Applied Soft Computing 133 (2023) 109905

Contents lists available at ScienceDirect

Applied Soft Computing

journal homepage: www.elsevier.com/locate/asoc

Multi-scale random walk driven adaptive graph neural network with dual-head neighboring node attention for CT segmentation

Ping Xuan [a,b], Xixi Wu [a], Hui Cui [c], Qiangguo Jin [d], Linlin Wang [e], Tiangang Zhang [f,*], Toshiya Nakaguchi [g], Henry B.L. Duh [c]

[a] School of Computer Science and Technology, Heilongjiang University, Harbin, China
[b] Department of Computer Science, School of Engineering, Shantou University, Shantou, China
[c] Department of Computer Science and Information Technology, La Trobe University, Melbourne, Australia
[d] School of Software, Northwestern Polytechnical University, Xi'an, China
[e] Department of Radiation Oncology, Shandong Cancer Hospital and Institute, Shandong First Medical University and Shandong Academy of Medical Sciences, Jinan, China
[f] School of Mathematical Science, Heilongjiang University, Harbin, China
[g] Center for Frontier Medical Engineering, Chiba University, Chiba, Japan

ARTICLE INFO

Article history:
Received 4 May 2022
Received in revised form 17 November 2022
Accepted 29 November 2022
Available online 1 December 2022

Keywords:
Multi-scale random walk
Adaptive graph neural network
Volumetric CT segmentation
Neighboring node attention
Graph-wise attention

ABSTRACT

Segmenting objects with indistinct boundaries and large variations from CT volumes is a challenging issue due to overlapping intensity distributions from neighboring tissues or long-distance semantic relations. We propose a multi-scale random walk (RW) driven graph neural network (GNN) to address this issue. A graph is first initialized to represent image regions and deep semantic features from the segmentation encoder by graph nodes and attributes. We then propose a multi-scale graph reasoning model where for each scale, graph node attribute embedding is obtained by an adaptive GNN with dual-head neighboring node attention, while graph topology is evolved by RW. The neighboring-node attention mechanism is designed to learn and incorporate the importance and influence of neighboring nodes on their connected nodes. Random walking to multi-order neighbors enhance the contextual information formulation and diffusion along graph edges. Finally, multi-scale knowledge learnt from graphs is adaptively fused by a new graph-wise attention fusion module before reshaping and feeding to the segmentation decoder. We evaluate the contributions of major innovations by ablation studies, comparison with other state-of-the-art models on public kidney and tumor segmentation dataset. The generalization ability of our model is validated by different segmentation backbones. Experimental results show that the novel multi-scale adaptive graph reasoning architecture and RW-enhanced GNN model improved the segmentation of objects from adjacent tissues.

1. Introduction

Kidney cancer is one of the most common malignant tumors worldwide with an increasing number of diagnosed cases each year [1]. Computed tomography (CT) imaging is widely used for kidney cancer diagnosis and pre-surgery assessment [2]. Accurate segmentation of kidney organs and tumor regions from three dimensional (3D) CT volumes are essential steps in cancer diagnosis and radiotherapy treatment planning. However, manual detection and delineation of kidney and cancer are time-consuming and subject to operators such as clinicians and radiologists. Thus, computerized methods to assist the automated detection and boundary delineation are of great necessity to improve the treatment efficiency and effectiveness.

Accurate automated kidney and tumor segmentations are challenging. The primary reason is that the tumor may exhibit inside or near the kidney boundary while having similar intensity distributions and appearance with the kidney. Besides, there are large variations between the locations, shapes, and sizes of tumors across different patients. Left and right kidneys have similar texture patterns and shapes, but they locate in distant regions in images. Thus, modeling long-distance dependence between two kidneys and contextual connection between kidney and tumor would help detect and segment kidneys and tumors.

Machine learning methods have been widely used in fields such as data science, image classification, and object recognition. Conventional classification is by defining and extracting various features, performing feature selection using statistical or feature engineering approaches, and finally feeding the selected descriptive features to a machine learning model. For instance, Bahaeddin et al. proposed a series of Artificial Algae Algorithm

* Corresponding author.
E-mail address: zhang@hlju.edu.cn (T. Zhang).

https://doi.org/10.1016/j.asoc.2022.109905
1568-4946/

for feature selection [3] and clustering analysis [4], and training neural networks [5]. However, a well-known challenge with the conventional classification approach is that the feature definition and selection process are mostly hand-crafted yet net generic. Recently, deep learning techniques, including convolutional neural networks (CNNs) received intensive research and industry attention because CNN can perform feature extraction and optimization in an end-to-end approach. CNN generates multi-depth discriminative features that can be learned and optimized during training, especially when there is massive data.

CNN based models have been widely used in a variety of image segmentation tasks [6–9], including kidney and tumor segmentation [10,11]. However,CNN-based segmentation models mainly focus on local, regional information of the image, making it challenging to capture the complex long-distance relationships between multiple image regions. There are some methods proposed to use multiple layers of knowledge propagation in the segmentation problem [12]. To introduce the global context of the image, a popular solution is to enlarge the perception field of convolution. For instance, DenseASPP [13] and PSPNet [14] expand the range of spatial location sampling to integrate a broader content of contextual connections between image regions to improve the segmentation results. A deformable convolutional network [15] is proposed, which learns the displacement of the convolution sampling positions to capture the context information within broader areas. Squeeze-and-Excitation network [16] uses global average pooling and full connection to integrate the global contextual features of the image. Self-attention mechanism [17], double attention networks [18], dual attention network [19], and non-local network [20] calculate the correlation between two positions or channels in the feature map to capture the global dependence of positions or channels. Recurrent neural network (RNN) models [21,22] encode and integrate contextual connections between spatial locations. However, this type of method is difficult to learn the spatial dependence and complex relations between distant image regions.

Recently, graph-based reasoning models have attracted increasing attention in various tasks, including node classification [23–25] and link prediction [26,27]. That is because graph models are flexible and dynamic in modeling multi-sourced heterogeneous relations such as disease and protein [28,29]. For instance, graph neural networks (GNN) [30,31] can deeply integrate the topological structure of the graph and the attributes of each graph node. Graph attention network (GAT) [32] can learn from neighboring node attributes by self-attention mechanism. Since the graph structure and node connections enable knowledge propagation and reasoning between local and long-range regions without the limitations of rigid feature space, graph models have also been used in image segmentation and classification tasks [33]. Existing graph-based models mainly focus on learning and updating node attributes during the training process. However, knowledge propagation and image feature representations in graphs are affected by node connections, also known as graph topology. Thus, static node connections and single-scale knowledge propagation are not sufficient in effective graph-based reasoning, especially for image segmentation tasks that require the modeling of long-distant regional relations and contextual dependencies.

We propose a new multi-scale random walk (RW) enhanced graph reasoning model to address the challenges and improve the tumor and kidney segmentation results. Traditional GNN can only learn the semantic features of neighbor nodes with specified distance. However, the semantic features of close neighbor nodes and distant nodes have different effects on image region nodes. Our new topological embedding strategy based on random walks can integrate the topological information of neighbors in different ranges of nodes, thereby forming multi-scale neighbor topological embeddings. Our model uses graph nodes and node attribute vectors to represent image regions and corresponding semantic features learnt by the segmentation encoder. Afterwards, the long-distance dependence and contextual relations between nodes are derived from the graph by evolving and integrating multi-scale neighbor topologies during the learning process. Our primary technical innovations are summarized below.

Firstly, we propose a new adaptive GNN with dual-head neighboring node attention to enhance node attribute embedding. Traditional GNNs use the same weights to integrate the information learned between multiple heads, which cannot maximize the use of important node attributes. Our new RW-based topological embedding strategy can integrate topological information of neighbors in different nodes ranges to generate multiscale neighbor topological embeddings. Our new neighboring-node attention mechanism can learn and incorporate the importance and influence of adjacent nodes on connected nodes, to enhance the contextual information formulation and diffusion along graph edges. Besides, different from multi-head attention in conventional GNN which concatenates different heads by equal weights, our adaptive GNN has a new head-level attention mechanism to learn the adaptive wights of two heads for the final node attribute embedding.

Secondly, we propose a multi-scale graph reasoning architecture where for each scale, graph node attribute embedding is learnt by adaptive dual-head GNN, and graph topology is evolved by RW. Traditional GNNs can only learn the features of neighboring nodes at a given distance, but the spatial relationships between nodes at different scales are different. Thus, we propose GNN at multiple scales to capture various spatial connections between nodes. The multi-scale RW based node topology embedding can assist GNN to integrate the information of these different perspectives and better encode the information of each image node. For the first time, RW is exploited to update graph topology in GNN and form multi-scale topological information. By performing RW on graph edges in multiple orders, information is carried by the walker to multi-order neighbors along evolved graph edges to enable diverse and broader knowledge propagation.

Finally, a graph-wise attentional fusion module is designed to adaptively integrate the knowledge learnt from multiple graphs. Ablation study results demonstrated the effectiveness and contributions of each of the innovations. The generalization ability of our new multi-scale RW enhanced GNN model is validated by using different segmentation backbones. The experimental results demonstrated the improved performance over the state-of-the-art methods for kidney and cancer segmentation.

