

## RESEARCH ARTICLE

# Graph Convolutional Networks for Semi-Supervised Image Segmentation

ANNA FABIJANŃKA 

Institute of Applied Computer Science, Lodz University of Technology, 90-537 Łódź, Poland

e-mail: anna.fabijanska@p.lodz.pl

**ABSTRACT** The problem of image segmentation is one of the most significant ones in computer vision. Recently, deep-learning methods have dominated state-of-the-art solutions that automatically or interactively divide an image into subregions. However, the limitation of deep-learning approaches is that they require a substantial amount of training data, which is costly to prepare. An alternative solution is semi-supervised image segmentation. It requires rough denotations to define constraints that are next generalized to precisely delimit relevant image regions without using train examples. Among semi-supervised strategies for image segmentation, the leading are graph-based techniques that define image segmentation as a result of pixel or region affinity graph partitioning. This paper revisits the problem of graph-based image segmentation. It approaches the problem as semi-supervised node classification in the SLIC superpixels region adjacency graph using a graph convolutional network (GCN). The performance of both spectral and spatial graph convolution operators is considered, represented by Chebyshev convolution operator and GraphSAGE respectively. The results of the proposed method applied to binary and multi-label segmentation are presented, numerically assessed, and analyzed. In its best variant, the proposed method scored the average DICE of 0.86 in the binary segmentation task and 0.79 in the multi-label segmentation task. Comparison with state-of-the-art graph-based methods, including Random Walker and GrabCut, shows that graph convolutional networks can represent an attractive alternative to the existing solutions to graph-based semi-supervised image segmentation.


**INDEX TERMS** GCN, graph convolutional networks, graph node clustering, region adjacency graph, semi-supervised image segmentation.

## I. INTRODUCTION

The problem of image segmentation has been continuously gaining the attention of the computer vision research community. The precise location and delineation of meaningful objects are crucial in many image processing and analysis pipelines. Therefore, various approaches have been proposed to partition an image into regions corresponding to diverse objects and precisely outline their borders.

Recently, state-of-the-art approaches to image segmentation have been dominated by convolutional neural networks (CNN) [1], [2], [3]. The vast majority of these automatic approaches focus either on semantic segmentation [4], [5],

[6], [7] or identifying object instances of particular characteristics (e.g., specific organs, tumorous regions, or disease symptoms in the case of medical images [8], [9], [10], [11]). However, convolutional neural network architectures were also leveraged to learn the interactive segmentation task [12], [13], [14], [15], [16], [17]. The CNN-based approaches to interactive segmentation additionally need image-user interaction pairs at the input in the form of bounding-boxes [15], user positive and negative clicks or scribbles [12], [13], [14], or object's four most extreme points [16]. Although CNN-based methods have demonstrated exceptional performance in various image segmentation problems, they usually require a substantial amount of precisely annotated train data which is costly to prepare and, therefore, frequently limited. Despite current efforts to save for the annotation workload,

The associate editor coordinating the review of this manuscript and approving it for publication was Mingbo Zhao .

semi-supervised CNN-based segmentation approaches still require some precisely labeled data to learn from unlabeled data [18]. At the same time, most self-supervised methods depend on fully-supervised pre-trained models [19], while weakly-supervised methods may still require roughly annotating region location (e.g., by drawing bounding boxes, scribbles, or points on many images [20]). Furthermore, convolutional neural networks cannot generalize to objects unrepresented in the train set.

Convolutional neural networks' exceptional performance and potential benefits have drawn attention away from semi-supervised image segmentation that uses user-provided labels to form a training set. These approaches need coarse indicating the position of relevant objects and background [12], [13], [14], [15], [16], [17], [21], [22], [23], [24]. The seed regions can be shown either interactively or in some other way, e.g., using a priori knowledge about the regions to be segmented. Initial constraints derived from rough denotations are generalized to provide label-likelihood for unlabelled pixels and precisely delimit relevant image regions. The semi-supervised approach allows adapting the segmentation process to a particular image and various objects. As a result, they can often provide a much more accurate object outline than automatic algorithms. Besides, semi-supervised segmentation does not require training examples, as CNN-based methods do. This property can be a significant advantage in many image processing pipelines.

The most prominent group of semi-supervised segmentation methods are graph-based algorithms. They define image segmentation as a result of pixel or region affinity graph partitioning. Several approaches to graph partitioning have been proposed over the years, including graph cuts [25], [26], [27], and random walker [28]. Even though they remain the most popular semi-supervised image segmentation methods, graph-based approaches exhibit limitations. The most significant one represents the considerable computational cost, which limits the application of graph-based algorithms in the case of high image resolutions.

This paper revisits the problem of graph-based image segmentation. It approaches the problem as semi-supervised node classification in the affinity graph using a graph convolutional network (GCN). As recent works regarding graph data show, GCNs exhibit the ability to discover structural patterns and learn graph data feature representation, which can be next used for node labeling [29], [30], [31]. The labeling results remain accurate, even for a limited number of labeled nodes.

Applications of the graph convolutional networks in image segmentation are yet limited. To the best of the author's knowledge, this is one of the first works exploring the performance of GCNs in the problem of semi-supervised image segmentation. A few related works regard semantic segmentation [14], [32], [33], [34], [35], which remains a diverse problem since segmentation is performed fully automatically in a fully supervised way by models trained from end-to-end with multiple train examples.

This paper brings the following main contributions:

- graph convolutional neural network is applied for semi-supervised image segmentation for the first time,
- performance of both spectral and spatial graph convolution operators is analyzed in the problem of binary and multi-label image segmentation, and their performance is compared,
- the advantages and limitations of graph convolutional networks in the problem of semi-supervised image segmentation are discussed.

The following part of this paper is organized as follows. Section II briefly revisits graph-based methods of image segmentation and reviews the basics of graph convolutional networks. Then follows the formulation of a semi-supervised image segmentation problem in Section III. Section IV details the proposed approach and the resulting graph convolutional network architecture. The results of the semi-supervised image segmentation with the use of GCN are demonstrated in Section V and discussed in Section VI. Finally, Section VII concludes the paper.

## II. RELATED WORKS

### A. GRAPH BASED IMAGE SEGMENTATION

Graph-based methods represent images as weighted graphs where nodes represent pixels, edges represent relations between pixels, and weights quantify these relations employing a similarity measure. Image segmentation is then defined as a graph partitioning into sub-graphs representing relevant image regions.

There exist diverse approaches for finding edges to be removed in graph partitioning. They can be divided into three main groups: minimal spanning tree (MST) methods, graph cuts, and the Random Walker approach.

MST-based methods for image segmentation perform graph division concerning the minimal spanning tree of pixel adjacency graph [36], [37]. These methods can divide an image into any number of classes. However, they usually result in significant over-segmentation and are primarily used to generate superpixels (i.e., homogeneous image regions).

Graph-cut methods partition affinity graph concerning a minimal cut (i.e., a set of edges with minimal total weights). The cut is found either via spectral graph partitioning [25], [38] or combinatorial graph cuts. In the first case, the generalized eigenvalue problem is solved to find eigenvectors of the graph Laplacian and utilize them to perform optimal graph bipartition. Finding the eigenvectors is, however, a highly complex problem. Therefore spectral graph partitioning methods remain mainly theoretical, and their application is limited to low-resolution images. Combinatorial graph cuts define subgraphs by solving the min-cut/max-flow problem [26], [27], [39]. Explicitly, the graph bipartition is determined by computing the global optimum among all possible bipartitions that satisfy constraints imposed on the object and background. The optimum is determined by minimizing the energy that incorporates regional and boundary conditions.

Extensions to the original min-cut/max-flow approach like GrabCut [40], and Lazy Snapping [41] up to today remain the most effective and prevailing approaches to semi-supervised image segmentation [42], [43], [44].

Random Walker [28] presents an alternative approach to graph-based segmentation. It assigns labels to nodes based on the probability that a random walker released from unlabeled nodes first arrives in nodes with a particular label assigned. The unlabeled nodes are assigned a label for which the highest probability exists. Contrary to graph-cuts, the Random Walker can partition an image into multiple classes.

When applied pixel-wise, the graph-based approaches to image segmentation experience significant problems with efficiency. Explicitly, for high image resolutions, an enormous amount of computer memory is required to represent the corresponding massive image graph. This overhead makes graph partitioning highly complex in computations, providing segmentation results in an unacceptable time. To alleviate these limitations, graph-based approaches are routinely used region-wise where nodes represent so-called superpixels, i.e., groups of connected pixels sharing some properties [45], [46]. Processing region adjacency graph reduces time and memory overhead. However, it may decrease segmentation accuracy near image boundaries.

## B. GRAPH CONVOLUTIONAL NETWORKS

Graph convolutional networks (GCN) have been proposed in the last few years to generalize convolutional neural networks into irregular or non-Euclidean domains. Particularly, GCNs attempt to deal with general (irregular) graphs where computing convolution is not as straightforward as for the grid graphs modeling pixel relations. Such irregular graphs may encode complex geometric data, structure, and pairwise relationships in numerous problems [47], [48], examples of which are social networks [49], [50], citation networks [51], [52], [53], gene data and protein structure [54], chemical molecular structures [55] or transportation networks [56], [57].

