Mathematical Biosciences and Engineering

http://www.aimspress.com/journal/mbe

MBE, 21(2): 2163–2188.
DOI: 10.3934/mbe.2024095
Received: 14 October 2023
Revised: 19 December 2023
Accepted: 22 December 2023
Published: 10 January 2024

Research article

Automatic recognition of white blood cell images with memory efficient superpixel metric GNN: SMGNN

Yuanhong Jiang[1], Yiqing Shen[3], Yuguang Wang[2,4] and Qiaoqiao Ding[2,*]

[1] School of Mathematical Sciences, MOE-LSC, Shanghai Jiao Tong University, Shanghai 200030, China

[2] Institute of Natural Sciences, Shanghai Jiao Tong University, Shanghai 200030, China

[3] Department of Computer Science, Johns Hopkins University, USA

[4] Shanghai Artificial Intelligence Laboratory, Shanghai 200433, China

* **Correspondence:** Email: dingqiaoqiao@sjtu.edu.cn.

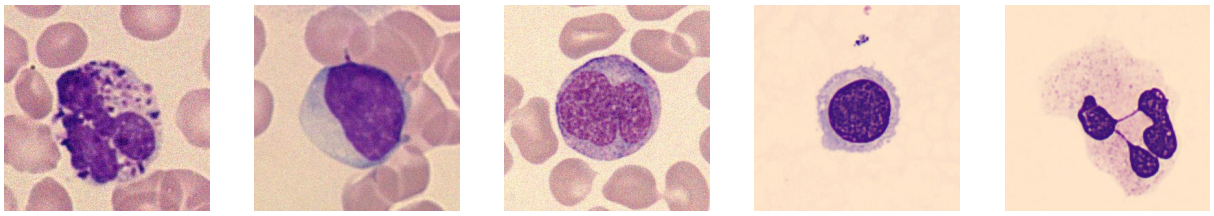
Abstract: An automatic recognizing system of white blood cells can assist hematologists in the diagnosis of many diseases, where accuracy and efficiency are paramount for computer-based systems. In this paper, we presented a new image processing system to recognize the five types of white blood cells in peripheral blood with marked improvement in efficiency when juxtaposed against mainstream methods. The prevailing deep learning segmentation solutions often utilize millions of parameters to extract high-level image features and neglect the incorporation of prior domain knowledge, which consequently consumes substantial computational resources and increases the risk of overfitting, especially when limited medical image samples are available for training. To address these challenges, we proposed a novel memory-efficient strategy that exploits graph structures derived from the images. Specifically, we introduced a lightweight superpixel-based graph neural network (GNN) and broke new ground by introducing superpixel metric learning to segment nucleus and cytoplasm. Remarkably, our proposed segmentation model superpixel metric graph neural network (SMGNN) achieved state of the art segmentation performance while utilizing at most 10000× less than the parameters compared to existing approaches. The subsequent segmentation-based cell type classification processes showed satisfactory results that such automatic recognizing algorithms are accurate and efficient to execeute in hematological laboratories. Our code is publicly available at https://github.com/jyh6681/SPXL-GNN.

Keywords: GNN; superpixel metric learning; memory efficient model; white blood cell segmentation; cell type classification

1. Introduction

White blood cells (WBCs), also known as leukocytes, play a pivotal role in the immune system's defense against various infections. Accurate quantification and classification of WBCs can provide valuable insights for diagnosing a wide range of diseases, including infections, leukemia [1], and AIDS [2]. In the laboratory environment, the traditional examination of blood smears was a laborious and time-consuming manual task. However, with the advent of computer-aided automatic cell analysis systems, rapid and high-throughput image analysis tasks can now be accomplished [3]. Some automatic recognizing system of white blood cells typically entails three major steps: Image acquisition, cell segmentation and cell type classification. Among these steps, cell segmentation is widely recognized as the most crucial and challenging one, as it significantly influences the accuracy and computational complexity of subsequent processes [4]. Some segmentation-free methods take the whole image as input to the classifier without extracting region of interest (ROI) [5–7].