In the following Section 2, we present the related graph based models for image segmentation and methods designed for kidney and tumor segmentation. Our proposed new model is given in Section 3, with experimental results and discussions in Section 4. Finally, Section 4 concludes the paper and future work.

2. Related work

2.1. Graph-based image segmentation

Graph based models have been widely used for node classification [23,24], link prediction [26,27] and graph classification [25]. Since the nodes in the graph can spread information, graph reasoning is also introduced into visual recognition tasks [23,34–36]. Conventional graph-based models for image segmentation construct graphs by grid structures where each pixel corresponds to a node. Cui et al. proposed a polymorphic graph model with three types of edges to reflect the similarities and relations between

nodes [33] for CT based lymph node classification. A method based on Random Walk (RW) [37] and conditional random field (CRF) was proposed to integrate the correlation between nodes in the image for lung tumor segmentation from CT volumes. However, the conventional graph-based models mentioned above are handcrafted for specific tasks using shallow features such as the absolute differences between intensity values and prior knowledge formulated by Gaussian distributions. With the increasing complexity of segmentation tasks, satisfactory results and generalization ability cannot be guaranteed.

GNN [29,30] can deeply integrate the topological structure of the graph and attributes of each graph node. Thus, through the dissemination and aggregation of the node-wise information in the graph, the topological structure of the graph and the attribute information of the graph nodes are integrated to infer the connections between the long-distance areas in images. In existing methods GNN based image segmentations, the first step is to construct a fully connected graph composed of a small number of nodes by projecting the image nodes to an interaction space. The nodes in the image conduct relational reasoning based on mutual dissemination and aggregation of node information. For instance, Saha et al. [38] proposed a GNN based model for detecting COVID-19 from CT, which represented CT using an undirected graph and considered only edge information. GNN based refinement was used for airway extraction from 3D CT [39]. However, those two methods are not end-to-end optimized, which rely on handcrafted features as input to the graph. Garcia-Uceda et al. [40] designed a U-Net GNN architecture by replacing the deepest convolutional layer of 3D U-Net with a GNN module. However, this method can only learn the semantic features of the specified distance neighbor nodes. Another challenging issue with GNN is that the relation-sensitive features need to be projected back to the original coordinate space for subsequent image segmentation. The projection process, however, results in losing the original spatial relations between nodes. Li et al. [24] proposed to directly reduce the dimensionality of the features of the nodes in the original coordinate space, and then perform relational learning based on graph convolution. DGC-Seg [41] proposed a graph convolutional autoencoder based to dynamically update graph topology to improve the segmentation performance of prostate from CT.

2.2. Kidney cancer segmentation from CT

Kidney tumors usually have different sizes, irregular shapes, and the boundary between the kidney and the tumor is relatively blurry. Therefore, it is challenging to accurately delineate the kidney and tumor boundaries from 3D CT volumes. A method based on 3D CNN and pyramid pooling is proposed for kidney and tumor segmentation [42]. This method firstly cropped regions of interest from the whole CT volume before segmentation, which is difficult to be applied directly during the clinical treatment and diagnosis process. Hu et al. leveraged boundary information to enhance the segmentation performance [43]. Fabian et al. [44] proposed nnU-Net based on 3D U-Net, which unified data pre-processing, network selection, training and post-processing procedures. nnU-Net achieved the best kidney and tumor segmentation results in the 2019 MICCAI competition. MSSU-net [10] is further proposed using nnU-net as the backbone and brought in the idea of multi-scale supervision to the decoder. RAU-Net [45] further improved U-Net architecture by residual learning and gated attention mechanism. VNet-AG-DSV [46] is later developed by introducing deep self-supervision strategies. A graph convolutional network (GCN) based model, DGC-Seg [41], integrated graph topology and node attribute information to U-Net architecture. Although DGC-Seg improved the segmentation performance, it cannot fully extract multi-scale neighbor topology. Image regional nodes contain neighbors reachable by one hop and adjacent nodes reachable by multiple hops, forming neighboring topologies of multi-scales. The previous methods did not fully utilize and integrate the multi-scale neighboring topology.

3. Method

The proposed multi-scale RW enhanced adaptive GNN model for volumetric CT segmentation is given in Fig. 1. As shown, given input CT volumetric image and output features from segmentation encoder, our new model consists of four major components. The first component is a new graph initialization strategy to represent the learnt image features by graph nodes and edges. In our graph, each node corresponds to an image region and has an attribute vector containing the learnt semantic features by the segmentation encoder. To embed the importance and semantic influence of neighboring nodes, our second innovation is an adaptive GNN model with dual-head neighboring node attention (dn_atten). The adaptive GNN with dn_atten takes node attribute vectors and nodes connections as input and gives a node embedding matrix as output. The third component is graph edge and similarity matrix evolution by RW to dynamically optimize graph edges and propagate information across multi-scale neighboring nodes along edges. We update the initial graph using the new similarity matrix and feed the graph with evolved topology into adaptive GNN with dn_atten module for a new level node embedding matrix extraction. Similarly, we extract another level node embedding matrix. The last major component is a graph-wise attention fusion module to adaptively fuse three-level embedding matrices before reshaping and feeding to the segmentation decoder. The segmentation encoder and decoder can be any 3D segmentation backbones such as 3D UNet and ResNet. In the following sections, we explain each of the major components and innovations.

3.1. Graph initialization

A graph is initialized as $G_0 = (V, \mathbf{A}_0, \mathbf{X}_0, E_0)$, where V denotes graph nodes, $\mathbf{A}_0$ is graph similarity matrix, $\mathbf{X}_0$ denotes graph node attributes,nd E_0 are graph edges. Since each position of output from Seg-Encoder $\mathbf{F} \in \mathbb{R}^{320\times5\times6\times5}$ is related to a specific image region, we reshape $\mathbf{F}$ and obtain $\mathbf{X}_0 \in \mathbb{R}^{150\times320}$ where each row belongs to a graph node v_i, $i = [1,\ 150]$. By such, each graph node v_i can be considered as representing an image region and is associated with an attribute vector $\mathbf{x}_{0i}$. Similarity matrix $\mathbf{A}_0 \in \mathbb{R}^{150\times150}$ between nodes is calculated by following the commonly used similarity measure [41,47] as

$$\mathbf{A}_{0ij} = \exp\{-\|\mathbf{x}_{0i} - \mathbf{x}_{0j}\|_1\} \tag{1}$$

If the similarity between v_i and v_j is ranked as top K = 30 higher among all the similarities in $\mathbf{A}_0$, we define an edge $e_{0ij} \in E_0$ to connect the two nodes as

$$(e_0)_{ij} \textit{ exists if } (\mathbf{A}_0)_{ij} \textit{ is ranked as top K} \tag{2}$$

3.2. Node attribute embedding by adaptive GNN with dual-head neighboring node attention

Given node attribute vector $\mathbf{X}$, we further learn node embedding matrix using a newly proposed adaptive GNN with *dn_atten* neighboring node attention, as shown in Fig. 2(a). Compared with conventional multi-head attention in GNN, our new dual-head attention enhanced adaptive GNN module has two major innovations. Firstly, our GNN has a *dn_atten* neighboring node