Extending convolutions, which are easy to compute for regular grids, into non-regular domains remains not a straightforward problem. It is mainly due to the necessity of preserving the weight-sharing property in the case of different size node neighborhoods. Therefore, many attempts have recently been made to extend convolution operation to irregular graphs. Most of them utilize neighborhood aggregation schemes or message passing [55] that generalize graph embedding for the node classification task.

There exist several concepts that generalize convolution operation into irregular graph domains. They can be classified into spectral methods and spatial methods [58].

Spectral methods define a convolution through the graph Fourier transform from graph signal filtering after converting to spectral-domain representation. This approach involves eigenvectors derived from the spectral decomposition of the graph Laplacian used as a Fourier basis [29], [59], [60]. To avoid direct computation of the graph

Laplacian eigenvectors, [61] proposed ChebNet where  $k$ -th order Chebyshev polynomials are used to approximate spectral filters that learn on  $k$ -hop neighborhoods of the graph. In [30] this approach was further simplified by using first-order polynomials fused with original feature information of data. The significant drawback of spectral convolution methods is that they are restricted to a fixed graph structure, including the number of nodes and their degrees.

The spatial graph convolution performs computations directly on the graph dependent on nodes' spatial relations [59]. Explicitly, the representation of a central node is updated concerning the representation of its neighbor. The node information is next propagated along graph edges. In this approach, two challenges need to be faced. These are receptive field selection and node ordering. Main approaches to this problem include: assuming a predefined contribution of each node [62], employing a graph diffusion process to incorporate the node context information [63], converting the graph locally to a linear vector space [64], adopting attention mechanism [65] or sampling node neighborhood to obtain a fixed number of neighbors [66]. Compared to spectral approaches, spatial graph convolution methods are more flexible, allowing variable graph structures. They are also more efficient in computing than their spectral counterparts.

This paper's approach to semi-supervised image segmentation combines graph-based image segmentation with graph convolutional networks. However, instead of graph-partitioning, it approaches the image segmentation problem as node clustering in the SLIC region adjacency graph (RAG). A graph convolutional network performs the classification, which uses image region relations expressed by RAG to perform image segmentation. The performance of spectral and spatial convolutions applied in image segmentation is compared and analyzed.

## III. PROBLEM FORMULATION

Let  $\mathcal{G} = (\mathcal{V}, \mathcal{E})$  be a weighted and undirected graph where  $\mathcal{V}$  is a vertex (node) set of cardinality  $N = |\mathcal{V}|$  and  $\mathcal{E} \subseteq \mathcal{V} \times \mathcal{V}$  is a weighted edge set of cardinality  $M = |\mathcal{E}|$ . Nodes  $v_i \in \mathcal{V}$  represent either single pixels or groups of connected pixels sharing similar properties (i.e., superpixels). Edges  $e_{ij} = (v_i, v_j) \in \mathcal{E}$  where  $i \neq j$  connect neighboring nodes  $v_i$  and  $v_j$ . Depending on representation, graph  $\mathcal{G}$  is thus either a regular pixel adjacency graph or irregular region adjacency graph. Each node  $v_i$  is a  $C$ -dimensional feature vector  $\mathbf{f}_i$ , of color (or intensity) features. Features of all nodes are thus represented as  $N \times C$  dimensional matrix  $X$ . Additionally, each edge  $e_{ij}$  has its corresponding weight describing similarity of the adjacent nodes  $v_i$  and  $v_j$ . All edges in a graph are represented as an adjacency matrix  $\mathcal{A}$  of size  $N \times N$  where entry  $a_{ij}$  indicates if nodes  $i$  and  $j$  are connected.

Let  $\mathcal{L}$  and  $\mathcal{U}$  denote subsets of labeled and unlabeled nodes such that  $\mathcal{L} \cup \mathcal{U} = \mathcal{V}$ ,  $\mathcal{L} \cap \mathcal{U} = \emptyset$ , and  $|\mathcal{L}| \ll |\mathcal{U}|$ . Each labeled node  $v_i \in \mathcal{L}$  has its corresponding label  $l_i \in \{1, 2, \dots, N_C\}$  where  $N_C$  is the total number of classes. With respect to the

above, the problem of image segmentation can be formulated as a semi-supervised learning of a classifier  $F : v \rightarrow y$  to assign a class label  $l_i$  to each unlabeled node  $v_i \in \mathcal{U}$  based on both labeled and unlabeled nodes (transductive learning settings).

#### IV. PROPOSED METHOD

##### A. GENERAL IDEA

The key idea behind the proposed approach is to use a graph convolutional neural network (GCN) to perform image segmentation concerning some constraints imposed on the regions of interest. Mainly, a GCN is applied to the image region adjacency graph (RAG) to predict region labels given sample nodes representing resulting regions. A GCN is thus trained in a semi-supervised learning framework to perform node-level classification.

Region constraints may be input in many ways. In this work, they are given as scribbles indicating the rough location (seed points) of regions of interest and thus representing their properties. Graph nodes covered by scribbles are considered labeled nodes  $\mathcal{L}$  with region labels  $l_i$  assigned regarding the unique scribble colors. Nodes not covered by scribbles remain unlabeled. A GCN predicts their region labels. The general idea of the proposed approach is presented in Fig. 1. The proposed approach can segment several regions of interest at a time, providing that the seeds for each of them are given (multi-label segmentation).

##### B. REGION ADJACENCY GRAPH

To build region adjacency graph  $\mathcal{G}$ , the input image  $I$  was divided into multiple uniform disjoint regions  $\{R_i\}_{i=1}^N$  using the SLIC (Simple Linear Iterative Clustering) superpixel approach (see Fig. 2). The number of resulting regions was arbitrarily set to  $N = \lceil \frac{PQ}{100} \rceil$ , where  $P \times Q$  is image spatial resolution. Prior division image features were normalized to the range of  $[0, 1]$  simply by dividing all channels by  $2^B - 1$  where  $B$  stands for a bit depth of each channel. The image was next smoothed with a Gaussian kernel of size  $\sigma = 2$ . In the resulting region adjacency graph, the superpixels  $R_i$  were represented by vertices  $v_i$ , each described by a  $N_C$ -dimensional feature vector  $\mathbf{f}_i$  derived from the corresponding region. Mean color (or intensity) within the region  $R_i$  was considered by averaging each color channel separately.

Weighted edges connected nodes representing adjacent regions. Weights are derived from a Gaussian weighting function given by the following formula:

$$w_{ij} = e^{-\beta d(\mathbf{f}_i - \mathbf{f}_j)} \quad (1)$$

where:

$$d(\mathbf{f}_i - \mathbf{f}_j) = \sum_{k=1}^C (f_i^k - f_j^k)^2 \quad (2)$$

and  $\beta$  is a free parameter of the method.

An image RAG is represented by a region adjacency matrix  $A = [a_{ij}] \in \mathbb{R}^{N \times N}$  such that:

$$a_{ij} = \begin{cases} 1 & \text{if regions } R_i \text{ and } R_j \text{ are adjacent} \\ 0 & \text{otherwise} \end{cases} \quad (3)$$

and a region feature matrix  $X \in \mathbb{R}^{N \times C}$  representing the  $N$  vertices each having features in  $\mathbb{R}^{1 \times C}$ .

##### C. GRAPH CONVOLUTION NETWORK MODEL

###### a: NETWORK GENERAL ARCHITECTURE

The architecture of the graph convolutional neural network used in this study is detailed in Table 1. The resulting GCN architecture stacks two convolutional layers such that:

$$\mathcal{H}^{(p)} = \sigma(\mathcal{H}^{(p-1)} *_{\mathcal{G}} g) \quad (4)$$

where:

- $\mathcal{H}^{(p)} \in \mathbb{R}^{N \times n_p}$  denotes feature matrix of  $p$ -th layer,  $n_p$  is the number of feature maps and  $\mathcal{H}^0 = X$ ;
- $\sigma(\cdot)$  stands for an activation function;
- $*_{\mathcal{G}}$  denotes graph convolution operator;
- $g$  is some function which aggregates node neighborhoods.