Accurately segmenting WBCs in cell images, thereby distinguishing between lymphocytes, monocytes, eosinophils, basophils, and neutrophils as shown in Figure 1, provides a wealth of crucial information for hematological diagnostics [8, 9]. However, achieving high-quality images requires careful consideration of various factors, including image resolution, exposure duration, illumination levels, and the proper utilization of optical filters. If inappropriate factors are chosen in the imaging process, it can adversely affect image quality, thereby posing challenges for analyzing WBC images.



(a) Basophils (b) Lymphocytes (c) Eosinophils (d) monocytes (e) Neutrophils

Figure 1. Sample images of five different types of WBCs. The colors of different images exhibit significant variations, and the boundaries of the cytoplasm are often ambiguous, posing a considerable challenge in accurately recognizing the shape of WBCs.

Deep learning, particularly convolutional neural networks (CNNs), has revolutionized medical image segmentation [10]. The U-shaped network (U-Net) [11], a symmetrical encoder-decoder convolutional network featuring skip connections, stands as a prime example. The U-Net has gained significant popularity in medical image processing, especially for datasets with limited samples. Extensive research has demonstrated the effectiveness of this architecture in extracting multi-scale image features [12].

Subsequent iterations, such as U-Net++ [13] and U-Net3+ [14], have been proposed to further enhance performance. U-Net++ introduces nested and dense skip connections to address the semantic gap and incorporates deep supervision learning techniques to improve segmentation performance [13]. Regularized U-Net (RU-Net) [15] proposed a new activation function with piecewise smooth effect to solve the under-segmentation problem. There have been various advancements in CNN-based architectures for image segmentation. One notable example is the multiscale information fusion network (MIF-Net), which incorporates the boundary splitter and information fusion mechanisms using

strided convolutions [16]. These techniques contribute to improved segmentation accuracy. In parallel, Transformer-based approaches, such as Vision Transformer (ViT) [17] and Swin Transformer [18], have emerged, leveraging self-attention mechanisms for better feature extraction. Specifically, the ViT partitions images into nonoverlapping patches and treats the patches as sequence data, where the self-attention mechanism is subsequently used to extract long-range information among patches. Furthermore, the Swin Transformer applies a shifted window to make ViT more computationally efficient. Though the original ViT model exhibits significantly better performance on large objects, it obtains lower performance on small objects [19]. This limitation might arise from the fixed scale of patches generated by transformer-based methods. A potential solution to enhance the small object detection (SOD) capability is to explore more refined patch sizes, which might increase the computational cost. Notably, U-Net architectures have also incorporated transformers, yielding U-Net Transformer [20], Medical Transformer [21], and Swin-Unet [22]—all of which have set new performance benchmarks in medical image segmentation. However, these architectures, rooted in pixel-based learning demand substantial memory resources, leading to inefficiencies, especially when available training samples are scanty [23]. Here, large models might face narrowing expressivity gaps against parameter-efficient counterparts. To mitigate this, embedding prior knowledge can reduce the computational burden.

Graph structure data provides an elegant way off describing the geometry of data, which contains abundant relational information. For example, diverse types of relational systems or structured entities can be described by graphs to include the interior connections, where some typical examples include particle system analysis [24, 25], social networks [26], and molecular properties prediction [27]. Correspondingly, graph neural networks (GNNs) are specifically designed to process graph data [28–32], where researchers have developed graph convolutional networks (GCNs) and various variants to update node features by aggregating information from neighboring nodes.

The transformation of image data, particularly those without an inherent geometric structure, into graph data, represents a substantial challenge. This challenge is twofold: Encoding Euclidean space data into graph representations and decoding them back to their original image domain. A prevalent approach to address this involves the use of a patch graph method, where image patches are treated as graph nodes. For example, the Graph-FCN [33] applies a fully convolutional network (FCN) to extract image features, and the graph structure is constructed based on the k nearest neighbor (kNN) methods where the weight adjacent matrix is generated with the Gaussian kernel function. In [34], the dual graph convolution network (DGCNet) constructs the graph structure not only on the spatial domain but also on the feature domain. In the semantic segmentation task, the bilinear interpolation upsampling operation acted on the downsampled output of the DGCNet to recover the same image size as the label. There has also been recent work aiming to combine the local feature extraction ability of CNNs with the long-range interaction ability of GNNs. The vision graph U-Net (VGU-Net) model was proposed to construct multi-scale graph structures, enhancing the model's learning capacity [35].