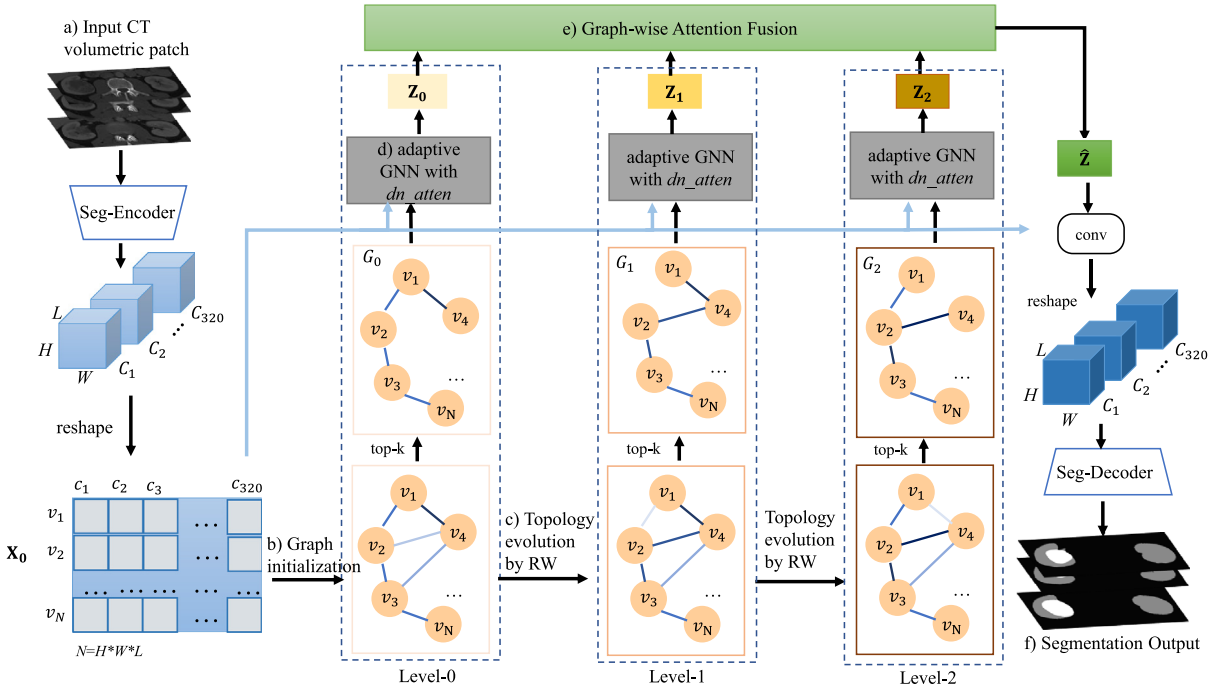

Fig. 1. The framework of Multi-scale random walk driven adaptive graph neural network with dual-head neighboring node attention for CT segmentation. Given (a) input CT volumetric image and output features $\mathbf{X}_0$ from segmentation encoder, (b) a graph is initialized to represent learnt image features by graph nodes and edges. Using initialized graph node attributes and edges as input, (d) an adaptive GNN model with dual-head neighboring node attention (dn_atten) outputs level-0 node embedding matrix $\mathbf{Z}_0$ which embeds the importance and semantic influence of neighboring nodes. Afterwards, the graph edges are dynamically optimized during the training process by (c) RW algorithm. The evolved graph with updated neighboring node connections is fed into adaptive GNN with dn_atten module for level-1 node embedding matrix extraction. Similarly, we extract level-2 node embedding matrix. The multi-level matrices are fused adaptively using (e) an innovative graph-wise attention fusion module. The output $\hat{\mathbf{Z}}$ is concatenated with $\mathbf{X}_0$, fused by 1-D convolutional operation, reshaped and fed to the segmentation decoder to obtain the final segmentation results.

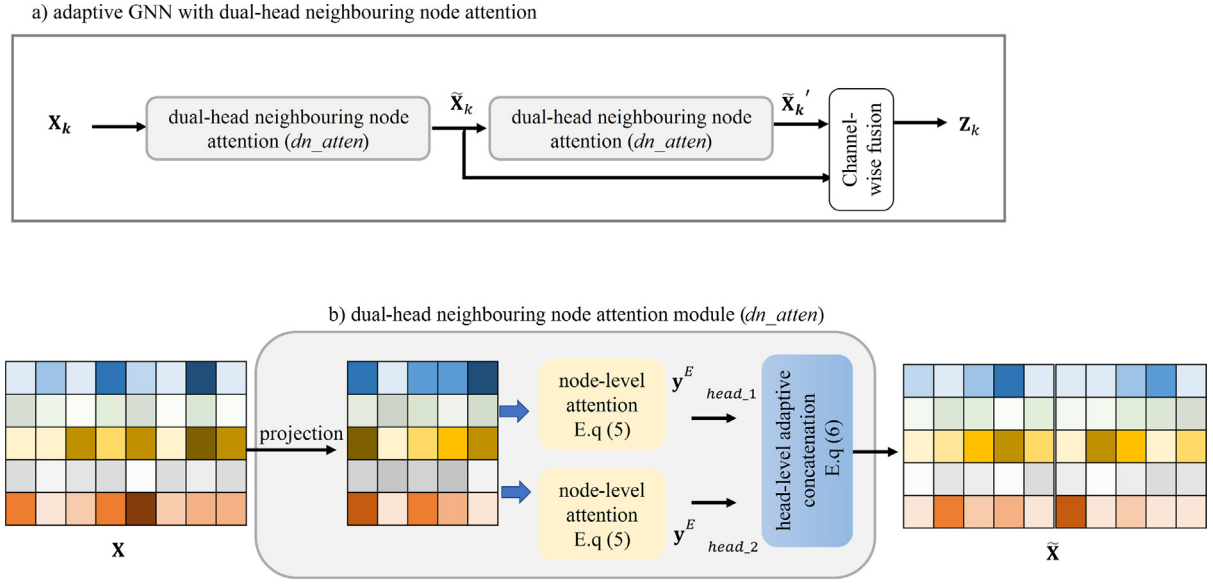

Fig. 2. Adaptive GNN with dual-head neighboring node attention (dn_atten) for node embedding matrix extraction.

attention mechanism (as shown in Fig. 2(b)) which can learn and incorporate the importance and influence of adjacent nodes on v_i to enhance the contextual information diffusion along graph edges. Secondly, unlike conventional GNN that concatenates the outputs from multiple heads using equal weights, our new model can learn adaptive weights of two heads. The contributions of new dual-head attention and adaptive GNN are validated with experimental results given in ablation study.

As shown in Fig. 2(a), let *dn_atten* denotes dual-head neighboring node attention, adaptive GNN takes graph, e.g. $\mathbf{G}_0$ and its graph node attribute $\mathbf{X}_0$ as input and is composed of two *dn_atten* modules $\widetilde{X}_0 = dn\_atten\ (\mathbf{X}_0,\ E_0)$ and $\widetilde{X}_0' = dn\_atten\ (\widetilde{\mathbf{X}}_0,\ E_0)$ to learn deeper image regional nodes features. The outputs from first and second-order *dn_atten* are fused by 1-D convolution to obtain the final node embedding matrix $\mathbf{Z}_0$. The detailed algorithms of

Table 1
Parameter settings and sizes in GNN with dual-head neighboring node attention.

	Input	Node_self_attention	Output size	dn_atten	Output size	Operations	Output size
GNN_0	$\mathbf{X}$: 150 × 320	$\mathbf{W}_{0head_0}$: 320 × 200 $\mathbf{W}_{0head_1}$: 320 × 200	$\mathbf{y}_{0head_0}$: 150 × 200 $\mathbf{y}_{0head_1}$: 150 × 200	$S_{att\,0}$: 200 × 1	$\widetilde{\mathbf{X}}$: 150 × 200	Conv1d, in:300, out:100, K = 1,s = 1	$\mathbf{Z}$: 150 × 100
GNN_1	$\widetilde{\mathbf{X}}$: 150 × 200	$\mathbf{W}_{1head_0}$: 200 × 100 $\mathbf{W}_{1head_1}$: 200 × 100	$\mathbf{y}_{0head_0}$: 150 × 200 $\mathbf{y}_{0head_1}$: 150 × 100	$S_{att\,0}$: 100 × 1	$\widetilde{\mathbf{X}}'$: 150 × 100		

dn_atten and the process of learning node embedding matrix are explained below with parameter settings given in Table 1.

***dn_atten* algorithm.** Given graph G_0 and edges E_0, we first calculate the importance and contextual influence of other nodes v_j on node v_i. Let $\alpha_{ij}^{E_0}$ denote the importance score, $\mathbf{X}_0$ is firstly projected to $\mathbf{X}_0{}'$ as

$$\mathbf{X}_{0i}' = \mathbf{X}_{0i} \cdot W_0 \tag{3}$$

where $'\cdot'$ denotes matrix multiplication, the weight matrix $W_0 \in \mathbb{R}^{320\times200}$ is randomly initialized and learnt during the training process. $\alpha_{ij}^{E_0}$ is then obtained and normalized by

$$\alpha_{ij}^{E_0} = \frac{exp\Big(LeakyRelu\left(a_{E_0}^{\mathrm{T}} \cdot \left[\mathbf{x}_{0i}' \parallel \mathbf{x}_{0j}'\right]\right)\Big)}{\sum_{k\in N_i^{E_0}} exp\Big(LeakyRelu\left(a_{E_0}^{T} \cdot \left[\mathbf{x}_{0i}' \parallel \mathbf{x}_{0k}'\right]\right)\Big)} \tag{4}$$

where *LeakyRelu* denotes activation function, a_{E_0} is node-level attention vector learnt during the training process, T denotes matrix transposition, and $\parallel$ denotes concatenation operation. Therefore, the learned node-level attention enhanced embedding of v_i is obtained as

$$\mathbf{y}_i^{E_0} = LeakyRelu(\sum_{j\in N_i^{E_0}} \alpha_{ij}^{E_0} \cdot \mathbf{X}_{0j}') \tag{5}$$

Inspired by multi-head attention [48], we repeat the above calculation of $\mathbf{X}_i^{E_0}$ twice and obtain two embeddings $\mathbf{y}_{i\ head_1}^{E_0}$ and $\mathbf{y}_{i\ head_2}^{E_0}$. Different from multi-head attention which concatenates two heads by equal weights, we propose head-level attention for the adaptive integration of dual heads as

$$\widetilde{X}_0 = (h_1 \circ \mathbf{y}_{i\ head_1}^{E_0}) \parallel (h_2 \circ \mathbf{y}_{i\ head_2}^{E_0}) \tag{6}$$

where $\circ$ denotes element-wise product operation. h_m, m = 1, 2, is head-level attention score which is obtained by

$$h_{m_{(m=1,2)}} = \frac{\exp(s_m{}^T s_{att})}{\exp\left(s_1{}^T s_{att}\right) + \exp(s_2{}^T s_{att})} \tag{7}$$

where s_{att} is learnt during the training process to capture the contextual relation between two heads, s_m is the informative score of mth head and defined as

$$s_m = \tanh(\mathbf{w}_{att} y_{i\ head_m}^{E_0} + b_{att}) \tag{8}$$

where weight matrix and bias vector $\mathbf{W}_{att}$ and b_{att} are randomly initialized and learnt during the training process. The role of attention score hm calculated by Eq. (7) is to adaptively learn the weight of each head in the multi-head attention during the integration process to obtain a better contextual relationship between the two heads. hm is used as the weight of each head when multi-head attention is integrated by Eq. (6).