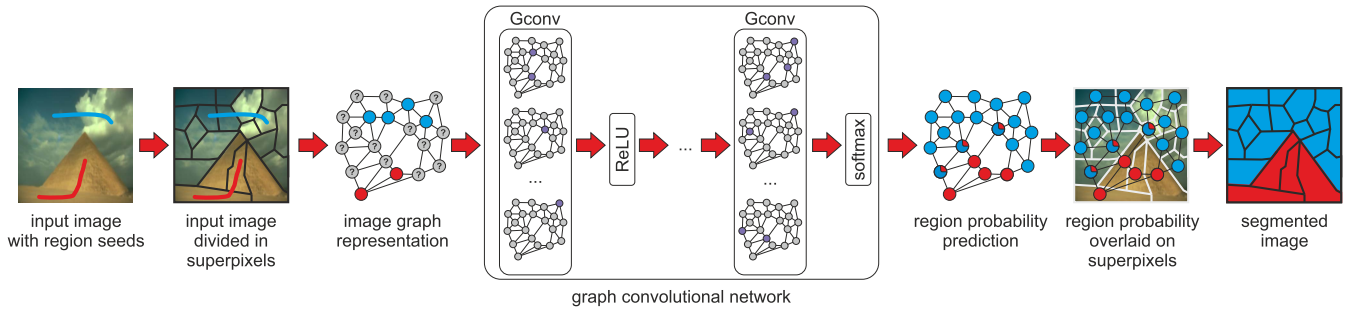
In the considered architecture graph convolutional layers are separated by a dropout layer used to reduce overfitting. The number of input channels to the first layer  $\mathcal{H}^{(0)} \in \mathbb{R}^{N \times C}$  is  $C$  (which in the case of the RGB images as used in this study equals 3) and the number of output channels is 16. The ReLU activation, such that  $f(\mathcal{H}) = \max(0, \mathcal{H})$ , is applied to the output of the first layer. The second convolutional layer  $\mathcal{H}^{(1)} \in \mathbb{R}^{N \times 16}$  takes 16 channels at the input and outputs  $N_C$  channels, where  $N_C$  is the number of labels in the segmented image. Softmax activation is applied to the output of the second convolutional layer to obtain the probability that the region  $R_i$  represented by node  $v_i$  belongs to each of the  $N_C$  regions. Finally, node  $v_i$  is the assigned label of a region for which the greatest probability exists.

**TABLE 1. Architecture of the GCN used in this study;  $C$  - the number input feature channels ( $C = 3$  for RGB images,  $C = 1$  for grayscale images);  $N_C$  - the number of labels in the segmented image.**

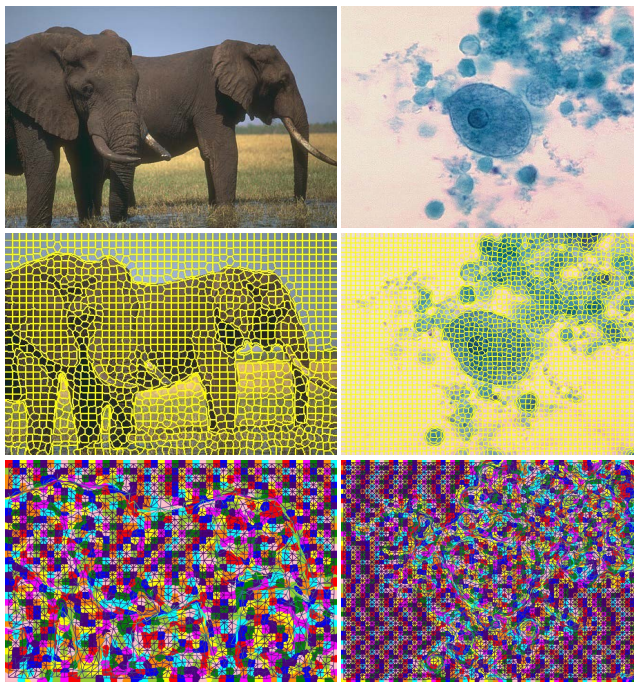
Layer	Input size	Output size
Conv1	$C$	16
ReLU	16	16
Dropout	16	16
Conv2	16	$N_C$
Softmax	$N_C$	$N_C$

The GCN architecture outlined in Table 1 was tested in two variants, the spectral and the spatial one. They differ in the approach used for calculating convolution on a graph in the convolutional layers of the neural network. The most popular graph convolution operators representing the spatial and the spectral approach were considered and comparatively tested. Particularly, the GraphSAGE and Chebyshev convolutional operator were considered.





**FIGURE 1.** Method is initialized with user-provided seeds (scribbles) corresponding to regions of interest. The input image is divided into superpixels using the SLIC approach. A superpixels adjacency graph is next generated to represent an image. Graph nodes correspond to superpixels. Edges connect neighboring superpixels. Graph nodes corresponding to superpixels that overlap with user-provided seeds are assigned labels of the corresponding regions. Labels of the remaining graph nodes are determined by the graph convolutional network. The final segmentation result is obtained by assigning predicted labels to the corresponding superpixels.



**FIGURE 2.** Image division into superpixels; sample input image (top panel) image divided into SLIC superpixels (middle panel); the corresponding region adjacency graph overlaid on superpixels (bottom panel).

*b: ChebNet-THE SPECTRAL GCN VARIANT*

The spectral variant of the considered GCN architecture (referred later herein as ChebNet) uses the Chebyshev convolutional layers (ChebConv) as proposed in [61]. The information between layers was propagated following the equation:

$$\mathcal{H}_{j=\{1,2,\dots,n_p\}}^{(p)} = \sigma\left(\sum_{i=1}^{n_{p-1}} \left(\sum_{k=0}^K \theta_k T_k(\tilde{L}) \mathcal{H}_i^{(p-1)}\right)\right) \quad (5)$$

where:

- $T_k(\cdot)$  denotes recursive series of Chebyshev polynomials, where  $T_0(X) = 1$ ,  $T_1(X) = X$ , and  $T_i(X) = 2XT_{i-1}(X) - T_{i-2}(X)$ ;

- $\theta$  is learnable filter coefficient matrix (namely, the vectors of Chebyshev coefficients  $\theta_k$ );
- $K$  stands for the order of network neighborhood on each convolutional layer; for of the first order neighborhood ( $K = 1$ ) only the nodes' immediate neighbors are considered when computing convolution, higher orders indicate that nodes  $K$ -hoops away are considered.
- $\tilde{L} = \frac{2L}{\lambda_{max}} - \mathcal{I}_n$ , where  $L$  is a normalized graph Laplacian, such that  $L = \mathcal{I}_n - \mathcal{D}^{-\frac{1}{2}} \mathcal{A} \mathcal{D}^{-\frac{1}{2}}$ ,  $\lambda_{max}$  is the largest eigenvalue of graph Laplacian, and  $\mathcal{D}$  is graph degree matrix with  $\mathcal{D}_i = \sum_j a_{ij}$ , and  $\mathcal{A} = [a_{ij}]$  denoting graph adjacency matrix.

*c: SageNet-THE SPATIAL GCN VARIANT*

The spatial variant of the considered GCN architecture (referred later herein as SageNet) uses the GraphSAGE operator as proposed in [66]. The information between layers was propagated following the equation:

$$\mathcal{H}_v^{(p)} = \frac{\tilde{\mathcal{H}}_v^{(p)}}{\|\tilde{\mathcal{H}}_v^{(p)}\|_2} \quad \forall v \in \mathcal{V} \quad (6)$$

where:

$$\tilde{\mathcal{H}}_v^{(p)} = \sigma(\mathcal{W} \cdot \text{MEAN}(\{\mathcal{H}_v^{(p-1)}\} \cup \{\mathcal{H}_u^{(p-1)} \forall u \in \mathcal{N}(v)\})) \quad (7)$$

and  $\mathcal{W}$  is a weight matrix,  $\mathcal{N}(\cdot)$  is a node neighborhood function, and  $\mathcal{H}_v^{(p)}$  is a feature vector of node  $v$  in  $p$ -th layer.

**D. EXPERIMENTAL SETUP**

The method was implemented in Python 3.7 programming language. Mainly, the GCN was implemented using PyTorch Geometric, i.e., a geometric deep learning extension library for PyTorch. PyTorch ran at the top of TensorFlow. Experiments were performed on a desktop computer with Intel Core i9-7940X (3.10 GHz) processor, 128 GB RAM, and NVidia GeForce GTX Titan X GPU.

The GCN was trained for 5000 epochs with the patience of 500 epochs (i.e., the training was stopped after 1000 epochs with no improvement). The Adam optimizer with the learning

rate of 0.005 and weight decay of 0.0005 was applied to minimize the negative log-likelihood loss. The weights ensuring the lowest loss were used for graph node classification.

## V. EXPERIMENTAL RESULTS

### A. EVALUATION PROCEDURE

The assessment of the proposed approach was performed concerning two available publicity datasets, namely the Berkeley Segmentation Dataset [67] and the GrabCut Vision Dataset [40]. In total, 100 images and the corresponding ground-truth segmentation results, were considered. The performance of the GCN-based approach was assessed in both binary (single object vs. background) and multi-label segmentation task. In the first case, the GrabCut dataset, containing 50 images in total, was used. The evaluation of multi-region segmentation accuracy was performed concerning 50 images randomly selected from the Berkeley dataset. The test images represented challenging scenes in both experiments, including complex content, multiple objects, and similar foreground and background. The rough location of regions of interest was given as scribbles with different scribble colors representing different labels.