However, while the patch-based method offers convenience in graph construction, it has its limitations. The fixed structure of image patches can lead to the omission of critical boundary details. An alternative lies in the superpixel approach. Superpixels, by design, can dramatically lower both computational and memory costs for image processing. Since image superpixel can significantly reduce the computational and memory overhead for image processing tasks, superpixel methods are commonly implemented as a preprocessing step before the deep reasoning models [36–43]. Various superpixel methods over-segment the image into multiple nonoverlapped regions based on the pixel features and homogeneous pixels are

grouped inside single superpixel. Traditional superpixel generation method can be roughly divided into graph-based [44–46] and clustering-based [47–50] methods. These methods are efficient and fast to generate high-quality superpixels and require no human label and less memory in computing. Recently, the deep learning-based approaches are employed in superpixel sampling [51–55]. These methods are accurate but not efficient in memory saving since learning high-level image features requires a relatively large amount of convolution kernel parameters. Based on the pre-computed superpixel graph and GNN, [56] captures global feature interactions for brain tumor segmentation. In [57], superpixel-based graph data and an edge-labeling graph neural network (EGNN) [58] are implemented for biological image segmentation.

Medical image segmentation has long depended on the precision achieved by supervised learning methods. Yet, the perennial issue remains; that is, the paucity of richly labeled datasets in clinical contexts. This limitation has driven the pivot to metric learning, which disrupts the entrenched belief that robust intelligence is the sole preserve of abundant labeled data [59–62]. Metric learning approaches, such as contrastive methods, learn representations in a discriminative manner by contrasting positive sample pairs against negative pairs. By tapping into vast reservoirs of unlabeled samples, they set the stage for pretraining deep learning models. The subsequent phase involves meticulous fine-tuning, utilizing just a fraction of labeled samples. Remarkably, the outcome is a model performance that stands shoulder to shoulder with traditional supervised strategies.

Notably, there's a burgeoning interest in supervised metric learning methods, specifically tailored to unravel cross-image intricacies. These techniques use sample labels as the blueprint to categorize them into positive and negative sets [63]. The confluence of metric learning methods offers deep learning models a unique advantage. By bridging labeled and unlabeled data, they are empowered to deliver stellar results, even when navigating the constraints of scantily labeled samples.

In this work, we propose a novel approach for WBCs segmentation, namely superpixel metric graph neural network (SMGNN). The core strength of SMGNN lies in its dual promise: delivering unparalleled accuracy while simultaneously optimizing memory efficiency. The foundation of our technique is a superpixel graph constructed from image data. This restructuring drastically diminishes the problem's dimensionality and serves as a conduit for infusing abundant prior information into the graph data. In addition to leveraging prior knowledge on a single training sample, our proposed approach introduces superpixel metric learning to capture "global" context across the training samples. In clinical image segmentation scenarios with limited training samples, we believe incorporating this "global" context can enhance the expressivity of deep learning models. Our proposed metric learning operates on the superpixel embeddings rather than the vast number of pixel embeddings, which offers the advantage of memory saving.

The contributions of this paper can be summarized as follows.

- Our proposed lightweight SMGNN significantly reduces the learnable parameters by at most 10000 times compared with mainstream segmentation models.
- Our proposed superpixel-based model reduces the problem size and poses rich prior knowledge to the rarely considered graph structure data, which helps SMGNN achieve state of the art (SOTA) performance on WBC images.
- We innovatively propose superpixel metric learning according to the definition of superpixel metric score, which is more efficient than pixel-level metric learning.

- The whole deep learning-based nucleus and cytoplasm segmentation and cell type classification system is accurate and efficient to execute in hematological laboratories.