To learn multi-scale and deeper information, we further perform *dn_atten* over $\widetilde{\mathbf{X}}_0$ and obtain $\widetilde{\mathbf{X}}_0' = dn\_atten(\widetilde{\mathbf{X}}_0,\ E_0)$. During this round of attention calculation, the projection weight matrix in Eq. (3) is of dimension $\mathbb{R}^{200\times100}$. Finally, node embedding matrix $\mathbf{Z}$ is obtained by feeding $\widetilde{\mathbf{X}}_0$ and $\widetilde{\mathbf{X}}_0'$ into 1*1 convolutional operation for channel-wise fusion as

$$\mathbf{Z}_0 = f_{conv_{1*1}}(\widetilde{\mathbf{X}}_0 \parallel \widetilde{\mathbf{X}}_0') \tag{9}$$

where $\parallel$ denotes concatenation.

3.3. Graph topology evolution by Random Walks (RW)

To learn and update nodes connections that reflect graph topology, we perform RW on the graph. The walkers start from each node with equal probability and travel along graph edges. Upon finishing $t+1$ th hops, we can obtain pairwise similarity between nodes. The iterative RW process and the status $\boldsymbol{\pi}$ at a time $t+1$ can be formulated by

$$\boldsymbol{\pi}(t+1) = (1-\gamma)\,\mathbf{P}^{\mathbf{T}}\boldsymbol{\pi}(t) + \gamma\boldsymbol{\pi}(0) \tag{10}$$

where γ is the probability that the walker restarts from initial status $\boldsymbol{\pi}(0)$; $\boldsymbol{\pi}(0)$ is a identity matrix. $\mathbf{P}$ is the transition probability matrix where $p_{ij} \in \mathbf{P}$ denotes the probability the walker moves to node v_j from v_i. $\mathbf{P}$ is obtained by performing Laplacian operation on $\mathbf{A}_0$ as

$$\mathbf{P} = \mathbf{D}_0^{-1/2}\mathbf{A}_0\mathbf{D}_0^{-1/2} \tag{11}$$

where $(\mathbf{D}_0)_{ii} = \sum_j \mathbf{A}_{0ij}$. After the random walker hops along the graph edges E_0 once, the new similarity between nodes v_i and v_j is calculated by Eq. (1) as

$$\mathbf{A}_{1ij} = \exp\{-\|\boldsymbol{\pi}(\mathbf{1})_i - \boldsymbol{\pi}(\mathbf{1})_j\|_1\} \tag{12}$$

where $\boldsymbol{\pi}(1) = \mathbf{P}^{\mathbf{T}}\boldsymbol{\pi}(0)$ according to Eq. (10).

Given the new similarity matrix $\mathbf{A}_1$, the node connections and graph edges in G_0 are updated, resulting in a new edge set E_1 where

$$(e_1)_{ij} \in E_1 \ exists\ if\ (\mathbf{A}_1)_{ij}\ is\ ranked\ as\ top\ K \tag{13}$$

3.4. Multi-scale graphs and weighted fusion

To learn and integrate information derived from the evolved graph, we perform the above node embedding matrix extraction and RW topology evolution process two more times and obtain multi-scale embedding matrices $\mathbf{Z}_1$ and $\mathbf{Z}_2$. As shown in Fig. 1, using the updated edge set E_1, similarity matrix $\mathbf{A}_1$, nodes V and attributes $\mathbf{X}_0$, we can construct a new graph $G_1 = (V, \mathbf{A}_1, \mathbf{X}_0, E_1)$. By feeding G_1 to adaptive GNN with dn_atten, embedding matrix $\mathbf{Z}_1$ is obtained as

$$\mathbf{Z}_1 = f_{conv_{1*1}}(\widetilde{\mathbf{X}}_1 \parallel \widetilde{\mathbf{X}}_1') \tag{14}$$

where $\widetilde{X}_1 = dn\_atten\ (\mathbf{X}_0,\ E_1)$ and $\widetilde{X}_1' = dn\_atten\ (\widetilde{X}_1,\ E_1)$.

Similarly, the random walker hops on the graph G_1 twice as

$$\boldsymbol{\pi}(2) = (1-\gamma)\,\mathbf{P}^{\mathbf{T}}\boldsymbol{\pi}(1) + \gamma\boldsymbol{\pi}(0) \tag{15}$$

and obtain $\mathbf{A}_2$ and corresponding new edge set E_2. Accordingly, a new graph $G_2 = (V, \mathbf{A}_2, \mathbf{X}_0, E_2)$ is constructed and fed into adaptive GNN with dn_atten to obtain

$$\mathbf{Z}_2 = f_{conv_{1*1}}(\widetilde{\mathbf{X}}_2 \parallel \widetilde{\mathbf{X}}_2') \tag{16}$$

where $\widetilde{X}_2 = dn\_atten\ (\mathbf{X}_0,\ E_2)$ and $\widetilde{X}_2' = dn\_atten\ (\widetilde{X}_2,\ E_2)$.

Inspired by [47], the three embeddings are adaptively fused with various importance as shown in Fig. 3 as

$$\hat{\mathbf{Z}} = \sum_{i=0}^{2} \beta_i \cdot \mathbf{Z}_i + \mathbf{Z}_i \tag{17}$$

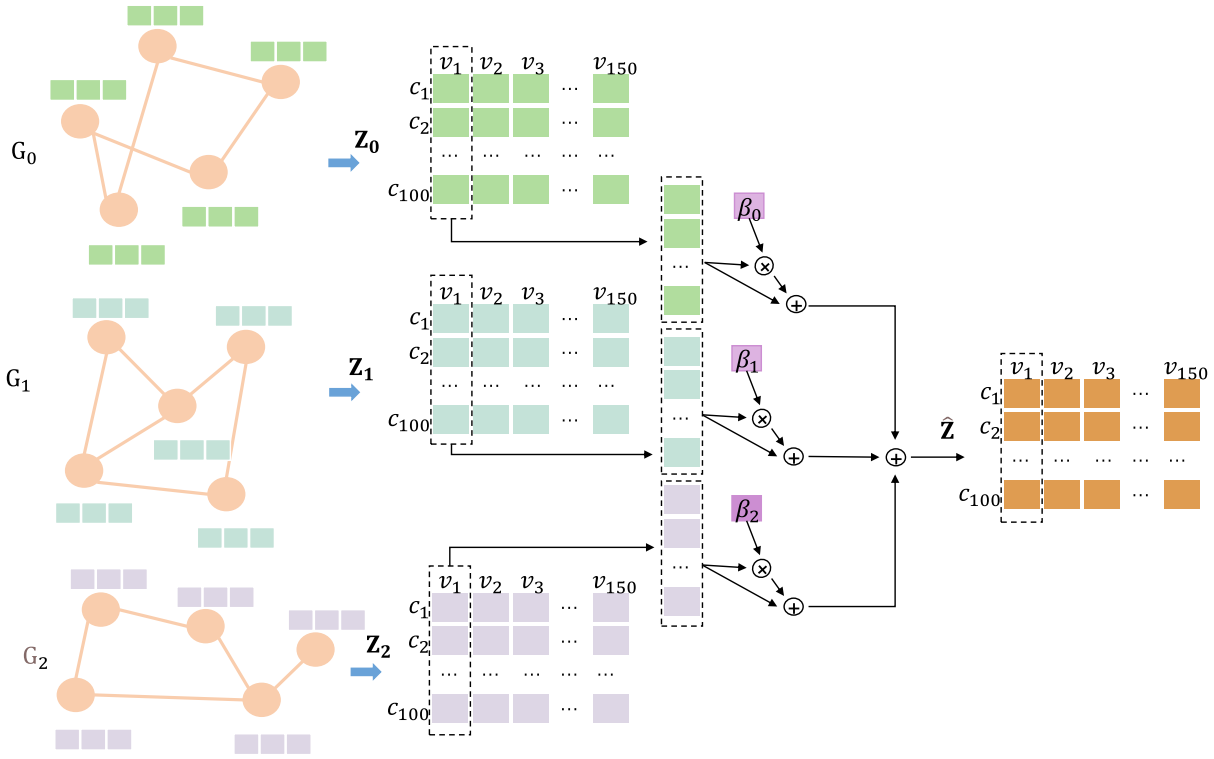

Fig. 3. Illustration of graph-wise attention for multi-scale graph fusion.

where β_i denotes the importance of ith scale,

$$\beta_i = softmax(\mathbf{q}^T \cdot \tanh(\mathbf{W} \cdot (\mathbf{Z}_i)^T + \mathbf{b})) \tag{18}$$

where **q** is a shared attention vector, **W** is weight matrix, and **b** is bias vector. **q**, **W** and **b** are learnt and updated during the training process.

4. Experimental results and discussions

4.1. Experimental settings and dataset

To validate the performance of multi-scale RW enhanced adaptive GNN in modeling semantic relations and long-distance dependence for volumetric image segmentation, we collected 210 patient studies from KiTS19 Challenge Dataset [49] and evaluated the performance using three different 3D segmentation backbones.