The DICE coefficient and Jaccard index (also known as Intersection over Union, IoU) were used for the numerical evaluation of the segmentation results. The measures are defined by (8) and (9) respectively.

$$DSC = \frac{2|S \cap T|}{|S| + |T|} \quad (8)$$

$$IoU = \frac{|S \cap T|}{|S \cup T|} \quad (9)$$

where  $S$  and  $T$  stand for segmentation result and ground truth, respectively.

In the case of multi-label image segmentation, individual scores were determined for each label and then averaged.

The proposed approach to semi-supervised image segmentation was also compared to the most popular state-of-the-art graph-based competitors. For binary segmentation, these included Random Walker and GrabCut. In the case of multi-label segmentation, only the Random Walker was considered since the GrabCut is intended for binary segmentation (object extraction from background). All considered methods were initialized with the same scribbles to facilitate a fair comparison.

A qualitative evaluation of the considered method was also performed for medical and microscopy images randomly collected over available publicity sources. For these images, the ground truths were not given. However visual assessment of the segmentation allowed to observe some properties of the proposed approach.

### B. BINARY SEGMENTATION

The visual results of object extraction from the background using the proposed approach are presented in Fig. 3. Sample images from the Microsoft dataset were selected for presentation. The top panel presents original images with user

input (scribbles) overlaid on them. Consecutive rows present segmentation results obtained by the considered methods with method names indicated on the left side of each row. ChebNet corresponds to a graph convolutional network with two Chebyshev spectral convolutional layers, while SageNet corresponds to a network of the same architecture but with SAGE spatial convolutional layers.

The distributions of the considered image segmentation accuracy measures obtained for each of the considered methods over the Microsoft dataset are presented in Fig. 4.

Finally, the visual results of the GCN-based approach applied to random histopathological images<sup>1</sup> are presented and compared to the results of the competitors in Fig. 5.

### C. MULTI-LABEL SEGMENTATION

The visual results of multi-label segmentation are presented in Fig. 6. Sample images were selected from the Berkeley dataset. As in the case of binary segmentation, the top panel presents original images with user input (scribbles) overlaid on them. Consecutive rows present segmentation results obtained by the considered methods with method names indicated on the left side of each row.

The distributions of the considered image segmentation accuracy measures obtained for each of the considered methods over the Berkeley dataset are presented in Fig. 7. The measures were determined for each resulting region separately and then averaged over the number of regions.

Finally, the visual results of the GCN-based multi-label segmentation applied to random microscopy, and medical images are presented and compared to the results of the competitors in Fig. 8.

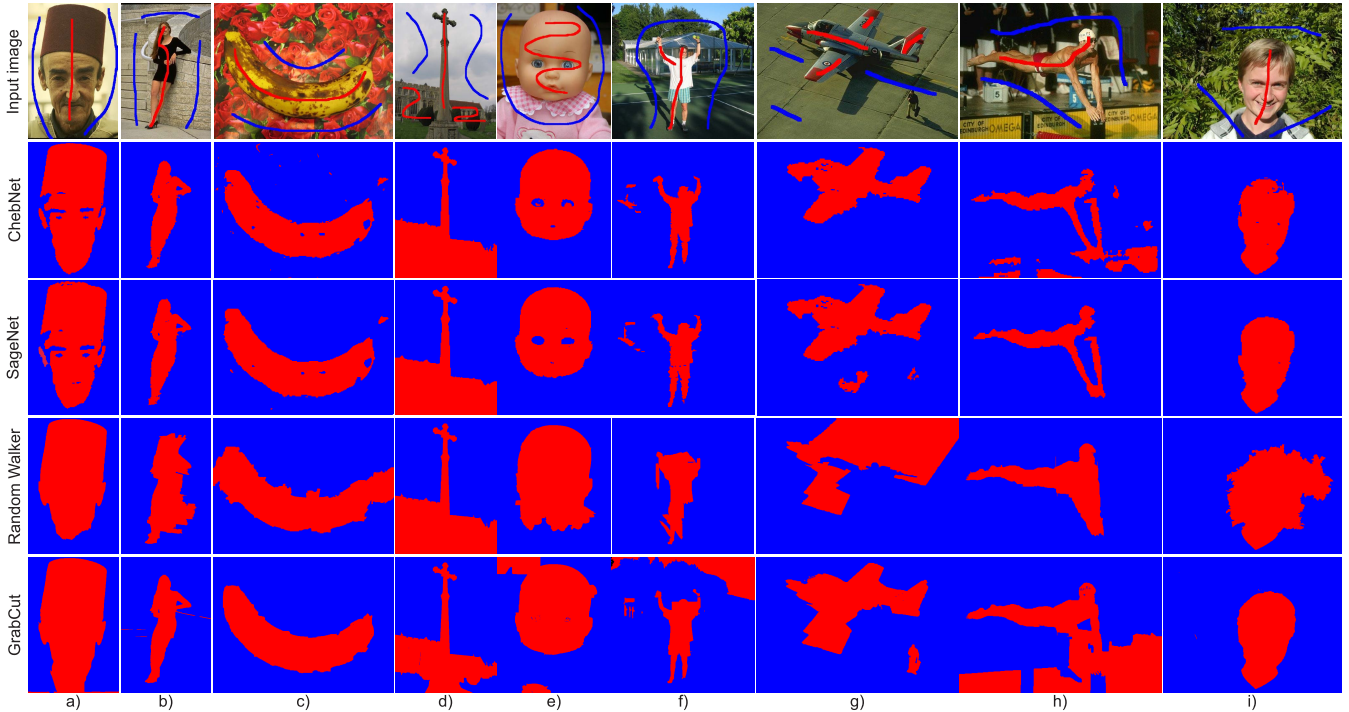
### D. SENSITIVITY TO INITIALIZATION

The influence of initial seeds on image segmentation results is presented in Fig. 9. Selected examples present cases with seeds changing gradually from precise and dense markings spreading across whole regions (cf. Fig. 9a) to very scarce annotations (cf. Fig. 9d). For both considered samples, the top panel presents seeds overlaid on the original image, while the panels below present the corresponding results with colors of the regions complying with the colors of initial annotations. Again, ChebNet and SageNet are considered with the model's name on each row's left side.

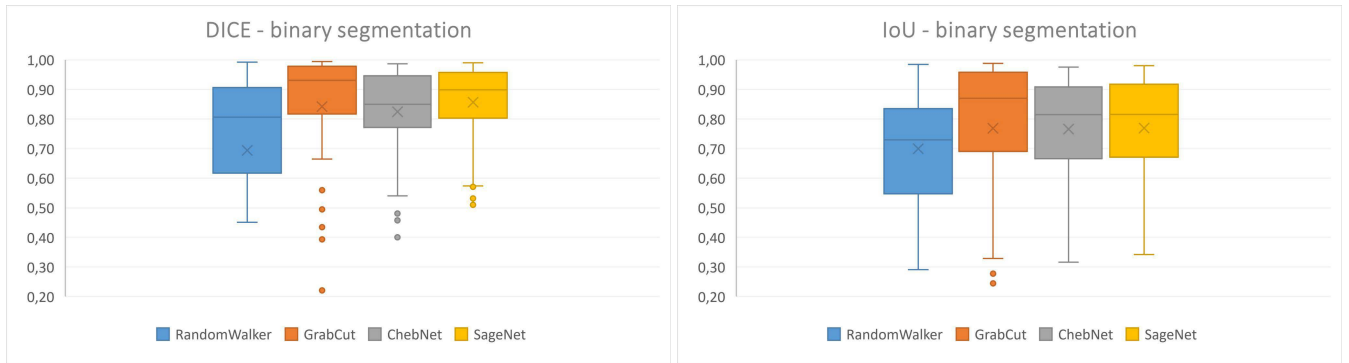
## VI. DISCUSSION

The results presented in Section V confirm that graph convolutional networks can be successfully applied for semi-supervised image segmentation in both binary (see Fig. 3, 5) and multi-label (see Fig. 6, 8) segmentation tasks. Based on the visual assessment, it can be observed that the GCN-based approach is universal. Particularly, it performs reasonably well for different images, representing diverse scenes and exhibiting different properties, including natural scene images, microscopic images, and medical images of

<sup>1</sup>credit to <http://www.publicdomainfiles.com/>



**FIGURE 3.** Results of semi-supervised binary segmentation of natural scene images using the graph convolutional neural network compared to the results of state-of-the-art methods. The method names are indicated on the left side of each row.



**FIGURE 4.** Distribution of image segmentation accuracy measures scored by the considered methods in the binary image segmentation task.