The remaining sections of this article are structured as follows. Section 2 describes our methodology in depth. Section 3 depicts the workflow and architecture of our proposed model. In Section 4, through extended segmentation experiments, the SMGNN model achieves SOTA segmentation performance in terms of both accuracy and memory efficiency. The cell type classification task is conducted with a lightweight residual network (ResNet) based on the segmentation result. The whole procedure of the proposed automatic recognition system is shown as Figure 3.

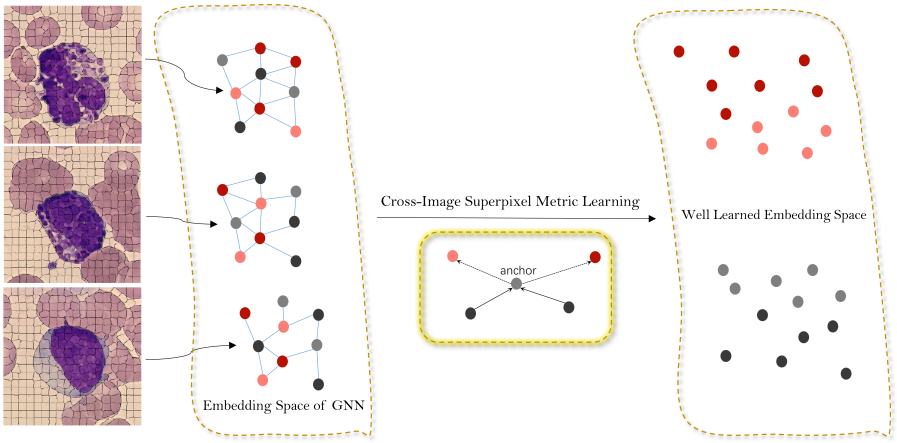


Figure 2. The main idea underlying our approach is to learn the distance between superpixel embeddings using the superpixel metric score, which is the ratio of the majority class inside the superpixel. Given the anchor embedding, similar embeddings with approximate metric scores will be pulled close and dissimilar embeddings will be pushed away. With the help of metric learning, the cross-image global context can be captured and a better embedding space will be learned.

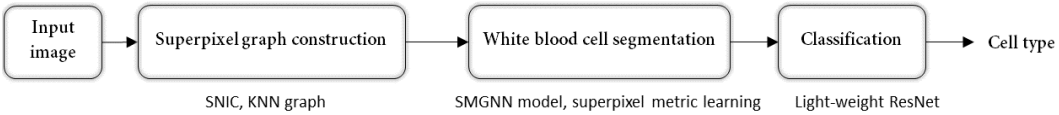


Figure 3. The overall work-flow of the proposed automatic WBC recognition system.

2. Methodolgy of superpixel metric

Deep learning methods for medical image processing have predominantly concentrated on discerning the local context, which refers to inter-pixel dependencies within individual images [63]. However, there's a missed opportunity: Capturing the "global" context that exists between training samples. While pixel-level contrast or metric learning provides a way to bridge this gap, the sheer computational and memory overheads—due to contrast or metric computations spanning every pixel pair—render them less feasible. We propose an innovative efficient superpixel-level metric learning on metric loss, which not only captures the desired global context but does so while drastically cutting computational and memory costs.

2.1. Compression ratio on image data

The utility of superpixel methods in image data preprocessing is well acknowledged, particularly for their ability to condense data and reduce computational demands. Consequently, this study capitalizes on these advantages, transforming the image data into a more compact graph representation. Within this framework, each superpixel evolves into a graph node. The interconnectedness of these nodes—whether driven by spatial positioning or feature similarity—determines the graph's topology. These node features aren't rigid; their definition can range from basic five-dimensional attributes encompassing color (three dimensions) and location (two dimensions) to more intricate data points like histograms, positional variance and variations in pixel values.