Each patient study has CT images and corresponding manually delineated kidney and tumor regions by clinicians (ground truth (GT)). Since the images are from different hospitals and scanners using various protocols, the number of CT slices and pixel spacings vary among patients. We resampled all the images using a unified voxel size of $1 \times 1 \times 1$ mm^3. Training, validation and testing sets are randomly generated using 60%, 20% and 20% of the entire dataset, resulting in 127 patient studies in training set, 42 patient studies in validation set, and 41 testing cases. Data augmentation was performed on the training and validation datasets, including random scaling, rotation, mirroring operations, adjusting brightness, and adding gammas noises.

Data augmentation was performed on the training and validation datasets. We performed data augmentations which are exactly the same as nnU-Net framework which are given in [50]. The detailed operations are: (1) Spatial Augmentations mirroring including channel translation (to simulate registration errors), elastic deformations, rotations, scaling, and resampling; (2) Color Augmentations including brightness (additive, multiplication), contrast, gamma (like gamma correction in photo editing); (3) Noise Augmentations including Gaussian Noise, Rician Noise; Cropping: random crop, center crop, and padding.

The segmentation encoder and decoder can be any 3D segmentation backbones. In this work, we chose three well-recognized 3D segmentation networks including 3D U-Net [51], 3D nnU-Net [52] and 3D ResNet. All the models were implemented using PyTorch on a single NVIDIA RTX 2080Ti (11 GB RAM). Patch size was set as $160 \times 192 \times 80$ with a batch size of 2. The learning rate was initialized as $3e^{-4}$ and updated by $\left(1 - \frac{epoch}{total_epoch}\right)^{0.9}$. The optimizer is Adam. The maximum possible epoch is 600; the best model is saved and used for testing.

4.2. Evaluation metrics

The kidney and tumor segmentation results by computerized models are compared with GT and evaluated in terms of volumetric spatial overlap and shape similarities by Dice similarity coefficient (DSC) [53], intersection over union (IoU) [54] and Hausdorff distance (HD) [55].

DSC and IoU are defined as

$$DSC = \frac{2|Vol_{seg} \cap Vol_{GT}|}{|Vol_{seg}| + |Vol_{GT}|} \tag{19}$$

$$IoU = \frac{2|Vol_{seg} \cap Vol_{GT}|}{|Vol_{seg} \cup Vol_{GT}|} \tag{20}$$

where Vol_{seg} denotes the segmentation result and Vol_{GT} represents GT. DSC and IoU are both within the range of 0 and 1, where a greater DSC (IoU) value indicates better segmentation results in terms of spatial overlap. In this work, we can obtain DSC and IoU for kidney and tumor which are denoted by DSC_{kidney}, IoU_{kidney}, DSC_{tumor}, and IoU_{tumor}, respectively for each patient.

Table 2
Ablation study results of multi-scale wise graph reasoning attention by GNN and RW, dn_atten in the new adaptive GNN, and graph-wise attention fusion.

GNN	RW	dn_atten	Graph-wise attention fusion	$Dice_{kidney}$	IoU_{kidney}	HD_{kidney} (mm)	$Dice_{tumor}$	IoU_{tumor}	HD_{tumor} (mm)
×	×	×	×	0.9642	0.9319	24.2812	0.8577	0.7673	25.0405
✓	×	×	×	0.9642	0.9319	22.6055	0.8587	0.7709	21.8470
✓	✓	×	×	0.9651	0.9335	21.3676	0.8636	0.7765	26.3746
✓	✓	✓	×	**0.9652**	**0.9337**	23.0975	0.8647	0.7757	28.9353
✓	✓	✓	✓	0.9645	0.9326	**18.1211**	**0.8725**	**0.7852**	**19.0169**

HD is defined as

$$HD_{tumor}(Vol_{seg\,tumor}, Vol_{GT\,tumor}) = max\{dist(Vol_{seg_{tumor}}, Vol_{GT\,tumor}), dist(Vol_{GT\,tumor}, Vol_{seg_{tumor}})\} \quad (21)$$

where $dist(\,)$ denotes the distance between the surfaces of Vol_{GT} and Vol_{seg}, and is calculated as

$$dist(Vol_{seg_{tumor}}, Vol_{GT\,tumor}) = \max_{a \in Vol_{seg_{tumor}}} \min_{b \in Vol_{GT\,tumor}} \|a - b\| \quad (22)$$

In the experiments, we use HD_{kidney} and HD_{tumor} to denote the HD values obtained for kidney and tumor evaluations. The smaller the HD values, the better the segmentation results.

4.3. Ablation study results

We firstly perform ablation studies to validate the effectiveness and contributions of each of our major technical innovations, including multi-scale graph reasoning by GNN and RW, dn_atten in the new adaptive GNN, and graph-wise attention fusion. The results using nnUNet as backbone are given in Table 2.

As shown in Table 2, the multi-scale graph reasoning by GNN and RW improved the kidney and tumor segmentation results. Compared to the backbone model, GNN improved the tumor segmentation performance with respect to both spatial overlap and shape similarity. By adding multi-scale RW to the graph reasoning process, the model improved kidney and tumor segmentation results for all the evaluation measures except HD_{tumor}. dn_atten further enhanced the performance, especially kidney, with the best DSC_{kidney} of 0.9652 and IoU_{kidney} of 0.9337. By adaptively fusing multi-scale knowledge using graph-wise attention, the kidney and tumor HD results were boosted, especially tumor, leading to the best HD_{kidney} of 18.1211 mm and HD_{tumor} of 19.0169 mm. Although the DSC and IoU of the kidney were lower than the method without using graph-wise attention fusion, the HD was lifted to 18.1211 mm. Besides, the tumor segmentation performance was significantly improved with the best DSC of 0.8725, IoU of 0.7852, and HD of 19.0169 mm.

Six cases are given in Fig. 4. Compared with GT, nnU-Net failed to detect the tumor in cases (c) and (e) and obtained smaller segmented tumor regions in the remaining four cases. Our RW and GNN based graph reasoning architecture contributed to the localization of tumors in (c) and (e). dn_atten in adaptive GNN incorporated the neighboring influence, which enlarged the labeling information propagation area during the segmentation process such as (a), (d) and (e). By adapting multi-scale graph embeddings, our final model achieved the best results with the highest visual similarity to GT as shown in Fig. 4 and quantitative results as demonstrated by Table 2. We explain the findings for two reasons. The primary reason is that the dn_atten module incorporated the importance and influence of neighboring nodes on the connected nodes, which enhanced the contextual information formulation and diffusion along graph edges. As a result, the labeling information can be propagated to a broader area during the training and learning process. The second reason is that our adaptive multi-scale graph architecture improves semantic information propagation and long-range spatial dependence modeling by evolved graph topologies and adaptive attribute embedding fusion.

4.4. Investigations of different 3D segmentation backbones

We also validate the generalization ability of our newly proposed multi-scale RW enhanced GNN for volumetric image segmentation. We use the encoder and decoder from three segmentation networks, including 3D U-Net, 3D nnU-Net and 3D ResNet, to replace the Seg-Encoder and Seg-Decoder in Fig. 1. The quantitative and qualitative results are given in Table 3 and Fig. 5.

As shown by Table 3, our multi-scale RW enhanced adaptive GNN, denoted by ours, consistently improved the segmentation results by different backbones. Among these results, the best DSC and IoU of kidney were achieved using 3D nnU-NNet as the backbone, while the best kidney HD was observed using a 3D U-Net encoder and decoder. The best results in all the evaluation measures for tumor segmentation were achieved when using 3D ResNet as the backbone.

As shown by the 6 cases in Fig. 5, 3D U-Net failed to detect the tumor in cases (a) and (c) and obtained incomplete tumor segmentations in the remaining cases. Using our multi-scale RW enhanced adaptive GNN, 3D U-Net based model achieved better tumor detections and segmentation, especially for case (a), and case (d) and (e)with heterogeneous distributions. For 3D nnU-Net and 3D ResNet backbones, our adaptive multi-scale graph reasoning model contributed to the complete tumor delineations, especially in cases (a), (b), (d) and (e). The results and findings further proved our hypothesis that modeling semantic relations and long-range dependencies between different objects in the image would improve segmentation results, particularly for cases with large variations and overlapping intensity distributions.

4.5. Comparison with other methods

The comparison results with other state-of-the-art CT and kidney and tumor segmentation methods are given in Table 4. As shown, our model achieved the best DSC_{kidney} and IoU_{kidney} of 0.965 and 0.933 and HD_{kidney} of 18.121 mm. DGC [41] and Transformer based models TransUNet [56] and TransBTS [57] achieved the second best DSC_{kidney} of 0.964. Among those three models, UNet based DGC's achieved HD_{kidney} of 23.359 mm which was better than UNet architecture TransUNet with a HD_{kidney} of 25.825 mm and TransBTS's HD_{kidney} of 30.954 mm. Among all the methods with known HD_{kidney}, the second best HDkidney was achieved by another graph model nnU-Net_with_graph_reasoning [23].