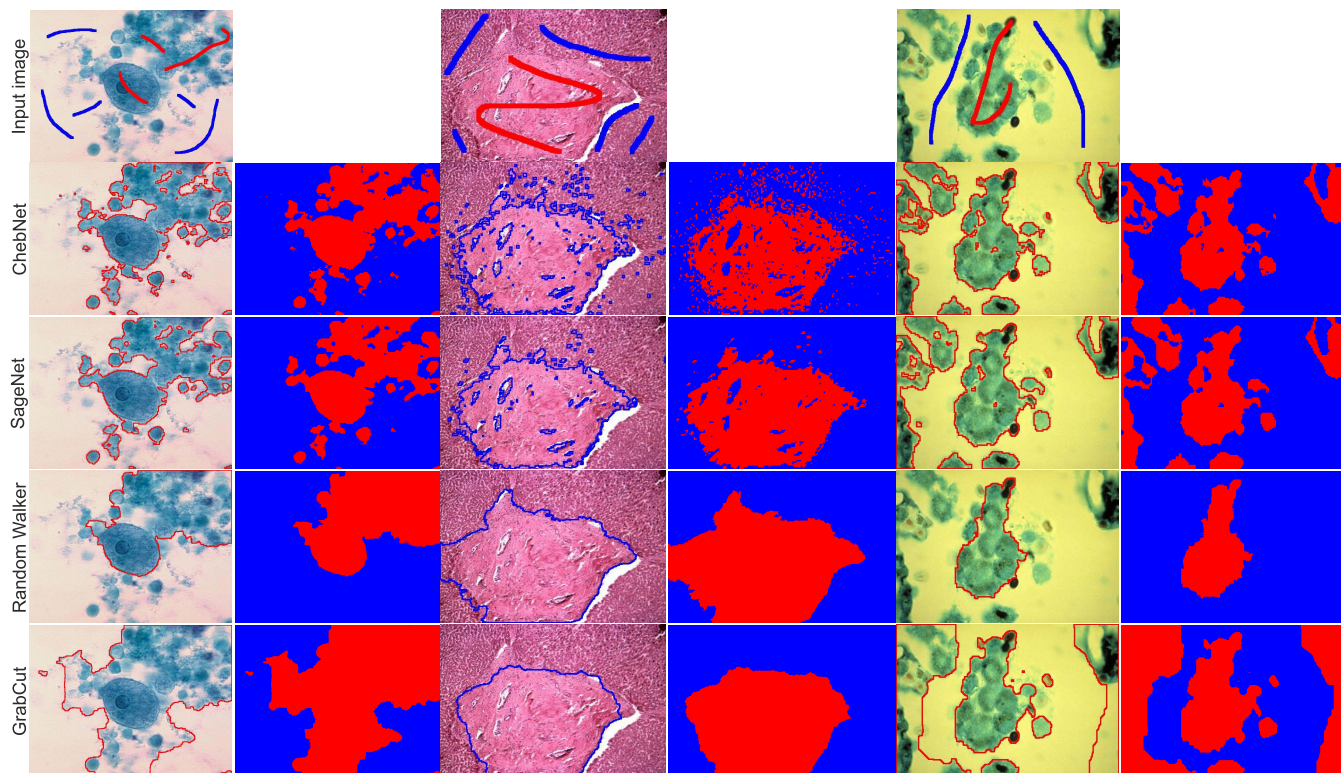
various modalities. The results of image partitioning by the GCN mostly precisely match the regions’ boundaries, even if the background, object, or both are nonuniform and complex. This observation regards both variants of the considered GCN architecture, i.e., the spectral one represented by the ChebNet and the spatial one represented by the SageNet.

Spectral and spatial convolution operators perform similarly well in the case of image segmentation. However, the SageNet (i.e., spatial approach) is less sensitive to local intensity and color variations. As a result, it provides slightly better segmentation results with more regular edges and less noise. Image segmentation quality metrics also confirm this visual observation. Notably, in the case of binary segmentation, the average and median values of the DICE coefficient scored by

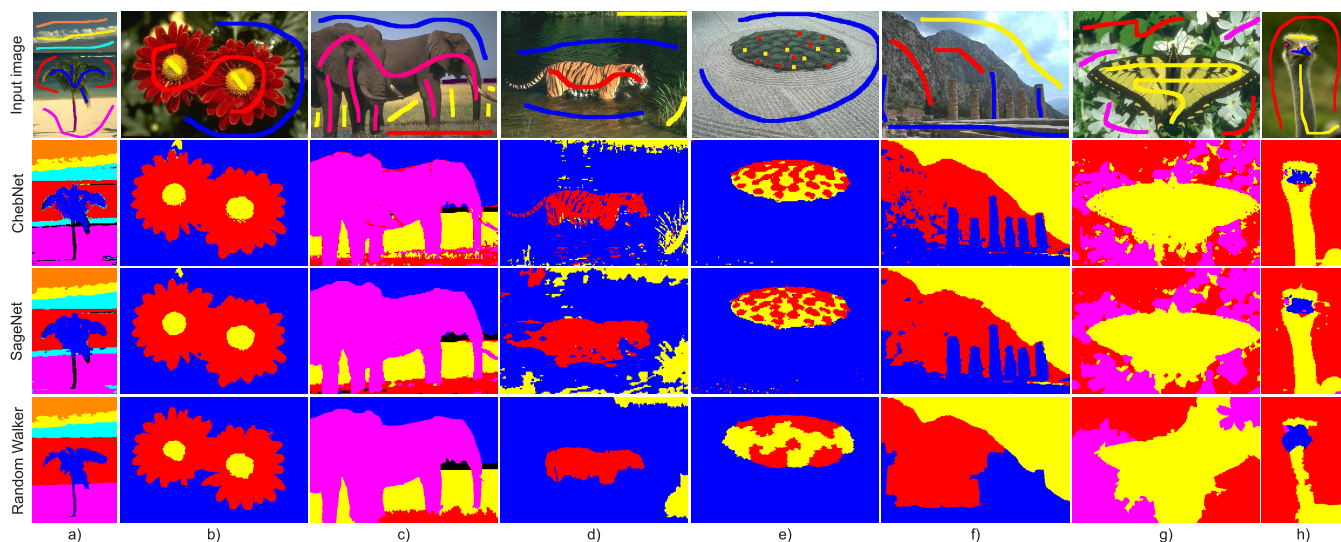
the SageNet were at the level of 0.86 and 0.90 vs. 0.82 and 0.85 scored respectively by the ChebNet. The corresponding values of the Jaccard index were almost equal for both variants and equal on average to 0.77, with the median value equal to 0.82. However, in the case of multi-label segmentation, both variants of the GCN performed equally well, scoring on average the DICE coefficient at the level of 0.79 and a median value equal to 0.83.

When compared to the competitive approaches, the GCN-based methods exhibit certain advantages. The advantage over the Random Walker is evident and confirmed by visual and numerical results. The Random Walker compromises details of region shapes over the compactness of the resulting regions. As a result, object shape details are





**FIGURE 5.** Results of the proposed approach applied for binary segmentation of random microscopic and medical images compared to the results of the competitors. The method names are indicated on the left side of each row.



**FIGURE 6.** Results of semi-supervised multi-label segmentation of natural scene images using the graph convolutional neural network compared to the results of state-of-the-art methods. The method names are indicated on the left side of each row.

often lost while object boundaries are smoothed. At the same time, the GCN-based approach retains even the finer details of object shapes. This effect can be observed, for example, in Fig. 3a where eyes, ears, and eyebrows are visible in the segmentation results, in Fig. 3h where swimmer’s arms and legs were precisely retained in the resulting image or Fig. 5ab where the shape of the cells was

precisely represented. The proposed approach is also resistant to the nonuniform background color or intensity distribution. In such a case, it precisely extracts objects from the background while the Random Walker experiences serious problems (see Fig. 3bcei). In the case of the multi-label image segmentation task, the Random Walker approach is not able to segment small regions and merges them. In contrast, the



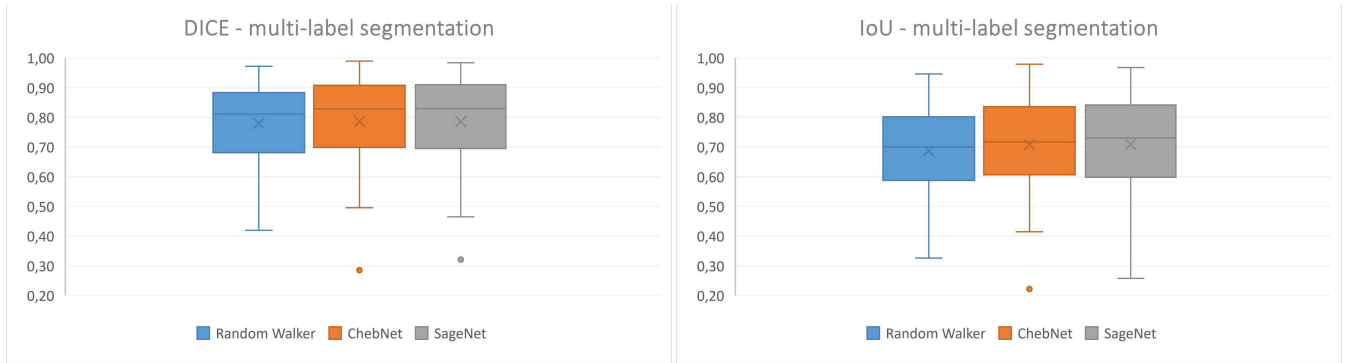


FIGURE 7. Distribution of image segmentation accuracy measures scored by the considered methods in the multi-label image segmentation task.

