wang, G. Wang and J. Liu, Fast and robust segmentation of white blood cell images by self-supervised learning, *Micron*, **107** (2018), 55-71.

[50] Z. Zhou, M. M. R. Siddiquee, N. Tajbakhsh and J. Liang, Unet++: A nested u-net architecture for medical image segmentation, *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support*, (2018), 3-11.

Received August 2023; 1st and final revision October 2023; early access November 2023.