Regarding tumor segmentation performance, our model achieved the best DSC_{tumor} of 0.873 and IoU_{tumor} of 0.785, and HD_{tumor} of 19.017 mm. The second-best spatial overlapping with respect to DSC and IoU was obtained by TransBTS with DSC_{tumor} of 0.867 and IoU_{tumor} of 0.777. DGC obtained the same third-best IoU_{tumor} of 0.773 as TransUNet. However, DGC's HD_{tumor}

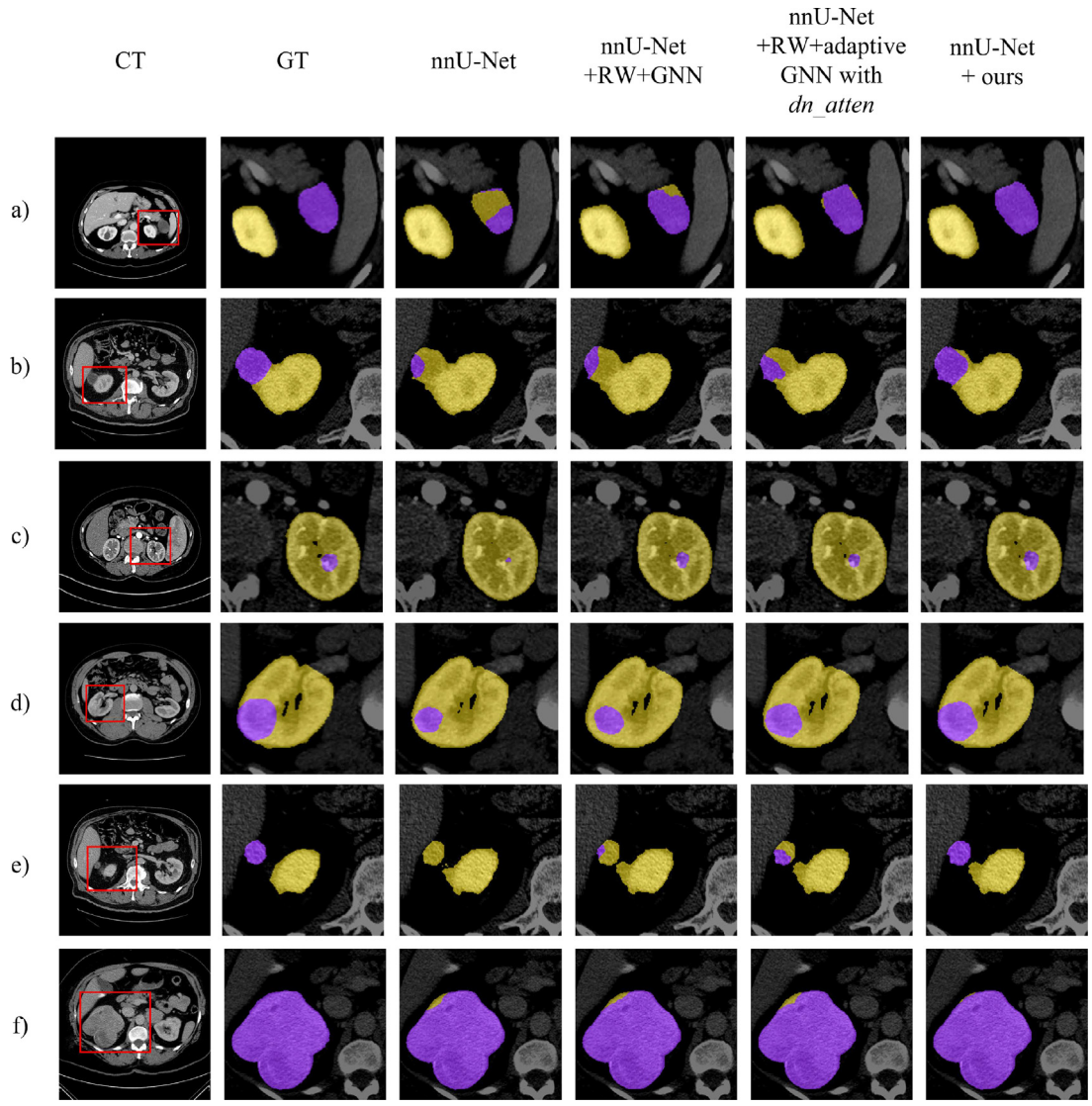

Fig. 4. Ablation study results of six cases. Yellow (purple) colormaps show the manual and automated segmentation of kidneys (tumor). Ours denotes the final model, which is RW+ adaptive GNN with dn_atten + graph wise attention fusion.

Table 3
Experimental results of embedding multi-scale RW enhanced adaptive GNN (denoted by ours) to different segmentation backbones.

Method	$Dice_{kidney}$	IoU_{kidney}	HD_{kidney} (mm)	$Dice_{tumor}$	IoU_{tumor}	HD_{tumor} (mm)
3D U-Net	0.9627	0.9292	24.8746	0.8543	0.7626	24.0168
3D U-Net+ours	0.9630	0.9299	18.2543	0.8623	0.7722	22.4800
3D nnU-Net	0.9642	0.9319	24.2812	0.8577	0.7674	25.0405
3D nnU-Net+ours	**0.9645**	**0.9326**	**18.1211**	**0.8725**	**0.7852**	**19.0169**
3D ResNet	0.9620	0.9277	25.3935	0.8508	0.7600	41.7243
3D ResNet+ours	0.9634	0.9304	21.8472	0.8718	0.7828	21.0062

of 22.773 mm was the second best among all the methods in comparison, followed by 3D U-Net's HD_{tumor} of 24.017 mm.

We also investigated the performance of our model using different segmentation backbones via 5-fold cross-validation. Using nnU-Net backbone, our model consistently achieved the best performance among all the comparing methods with $Dice_{kidney}$ of 0.965, IoU_{kidney} of 0.933, HD_{kidney} of 19.498 mm, $Dice_{tumor}$ of 0.869, IoU_{tumor} of 0.783, and HD_{tumor} of 20.863 mm. Compared with our results using random sampled 80% cases for training and 20% for testing, $Dice_{kidney}$ and IoU_{kidney} remained the same. HD_{kidney} increased by 1.37 mm, $Dice_{tumor}$ improved by 0.0035, IoU_{tumor} decreased by 0.0022, and HD_{tumor} increased by 1.8461 mm. Using U-Net and ResNet as backbones, the 5-fold cross-validation results for Kidney spatial overlap, as reflected by Dice and IoU, are the same as 80%/20% random sampling. At the same time, the shape similarity in terms of HD was slightly higher. For tumor segmentation, the Dice, IoU and HD results did not have statistically significant differences between different evaluation approaches, showing the robustness of our model.

Our main finding is that graph-based models and Transformers with the capacity to learn long-range dependency outperformed other methods in comparison. Compared with Transformer and GCN, our new multi-scale RW enhanced GNN enabled more effective long-distance dependence and contextual relations modeling between image nodes. The second finding is that knowledge propagation and learning in graph based non-rigid like

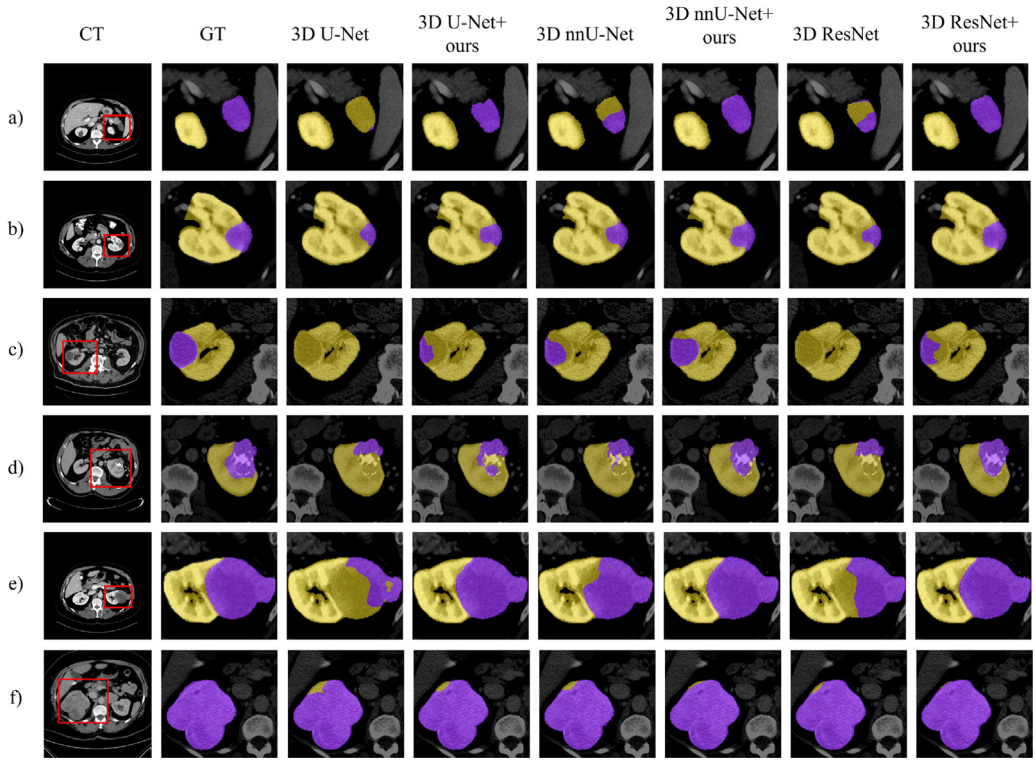

Fig. 5. Six cases of embedding multi-scale RW enhanced adaptive GNN model to different segmentation backbones. Yellow (purple) colormaps show the manual and automated segmentations of the kidney (tumor).