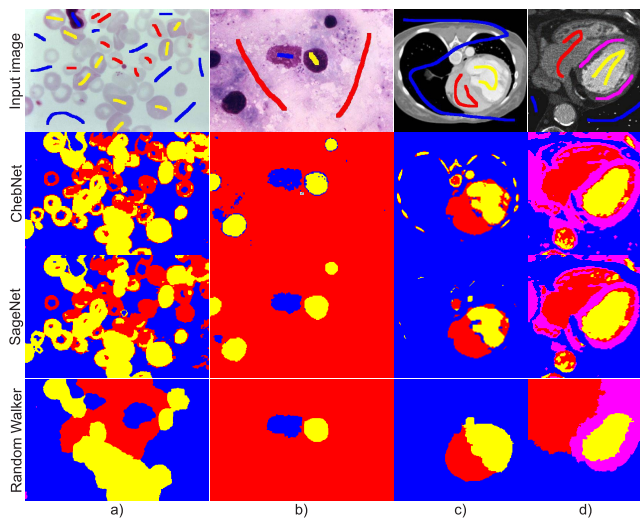


FIGURE 8. Results of semi-supervised multi-label segmentation of random medical and microscopy images using the graph convolutional neural network compared to the results of the state-of-the-art method. The method names are indicated on the left side of each row.

GCN approach retains even the smallest regions. This effect can be seen in the natural scene images in Fig. 3ef, as well as microscopy image in Fig. 8a. Overall, the GCN-based approach scores higher measures than the Random Walker in both binary (see Fig. 4) and multi-label segmentation tasks (see Fig. 7).

When the comparison with the GrabCut is considered, the main advantage of the proposed GCN-based approach is the ability to perform multi-label segmentation. The GrabCut is intended to extract an object from the background. As a result, its application is limited to binary segmentation tasks. When compared numerically, the GrabCut slightly outperforms the proposed approach in the natural scene segmentation task (see Fig. 4). This effect is caused mainly by the exclusion from the results of the proposed approach, some subregions like eyes in Fig. 3aei. Depending on the application, this can be both a disadvantage and an advantage of the method.

In the case of microscopy images presented in Fig. 5ab, the GrabCut performed significantly worse than in the case of

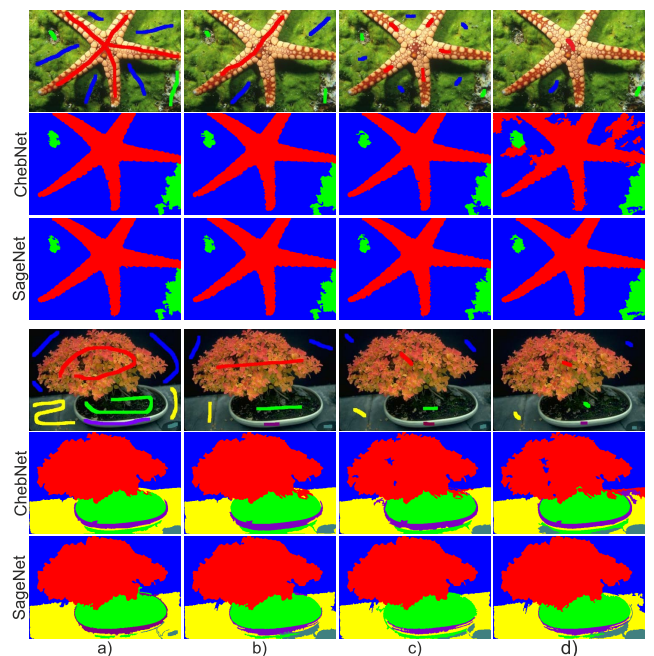
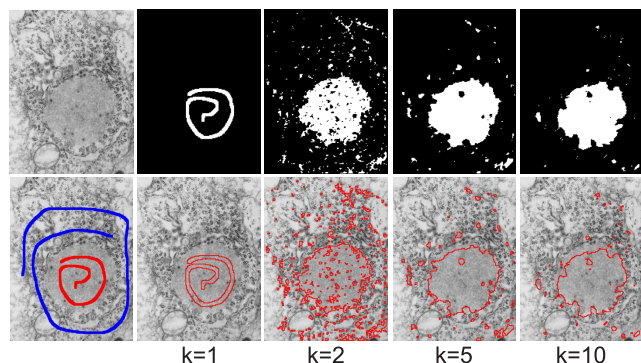


FIGURE 9. Influence of method initialization on the image segmentation result; a) precise and dense seeds; b) moderately accurate seeds; c) scarce seeds; d) very scarce seeds. Initial seeds (scribbles) are overlaid on original images in top panels. The method names are indicated on the left side of each row.

the natural scene images. Although the ground truth results are not given for these images, based on the visual assessment, it can be seen that the GrabCut causes significant over-segmentation. In contrast, the results of the GCN-based approach precisely match the tissue shape.

The main parameter of the method is  $\beta$  used to determine edge weights in (1). In this study, the values of  $\beta = \{0.001, 0.01, 0.1, 1, 10, 100, 1000, 10000\}$  were tested. In the case of ChebNet for the  $k$  values below 100, no significant influence on image segmentation results was observed. However, for values higher or equal to 100, the segmentation result became noisier with the increasing value of the parameter. The GCN mostly failed to produce consistent results for the highest considered value of  $\beta$ . All subpixels were



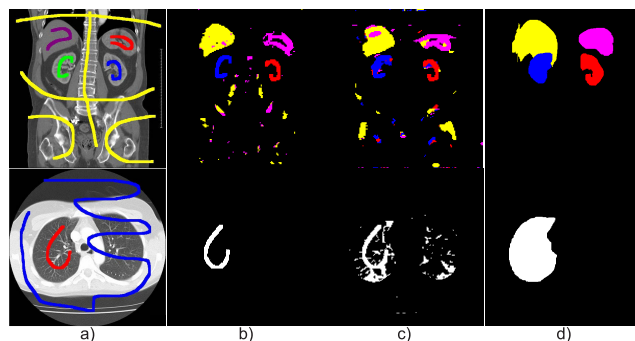
**FIGURE 10.** Influence of the Chebyshev polynomial order  $k$  on image segmentation results using the SegNet variant of the considered GCN architecture.

predominantly assigned one label, with clear scribbles, and other labels randomly distributed over the image. SageNet remained insensitive to all considered values of parameter  $\beta$ .

The additional parameter of the ChebNet variant of the considered GCN architecture was the order of Chebyshev polynomials used to approximate spectral filters that learn on  $k$ -hop neighborhoods of the graph. The best results (regarding the image segmentation accuracy measures) were obtained for  $k = 2$ . These results were presented in Section V. In the case of unambiguous scenes, for  $k = 1$ , the ChebNet was more sensitive to local intensity variations, in some cases producing a small amount of noise. The increasing values of  $k$  decreased the resulting noise level and the level of details in the segmentation result, with the borders of regions being softer and smoother. However, increasing the value of  $k$  may be necessary to perform the segmentation of unambiguous regions. These two effects can be observed in Fig. 10

The property of the proposed approach is that it does not preserve the resulting region connectivity, which can be both an advantage and a disadvantage depending on the application. Notably, the GCN assigns one label to regions similar to the region coarsely indicated by a user regardless of the location in the image and the connectivity to the seed region. As a result, it can extract objects split into disconnected regions. This effect (seen e.g. in Fig. 3f, Fig. 5ab or Fig. 8) is stronger for spectral variant of the GCN. It also discerns the GCN-based approach from the Random Walker, which always outputs the number of regions equal to the number of scribbles and does not assign a region a label when it is disconnected with a scribble (see, e.g., Fig. 8b).

As shown in Fig. 9 the segmentation results obtained for different initial seeds are very similar. They demonstrate that the proposed method moderately depends on the initialization. Good quality region segmentation results can be similarly obtained for scribbles accurately covering regions and scarce seeds (cf. Fig. 9a-c). The method can also perform accurately in the case of very scarce seeds. However, in such a case, image segmentation accuracy may deteriorate for some images with strongly heterogeneous regions, especially for the ChebNet variant of the proposed approach. SageNet



**FIGURE 11.** Results of the GCN-based multi-label segmentation of random medical and microscopic images compared to the results of the Random walker; a) input images; b) ChebNet; c) SageNet; d) Random walker.

remains even resistant to scarce initialization with seeds (cf. Fig. 9d).

The limitation of the GCN-based approach is that it cannot complete a segmentation task when similar regions are assigned different labels. This effect can be observed in Fig. 11 where both ChebNet (Fig. 11b) and SageNet (Fig. 11c) failed to extract regions of interest, while the Random Walker achieved this goal (Fig. 11d).

Depending on the image resolution and complexity of the segmented scene, image segmentation lasted up to a second for the spatial variant (SageNet) and from several seconds to three minutes for the spectral variant (ChebNet) in the case of computations performed on a GPU. For the ChebNet, the computation time significantly increased with the Chebyshev polynomial order. However, the computation time can be decreased by diminishing the number of superpixels. No computer memory limitations were observed regardless of image resolution (the resolution of images considered in this study was not higher than  $512 \times 512$  pixels).

## VII. CONCLUSION

The proposed method introduces graph convolutional neural networks to semi-supervised image segmentation. The approach is universal and performs equally well when applied to images of different characteristics. As a result, it seems to be a noteworthy alternative to existing graph-based segmentation methods. Notably, the proposed GCN-based image segmenter can be helpful when precise object-shaped delineation is required, fine shape details need to be preserved, or disconnected regions of similar characteristics need to be extracted.

The spatial variant seems to be a better choice from the considered graph convolution operators. Although both variants scored similar image segmentation accuracy, the spatial approach was much more efficient in computation time.

Even if the GCN-based approach performs well in binary and multi-label segmentation tasks, there is still room for improvements. They could potentially regard the development of dedicated graph weighting functions. The function used in this work considers only region color or intensity

differences. Extending the weighting function by a factor that considers the spatial location of regions and a mechanism to control the influence of this factor could probably allow adjusting the balance between regions' similarities and compactness of segmentation results. This issue will be the main objective of future works.

## REFERENCES

- [1] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, 2015.
- [2] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, A. C. Berg, and L. Fei-Fei, "ImageNet large scale visual recognition challenge," *Int. J. Comput. Vis.*, vol. 115, no. 3, pp. 211–252, Dec. 2015.
- [3] A. Voulodimos, N. Doulamis, A. Doulamis, and E. Protopapadakis, "Deep learning for computer vision: A brief review," *Comput. Intell. Neurosci.*, vol. 2018, pp. 1–13, Feb. 2018.
- [4] S. Hao, Y. Zhou, and Y. Guo, "A brief survey on semantic segmentation with deep learning," *Neurocomputing*, vol. 406, pp. 302–321, Sep. 2020.
- [5] Y. Zhang, X. Li, M. Lin, B. Chiu, and M. Zhao, "Deep-recursive residual network for image semantic segmentation," *Neural Comput. Appl.*, vol. 32, pp. 12935–12947, Jan. 2020.
- [6] Y. Mo, Y. Wu, X. Yang, F. Liu, and Y. Liao, "Review the state-of-the-art technologies of semantic segmentation based on deep learning," *Neurocomputing*, vol. 493, pp. 626–646, Jul. 2022.
- [7] I. Ulku and E. Akagunduz, "A survey on deep learning-based architectures for semantic segmentation on 2D images," *Appl. Artif. Intell.*, vol. 36, no. 1, 2022, Art. no. 2032924.
- [8] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *Medical Image Computing and Computer-Assisted Intervention—MICCAI 2015* (Lecture Notes in Computer Science), vol. 9351. Cham, Switzerland: Springer, 2015, pp. 234–241.
- [9] S. Minaee, Y. Boykov, F. Porikli, A. Plaza, N. Kehtarnavaz, and D. Terzopoulos, "Image segmentation using deep learning: A survey," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 44, no. 7, pp. 3523–3542, Jul. 2022.
- [10] M. Lin, S. Momin, Y. Lei, H. Wang, W. J. Curran, T. Liu, and X. Yang, "Fully automated segmentation of brain tumor from multiparametric MRI using 3D context deep supervised U-Net," *Med. Phys.*, vol. 48, no. 8, pp. 4365–4374, 2021.
- [11] M. Lin, M. Jiang, M. Zhao, E. Ukwatta, J. A. White, and B. Chiu, "Cascaded triplanar autoencoder M-Net for fully automatic segmentation of left ventricle myocardial scar from three-dimensional late gadolinium-enhanced MR images," *IEEE J. Biomed. Health Informat.*, vol. 26, no. 6, pp. 2582–2593, Jun. 2022.
- [12] N. Xu, B. Price, S. Cohen, J. Yang, and T. Huang, "Deep interactive object selection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 373–381.
- [13] J. Liew, Y. Wei, W. Xiong, S.-H. Ong, and J. Feng, "Regional interactive image segmentation networks," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2017, pp. 2746–2754.
- [14] Z. Li, Q. Chen, and V. Koltun, "Interactive image segmentation with latent diversity," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 577–585.
- [15] G. Wang, W. Li, M. A. Zuluaga, R. Pratt, P. A. Patel, M. Aertsen, T. Doel, A. L. David, J. Deprest, S. Ourselin, and T. Vercauteren, "Interactive medical image segmentation using deep learning with image-specific fine tuning," *IEEE Trans. Med. Imag.*, vol. 37, no. 7, pp. 1562–1573, Jul. 2018.
- [16] K.-K. Maninis, S. Caelles, J. Pont-Tuset, and L. Van Gool, "Deep extreme cut: From extreme points to object segmentation," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 616–625.
- [17] W.-D. Jang and C.-S. Kim, "Interactive image segmentation via back-propagating refinement scheme," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 5292–5301.
- [18] J. Peng, G. Estrada, M. Pedersoli, and C. Desrosiers, "Deep co-training for semi-supervised image segmentation," *Pattern Recognit.*, vol. 107, Nov. 2020, Art. no. 107269.
- [19] Z. Yang, H. Yu, Y. He, W. Sun, Z.-H. Mao, and A. Mian, "Fully convolutional network-based self-supervised learning for semantic segmentation," *IEEE Trans. Neural Netw. Learn. Syst.*, early access, May 11, 2022, doi: 10.1109/TNNLS.2022.3172423.
- [20] S. Wang, W. Chen, S. M. Xie, G. Azzari, and D. B. Lobell, "Weakly supervised deep learning for segmentation of remote sensing imagery," *Remote Sens.*, vol. 12, no. 2, p. 207, 2020.
- [21] J. Santner, T. Pock, and H. Bischof, "Interactive multi-label segmentation," in *Computer Vision—ACCV 2010*, R. Kimmel, R. Klette, and A. Sugimoto, Eds. Berlin, Germany: Springer, 2011, pp. 397–410.
- [22] C. Nieuwenhuis and D. Cremers, "Spatially varying color distributions for interactive multilabel segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 5, pp. 1234–1247, May 2013.
- [23] M. Mostajabi, P. Yadollahpour, and G. Shakhnarovich, "Feedforward semantic segmentation with zoom-out features," 2014, *arXiv:1412.0774*.
- [24] Y. Li and A. Gupta, "Beyond grids: Learning graph representations for visual recognition," in *Advances in Neural Information Processing System*, vol. 31, S. Bengio, H. Wallach, H. Larochelle, K. Grauman, N. Cesa-Bianchi, and R. Garnett, Eds. Red Hook, NY, USA: Curran Associates, 2018, pp. 9225–9235.
- [25] J. Shi and J. Malik, "Normalized cuts and image segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 22, no. 8, pp. 888–905, Aug. 2000.
- [26] Y. Y. Boykov and M. Jolly, "Interactive graph cuts for optimal boundary & region segmentation of objects in N-D images," in *Proc. IEEE Int. Conf. Comput. Vis.*, vol. 1, Jul. 2001, pp. 105–112.
- [27] Y. Boykov and V. Kolmogorov, "An experimental comparison of min-cut/max-flow algorithms for energy minimization in vision," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 26, no. 9, pp. 1124–1137, Sep. 2004.
- [28] L. Grady, "Random walks for image segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 28, no. 11, pp. 1768–1783, Nov. 2006.
- [29] M. Henaff, J. Bruna, and Y. LeCun, "Deep convolutional networks on graph-structured data," 2015, *arXiv:1506.05163*.
- [30] T. N. Kipf and M. Welling, "Semi-supervised classification with graph convolutional networks," in *Proc. Int. Conf. Learn. Represent.*, 2017, pp. 1–14.
- [31] C. Zhuang and Q. Ma, "Dual graph convolutional networks for graph-based semi-supervised classification," in *Proc. World Wide Web Conf. (WWW)*. Geneva, Switzerland: International World Wide Web Conferences Steering Committee, 2018, pp. 499–508.
- [32] L. Zhang, X. Li, A. Arnab, K. Yang, Y. Tong, and P. H. S. Torr, "Dual graph convolutional network for semantic segmentation," 2019, *arXiv:1909.06121*.
- [33] M. Edwards, X. Xie, R. I. Palmer, G. K. L. Tam, R. Alcock, and C. Roobottom, "Graph convolutional neural network for multi-scale feature learning," *Comput. Vis. Image Understand.*, vol. 194, May 2020, Art. no. 102881.
- [34] F. Ma, F. Gao, J. Sun, H. Zhou, and A. Hussain, "Attention graph convolution network for image segmentation in big SAR imagery data," *Remote Sens.*, vol. 11, no. 21, p. 2586, 2019.
- [35] Y. Lu, Y. Chen, D. Zhao, and J. Chen, "Graph-FCN for image semantic segmentation," in *Advances in Neural Networks—ISNN 2019*, H. Lu, H. Tang, and Z. Wang, Eds. Cham, Switzerland: Springer, 2019, pp. 97–105.
- [36] S. H. Kwok and A. G. Constantinides, "A fast recursive shortest spanning tree for image segmentation and edge detection," *IEEE Trans. Image Process.*, vol. 6, no. 2, pp. 328–332, Feb. 1997.
- [37] P. F. Felzenszwalb and D. P. Huttenlocher, "Efficient graph-based image segmentation," *Int. J. Comput. Vis.*, vol. 59, no. 2, pp. 167–181, Sep. 2004.
- [38] W. Tao, H. Jin, and Y. Zhang, "Color image segmentation based on mean shift and normalized cuts," *IEEE Trans. Syst., Man, Cybern. B, Cybern.*, vol. 37, no. 5, pp. 1382–1389, Oct. 2007.
- [39] Y. Boykov and G. Funka-Lea, "Graph cuts and efficient N-D image segmentation," *Int. J. Comput. Vis.*, vol. 70, no. 2, pp. 109–131, Nov. 2006.
- [40] C. Rother, V. Kolmogorov, and A. Blake, "GrabCut: Interactive foreground extraction using iterated graph cuts," *ACM Trans. Graph.*, vol. 23, no. 3, pp. 309–314, 2004.
- [41] Y. Li, J. Sun, C. Tang, and H. Shum, "Lazy snapping," *ACM Trans. Graph.*, vol. 23, no. 3, pp. 303–308, Aug. 2004.
- [42] R. Fu, B. Li, Y. Gao, and P. Wang, "Fully automatic figure-ground segmentation algorithm based on deep convolutional neural network and GrabCut," *IET Image Process.*, vol. 10, no. 12, pp. 937–942, 2016.
- [43] S. M. Jaisakthi, P. Mirunalini, and C. Aravindan, "Automated skin lesion segmentation of dermoscopic images using GrabCut and  $k$ -means algorithms," *IET Comput. Vis.*, vol. 12, no. 8, pp. 1088–1095, 2018.
- [44] J. Mao, K. Wang, Y. Hu, W. Sheng, and Q. Feng, "GrabCut algorithm for dental X-ray images based on full threshold segmentation," *IET Image Process.*, vol. 12, no. 12, pp. 2330–2335, 2018.