Table 4
Experimental results of comparing with other methods.

Method	$Dice_{kidney}$	IoU_{kidney}	HD_{kidney} (mm)	$Dice_{tumor}$	IoU_{tumor}	HD_{tumor} (mm)
RAU_Net[a] [45]	0.962	–	–	0.768	–	–
2D_PSPNET[b] [14]	0.902	–	–	0.638	–	–
3D_FCN_PPM[b] [42]	0.927	–	–	0.802	–	–
3D U-Net [51]	0.963	0.930	24.875	0.854	0.763	24.017
3D nnU-Net [52]	0.964	0.932	24.281	0.858	0.767	25.041
MSS U-net[a] [10]	0.958	0.920	21.123	0.821	0.720	49.347
nnU-Net_with_graph_reasoning [23]	0.938	0.890	20.131	0.853	0.756	36.574
DGC [41]	0.964	0.932	23.359	0.863	0.773	22.773
TransUNet [56]	0.964	0.931	25.825	0.862	0.773	37.066
TransBTS [57]	0.964	0.931	30.954	0.867	0.777	34.499
PM(3D nnU-Net backbone)	**0.965**	**0.933**	**18.121**	**0.873**	**0.785**	**19.017**
PM(3D nnU-Net backbone)[a]	0.965	0.933	19.498	0.870	0.783	20.863
PM(3D U-Net backbone)	0.963	0.930	18.254	0.862	0.772	22.480
PM(3D U-Net backbone)[a]	0.963	0.930	21.018	0.863	0.771	20.085
PM(3D ResNet backbone)	0.963	0.930	21.847	0.872	0.783	21.006
PM(3D ResNet backbone)[a]	0.963	0.930	24.763	0.868	0.782	21.501

[a]Denotes 5-fold cross validation.
[b]Denotes random sampling using 60% for training and 40% for testing.

image feature space has the capability to boost the segmentation performance especially the shape similarity as reflected by HD.

4.6. Execution complexity

To investigate the complexity of execution space and time of our newly proposed multiscale RW enhanced adaptive GNN, we calculated the GPU memory allocation, floating point operations per second (FLOPs), parameters, average training time per epoch, and inference time. As shown by Table 5, using nnU-Net as backbone, the GPU memory and FLOPs allocated by our embedded model are 8911 MB and 1252.41, respectively. The number of parameters is 31.0375 M. The average training time per epoch is 510.04 s. Removing multiscale topology evolution by RW, the GPU memory and FLOPs allocations decrease to 8845 MB and 1252.29, respectively. The number of parameters is 31.0344 M. The average training time per epoch is 506.37s. Therefore, the GPU memory occupied by our RW module is 66MB, the FLOPS is 0.12, the parameter size is 0.0031M, and the average training time per epoch is 3.67s. By further removing GNN module, we can obtain the execution complexity of GNN as occupying 2027MB GPU memory, 0.1 FLOPS, 0.2484M parameters, and 29.1s average training time per epoch.

5. Conclusion and future works

We present a new multi-scale RW enhanced adaptive GNN model for tumor and kidney segmentation from 3D CT. The novel GNN with dual-head neighboring node attention enhances node

Table 5
Complexity analysis of multiscale RW driven adaptive GNN.

GNN	RW	Memory (GPU) (MB)	FLOPs	Parameters (M)	Average training time (seconds per epoch)
×	×	6818	1252.19	30.7860	477.27
✓	×	8845	1252.29	31.0344	506.37
✓	✓	8911	1252.41	31.0375	510.04

attribute embedding by incorporating the importance of neighboring nodes. Multi-scale graph architecture and topology evolution by RW enable the diverse range knowledge propagation and effective graph reasoning. The ablation studies, comparison with other state-of-the-art methods, and investigations on different segmentation backbones show that our graph model improved the kidney and tumor segmentation performance, especially for tumor cases with large variations and overlapping intensity distributions. Our future work includes extending the new GNN algorithm to other fully supervised medical image segmentation tasks, and semi-supervised segmentation tasks where there are limited number of labels.

CRediT authorship contribution statement

Ping Xuan: Designed the method and participated in manuscript writing. **Xixi Wu:** Designed the experiments and edited the manuscript. **Hui Cui:** Participated in method design and manuscript writing. **Qiangguo Jin:** Participated in method design. **Linlin Wang:** Participated in experiment design. **Tiangang Zhang:** Participated in method design and manuscript writing. **Toshiya Nakaguchi:** Participated in experiment design. **Henry B.L. Duh:** Participated in method design.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgments

This work is supported by the Natural Science Foundation of China (61972135, 62172143, 82172865, 62201460), STU Scientific Research Initiation Grant (NTF22032), the Natural Science Foundation of Heilongjiang Province (LH2019F049), and the China Postdoctoral Science Foundation (2019M650069, 2020M670939). All authors contributed to the article and approved the submitted version.

Appendix A. Supplementary data

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.asoc.2022.109905.

References

[1] W.-H. Chow, L.M. Dong, S.S. Devesa, Epidemiology and risk factors for kidney cancer, Nat. Rev. Urol. 7 (5) (2010) 245–257.
[2] P.V. Glybochko, Y.G. Alyaev, S.B. Khokhlachev, D.N. Fiev, E.V. Shpot, N.V. Petrovsky, D. Zhang, A.V. Proskura, M. Yurova, E.L. Matz, et al., 3D reconstruction of CT scans aid in preoperative planning for sarcomatoid renal cancer: A case report and mini-review, J. X-Ray Sci. Technol. 27 (2) (2019) 389–395.
[3] B. Turkoglu, S.A. Uymaz, E. Kaya, Binary artificial algae algorithm for feature selection, Appl. Soft Comput. 120 (2022) 108630.
[4] B. Turkoglu, S.A. Uymaz, E. Kaya, Clustering analysis through artificial algae algorithm, Int. J. Mach. Learn. Cybern. 13 (4) (2022) 1179–1196.
[5] B. Turkoglu, E. Kaya, Training multi-layer perceptron with artificial algae algorithm, Eng. Sci. Technol., Int. J. 23 (6) (2020) 1342–1350.
[6] Q. Jin, H. Cui, C. Sun, Z. Meng, R. Su, Cascade knowledge diffusion network for skin lesion diagnosis and segmentation, Appl. Soft Comput. 99 (2021) 106881.
[7] H. Cui, Y. Xu, W. Li, L. Wang, H. Duh, Collaborative learning of cross-channel clinical attention for radiotherapy-related esophageal fistula prediction from CT, in: International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer, 2020, pp. 212–220.
[8] Y. Guo, Y. Liu, T. Georgiou, M.S. Lew, A review of semantic segmentation using deep neural networks, Int. J. Multimedia Inf. Retr. 7 (2) (2018) 87–93.
[9] D.M. Pelt, J.A. Sethian, A mixed-scale dense convolutional neural network for image analysis, Proc. Natl. Acad. Sci. 115 (2) (2018) 254–259.
[10] W. Zhao, D. Jiang, J.P. Queralta, T. Westerlund, MSS U-Net: 3D segmentation of kidneys and tumors from CT images with a multi-scale supervised U-Net, Inform. Med. Unlocked 19 (2020) 100357.
[11] Z. Li, J. Pan, H. Wu, Z. Wen, J. Qin, Memory-efficient automatic kidney and tumor segmentation based on non-local context guided 3D U-net, in: International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer, 2020, pp. 197–206.
[12] Z. Yang, Y. Wei, Y. Yang, Collaborative video object segmentation by multi-scale foreground-background integration, IEEE Trans. Pattern Anal. Mach. Intell. (2021) 1–12.
[13] H. Zhao, Y. Zhang, S. Liu, J. Shi, C.C. Loy, D. Lin, J. Jia, Psanet: Point-wise spatial attention network for scene parsing, in: Proceedings of the European Conference on Computer Vision, ECCV, 2018, pp. 267–283.
[14] H. Zhao, J. Shi, X. Qi, X. Wang, J. Jia, Pyramid scene parsing network, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017, pp. 2881–2890.
[15] J. Dai, H. Qi, Y. Xiong, Y. Li, G. Zhang, H. Hu, Y. Wei, Deformable convolutional networks, in: Proceedings of the IEEE International Conference on Computer Vision, 2017, pp. 764–773.
[16] J. Hu, L. Shen, G. Sun, Squeeze-and-excitation networks, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018, pp. 7132–7141.
[17] P. Shaw, J. Uszkoreit, A. Vaswani, Self-attention with relative position representations, 2018, arXiv preprint arXiv:1803.02155.
[18] Y. Chen, Y. Kalantidis, J. Li, S. Yan, J. Feng, A^2-nets: Double attention networks, 2018, arXiv preprint arXiv:1810.11579.
[19] J. Fu, J. Liu, H. Tian, Y. Li, Y. Bao, Z. Fang, H. Lu, Dual attention network for scene segmentation, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019, pp. 3146–3154.
[20] X. Wang, R. Girshick, A. Gupta, K. He, Non-local neural networks, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018, pp. 7794–7803.
[21] H. Fan, P. Chu, L.J. Latecki, H. Ling, Scene parsing via dense recurrent neural networks with attentional selection, in: 2019 IEEE Winter Conference on Applications of Computer Vision, WACV, IEEE, 2019, pp. 1816–1825.
[22] B. Shuai, Z. Zuo, B. Wang, G. Wang, Scene segmentation with dag-recurrent neural networks, IEEE Trans. Pattern Anal. Mach. Intell. 40 (6) (2017) 1480–1493.
[23] Y. Chen, M. Rohrbach, Z. Yan, Y. Shuicheng, J. Feng, Y. Kalantidis, Graph-based global reasoning networks, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019, pp. 433–442.
[24] X. Li, Y. Yang, Q. Zhao, T. Shen, Z. Lin, H. Liu, Spatial pyramid based graph reasoning for semantic segmentation, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020, pp. 8950–8959.
[25] R. Ying, J. You, C. Morris, X. Ren, W.L. Hamilton, J. Leskovec, Hierarchical graph representation learning with differentiable pooling, 2018, arXiv preprint arXiv:1806.08804.
[26] P. Xuan, S. Pan, T. Zhang, Y. Liu, H. Sun, Graph convolutional network and convolutional neural network based method for predicting lncRNA-disease associations, Cells 8 (9) (2019) 1012.