- [45] A. Fabijańska and J. Gocłowski, "The segmentation of 3D images using the random walking technique on a randomly created image adjacency graph," *IEEE Trans. Image Process.*, vol. 24, no. 2, pp. 524–537, Feb. 2015.
- [46] D. Ren, Z. Jia, Y. Jie, and N. K. Kasabov, "A practical GrabCut color image segmentation based on Bayes classification and simple linear iterative clustering," *IEEE Access*, vol. 5, pp. 18480–18487, 2017.
- [47] J. Zhou, G. Cui, S. Hu, Z. Zhang, C. Yang, Z. Liu, L. Wang, C. Li, and M. Sun, "Graph neural networks: A review of methods and applications," *AI Open*, vol. 1, pp. 57–81, Jan. 2020.
- [48] F. Liang, C. Qian, W. Yu, D. Griffith, and N. Golmie, "Survey of graph neural networks and applications," *Wireless Commun. Mobile Comput.*, vol. 2022, Jul. 2022, Art. no. 9261537.
- [49] S. Fu, G. Wang, S. Xia, and L. Liu, "Deep multi-granularity graph embedding for user identity linkage across social networks," *Knowl.-Based Syst.*, vol. 193, Apr. 2019, Art. no. 105301.
- [50] S. A. Ríos, F. Aguilera, J. D. Nuñez-Gonzalez, and M. Grana, "Semantically enhanced network analysis for influencer identification in online social networks," *Neurocomputing*, vols. 326–327, pp. 71–81, Jan. 2019.
- [51] P. Sen, G. Namata, M. Bilgic, L. Getoor, B. Gallagher, and T. Eliassirad, "Collective classification in network data," *AI Mag.*, vol. 29, no. 3, pp. 93–106, Sep. 2008.
- [52] A. Rodriguez, B. Kim, J.-M. Lee, B.-Y. Coh, and M. K. Jeong, "Graph kernel based measure for evaluating the influence of patents in a patent citation network," *Expert Syst. Appl.*, vol. 42, no. 3, pp. 1479–1486, 2015.
- [53] M. Eto, "Extended co-citation search: Graph-based document retrieval on a co-citation network containing citation context information," *Inf. Process. Manag.*, vol. 56, no. 6, 2019, Art. no. 102046.
- [54] S. Kearnes, K. McCloskey, M. Berndl, V. Pande, and P. Riley, "Molecular graph convolutions: Moving beyond fingerprints," *J. Comput.-Aided Mol. Des.*, vol. 30, no. 8, pp. 595–608, Aug. 2016.
- [55] J. Gilmer, S. S. Schoenholz, P. F. Riley, O. Vinyals, and G. E. Dahl, "Neural message passing for quantum chemistry," in *Proc. Int. Conf. Mach. Learn. (ICML)*, 2017, pp. 1263–1272.
- [56] W. Min and L. Wynter, "Real-time road traffic prediction with spatio-temporal correlations," *Transp. Res. C Emerg. Technol.*, vol. 19, no. 4, pp. 606–616, 2011.
- [57] L. Lin, Z. He, and S. Peeta, "Predicting station-level hourly demand in a large-scale bike-sharing network: A graph convolutional neural network approach," *Transp. Res. C, Emerg. Technol.*, vol. 97, pp. 258–276, Dec. 2018.
- [58] Z. Wu, S. Pan, F. Chen, G. Long, C. Zhang, and P. S. Yu, "A comprehensive survey on graph neural networks," *IEEE Trans. Neural Networks Learn. Syst.*, vol. 32, no. 1, pp. 4–24, 2021, doi: 10.1109/TNNLS.2020.2978386.
- [59] J. Bruna, W. Zaremba, A. Szlam, and Y. LeCun, "Spectral networks and locally connected networks on graphs," 2013, *arXiv:1312.6203*.
- [60] O. Rippel, J. Snoek, and R. P. Adams, "Spectral representations for convolutional neural networks," in *Proc. Adv. Neural Inf. Process. Syst.*, Jan. 2015, pp. 2449–2457.
- [61] M. Defferrard, X. Bresson, and P. Vandergheynst, "Convolutional neural networks on graphs with fast localized spectral filtering," in *Proc. Int. Conf. Neural Inf. Process. Syst. (NIPS)*. Red Hook, NY, USA: Curran Associates, 2016, pp. 3844–3852.
- [62] D. Duvenaud, D. Maclaurin, J. Aguilera-Iparraguirre, R. Gómez-Bombarelli, T. Hirzel, A. Aspuru-Guzik, and R. P. Adams, "Convolutional networks on graphs for learning molecular fingerprints," 2015, *arXiv:1509.09292*.
- [63] J. Atwood and D. Towsley, "Diffusion-convolutional neural networks," in *Proc. Int. Conf. Neural Inf. Process. Syst. (NIPS)*. Red Hook, NY, USA: Curran Associates, 2016, pp. 2001–2009.
- [64] M. Niepert, M. Ahmed, and K. Kutzkov, "Learning convolutional neural networks for graphs," in *Proc. Int. Conf. Mach. Learn.*, vol. 48, M. F. Balcan and K. Q. Weinberger, Eds., New York, NY, USA, 2016, pp. 2014–2023.
- [65] P. Velickovic, G. Cucurull, A. Casanova, A. Romero, P. Lio, and Y. Bengio, "Graph attention networks," 2017, *arXiv:1710.10903*.
- [66] W. L. Hamilton, R. Ying, and J. Leskovec, "Inductive representation learning on large graphs," 2017, *arXiv:1706.02216*.
- [67] D. Martin, C. Fowlkes, D. Tal, and J. Malik, "A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics," in *Proc. Int. Conf. Comput. Vis.*, vol. 2, 2001, pp. 416–423.



**ANNA FABIJAŃSKA** received the Master of Engineering, Doctoral, and Habilitation degrees in computer science from the Faculty of Electrical, Electronic, Computer and Control Engineering of the Lodz University of Technology, in 2006, 2007, and 2013, respectively. She is currently a Full Professor in computer science with the Institute of Applied Computer Science, Lodz University of Technology, Poland. She was granted the professor title, in 2022. She has authored/coauthored over 120 scientific papers. Her scientific research interests include digital image processing and analysis, machine vision, and artificial intelligence (especially deep learning). In particular, they are concerned with developing dedicated image processing pipelines for computer-aided diagnosis systems and computer vision applications in various fields of science and industry. She was a beneficiary of the Ministry of Science and Higher Education Fellowship for outstanding young scientists, a Recipient of the Foundation for Polish Science (FNP) START Fellowship, and the Leader of several scientific grants. From 2016 to 2021, she was a member of the Polish Young Academy of the Polish Academy of Sciences. Since 2017, she has been a member of the Committee on Informatics of the Polish Academy of Sciences.

• • •