[27] P. Xuan, L. Gao, N. Sheng, T. Zhang, T. Nakaguchi, Graph convolutional autoencoder and fully-connected autoencoder with attention mechanism based method for predicting drug-disease associations, IEEE J. Biomed. Health Inf. 25 (5) (2020) 1793–1804.
[28] P. Xuan, D. Wang, H. Cui, T. Zhang, T. Nakaguchi, Integration of pairwise neighbor topologies and miRNA family and cluster attributes for miRNA–disease association prediction, Brief. Bioinform. (2021).
[29] P. Xuan, K. Hu, H. Cui, T. Zhang, T. Nakaguchi, Learning multi-scale heterogeneous representations and global topology for drug-target interaction prediction, IEEE J. Biomed. Health Inf. (2021).
[30] T.N. Kipf, M. Welling, Semi-supervised classification with graph convolutional networks, 2016, arXiv preprint arXiv:1609.02907.
[31] Z. Wu, S. Pan, F. Chen, G. Long, C. Zhang, S.Y. Philip, A comprehensive survey on graph neural networks, IEEE Trans. Neural Netw. Learn. Syst. 32 (1) (2020) 4–24.
[32] P. Veličković, G. Cucurull, A. Casanova, A. Romero, P. Lio, Y. Bengio, Graph attention networks, 2017, arXiv preprint arXiv:1710.10903.
[33] H. Cui, P. Xuan, Q. Jin, M. Ding, B. Li, B. Zou, Y. Xu, B. Fan, W. Li, J. Yu, et al., Co-graph attention reasoning based imaging and clinical features integration for lymph node metastasis prediction, in: International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer, 2021, pp. 657–666.
[34] G. Bertasius, L. Torresani, S.X. Yu, J. Shi, Convolutional random walk networks for semantic image segmentation, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017, pp. 858–866.
[35] Y. Li, A. Gupta, Beyond grids: Learning graph representations for visual recognition, Adv. Neural Inf. Process. Syst. 31 (2018) 9225–9235.
[36] X. Liang, Z. Hu, H. Zhang, L. Lin, E.P. Xing, Symbolic graph reasoning meets convolutions, Adv. Neural Inf. Process. Syst. 31 (2018) 1853–1863.
[37] H. Cui, X. Wang, J. Zhou, S. Eberl, Y. Yin, D. Feng, M. Fulham, Topology polymorphism graph for lung tumor segmentation in PET-CT images, Phys. Med. Biol. 60 (12) (2015) 4893.
[38] G. Saha, A graph neural network based model for detecting COVID-19 from CT scans and X-rays of chest, Sci. Rep. (11) 1.
[39] R. Selvan, T. Kipf, M. Welling, J.H. Pedersen, J. Petersen, M. de Bruijne, Extraction of airways using graph neural networks, 2018, arXiv preprint arXiv:1804.04436.
[40] A. Garcia-Uceda Juarez, R. Selvan, Z. Saghir, M.d. Bruijne, A joint 3D UNet-graph neural network-based method for airway segmentation from chest CTs, in: International Workshop on Machine Learning in Medical Imaging, Springer, 2019, pp. 583–591.
[41] P. Xuan, H. Cui, H. Zhang, T. Zhang, L. Wang, T. Nakaguchi, H.B. Duh, Dynamic graph convolutional autoencoder with node attribute-wise attention for kidney and tumor segmentation from CT volumes, Knowl.-Based Syst. (2021) 107360.
[42] G. Yang, G. Li, T. Pan, Y. Kong, J. Wu, H. Shu, L. Luo, J.-L. Dillenseger, J.-L. Coatrieux, L. Tang, et al., Automatic segmentation of kidney and renal tumor in CT images based on 3d fully convolutional neural network with pyramid pooling module, in: 2018 24th International Conference on Pattern Recognition, ICPR, IEEE, 2018, pp. 3790–3795.
[43] S. Hu, J. Zhang, Y. Xia, Boundary-aware network for kidney tumor segmentation, in: International Workshop on Machine Learning in Medical Imaging, Springer, 2020, pp. 189–198.
[44] F. Isensee, J. Petersen, A. Klein, D. Zimmerer, P.F. Jaeger, S. Kohl, J. Wasserthal, G. Koehler, T. Norajitra, S. Wirkert, et al., Nnu-net: Self-adapting framework for u-net-based medical image segmentation, 2018, arXiv preprint arXiv:1809.10486.
[45] J. Guo, W. Zeng, S. Yu, J. Xiao, RAU-Net: U-Net model based on residual and attention for kidney and kidney tumor segmentation, in: 2021 IEEE International Conference on Consumer Electronics and Computer Engineering, ICCECE, IEEE, 2021, pp. 353–356.
[46] A. Tureckova, T. Turecek, Z. Kominkova Oplatkova, A.J. Rodriguez-Sanchez, Improving CT image tumor segmentation through deep supervision and attentional gates, Front. Robot. AI 7 (2020) 106.
[47] X. Wang, M. Zhu, D. Bo, P. Cui, C. Shi, J. Pei, Am-gcn: Adaptive multi-channel graph convolutional networks, in: Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, 2020, pp. 1243–1253.
[48] X. Wang, H. Ji, C. Shi, B. Wang, Y. Ye, P. Cui, P.S. Yu, Heterogeneous graph attention network, in: The World Wide Web Conference, 2019, pp. 2022–2032.
[49] N. Heller, N. Sathianathen, A. Kalapara, E. Walczak, K. Moore, H. Kaluzniak, J. Rosenberg, P. Blake, Z. Rengel, M. Oestreich, et al., The kits19 challenge data: 300 kidney tumor cases with clinical context, CT semantic segmentations, and surgical outcomes, 2019, arXiv preprint arXiv:1904.00445.
[50] F. Isensee, P. Jäger, J. Wasserthal, D. Zimmerer, J. Petersen, S. Kohl, J. Schock, A. Klein, T. RoSS, S. Wirkert, et al., Batchgenerators—A Python framework for data augmentation, 2020, http://dx.doi.org/10.5281/Zenodo, Vol. 3632567, Zenodo.
[51] Ö. Çiçek, A. Abdulkadir, S.S. Lienkamp, T. Brox, O. Ronneberger, 3D U-net: Learning dense volumetric segmentation from sparse annotation, in: International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer, 2016, pp. 424–432.
[52] F. Isensee, P.F. Jaeger, S.A. Kohl, J. Petersen, K.H. Maier-Hein, Nnu-Net: A self-configuring method for deep learning-based biomedical image segmentation, Nature Methods 18 (2) (2021) 203–211.
[53] K.H. Zou, S.K. Warfield, A. Bharatha, C.M. Tempany, M.R. Kaus, S.J. Haker, W.M. Wells III, F.A. Jolesz, R. Kikinis, Statistical validation of image segmentation quality based on a spatial overlap index1: Scientific reports, Acad. Radiol. 11 (2) (2004) 178–189.
[54] G. Csurka, D. Larlus, F. Perronnin, F. Meylan, What is a good evaluation measure for semantic segmentation? in: Bmvc, Vol. 27, no. 2013, 2013, pp. 10–5244.
[55] D.P. Huttenlocher, G.A. Klanderman, W.J. Rucklidge, Comparing images using the Hausdorff distance, IEEE Trans. Pattern Anal. Mach. Intell. 15 (9) (1993) 850–863.
[56] J. Chen, Y. Lu, Q. Yu, X. Luo, E. Adeli, Y. Wang, L. Lu, A.L. Yuille, Y. Zhou, Transunet: Transformers make strong encoders for medical image segmentation, 2021, arXiv preprint arXiv:2102.04306.
[57] W. Wang, C. Chen, M. Ding, H. Yu, S. Zha, J. Li, Transbts: Multimodal brain tumor segmentation using transformer, in: International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer, 2021, pp. 109–119.