Given that the adjacency matrix exhibits sparse characteristics, it's predominantly the node features that dictate memory consumption. Opting for the more straightforward features facilitates significant data compression. Specifically, for three-channel red-green-blue (RGB) images, we've discerned a compression ratio roughly represented as $c = \frac{\text{graph data}}{\text{image data}} \approx \frac{5K}{3n}$. Importantly, this efficient compression does not compromise on quality. Our subsequent numerical experiments demonstrate that this ratio is concomitant with optimal segmentation outcomes.

2.2. Quality of superpixel and reconstruction score

The segmentation task relies on the quality of the generated superpixels, and good superpixel results coherent with the boundary of labeled images. Suppose $\mathbf{X} \in \mathbb{R}^{H\times W\times 3}$ is the input image. Let $\mathcal{V}$ be the set of all superpixels, $N = H \times W$ be the number of pixels, $K = |\mathcal{V}|$ be the number of superpixels, and $\mathbf{Q} \in \mathbb{R}^{N\times K}$ be the association matrix between pixels and superpixels, then we have

$$\mathbf{Q}_{i,j} = \begin{cases} 1, & \text{if } \hat{\mathbf{X}}_i \in \mathcal{V}_j \\ 0, & \text{otherwise} \end{cases}, \hat{\mathbf{X}} = \text{Flatten}(\mathbf{X}) \in \mathbb{R}^{N\times 3},$$

where $\mathcal{V}_j$ is the jth superpixel. Flatten operation converts a two-dimensional image matrix into a one-dimensional vector. The role of the association matrix builds the bridge between image space and graph space. For computation convenience, we define the column normalized association matrix as

$$\overline{\mathbf{Q}}_{i,j} = \begin{cases} 1/[\mathcal{S}_j], & \text{if } \hat{\mathbf{X}}_i \in \mathcal{V}_j \\ 0, & \text{otherwise} \end{cases}, [\mathcal{S}_j] = \sum_{x\in\mathcal{V}_j} 1.$$

Let $\mathbf{Y} \in \mathbb{R}^N$ be the label of the image pixels and $\mathcal{Y} \in \mathbb{R}^K$ be superpixel metric or metric score of the graph data. Using the pixel label, we can formulate metric score as

$$\mathcal{Y}_j = \frac{\sum_{x\in\mathcal{V}_j} \mathbf{Y}_x}{\sum_{x\in\mathcal{V}_j} 1}, \tag{2.1}$$

for the jth superpixel. $\mathcal{Y}$ can also be efficiently computed with column normalized association matrix, i.e., $\mathcal{Y} = \overline{\mathbf{Q}}^T\mathbf{Y}$. We can back-project the supepixel label to image space by association matrix, i.e., $\overline{\mathbf{Y}} = \mathbf{Q}\mathcal{Y}$. We define the intersection of union (IoU) reconstruction score to evaluate the quality, which reads

$$IoU_r = \frac{1}{c}\sum_{i=0}^{c} \frac{\{x|\overline{\mathbf{Y}}_x = i\} \cap \{x|\mathbf{Y}_x = i\}}{\{x|\overline{\mathbf{Y}}_x = i\} \cup \{x|\mathbf{Y}_x = i\}}, \tag{2.2}$$

where $\{x|\mathbf{Y}_x = i\}$ denotes the set of pixels whose class label equals i and c is the number of classes.

2.3. Lightweight GNNs for superpixel embedding

GNNs have the advantage of being lightweight compared to other deep learning models that often require deeper and larger networks to achieve higher performance. GNNs leverage relational information and utilize shallow layers to achieve satisfactory results.

Given an undirected attributed graph $\mathcal{G} = (\mathcal{V}, \mathcal{E}, \mathcal{S})$, $\mathcal{G}$ consists of a nonempty finite set of $K = |\mathcal{V}|$ nodes $\mathcal{V}$ and a set of edges $\mathcal{E}$ between node pairs. Denote $\mathcal{A} \in \mathbb{R}^{K\times K}$ the graph adjacency matrix and $\mathcal{S} \in \mathbb{R}^{K\times d}$ the node attributes. A graph convolution learns a matrix representation $\mathcal{H}$ that embeds the structure $\mathcal{A}$ and feature matrix $\mathcal{S} = \left\{\mathcal{S}_j\right\}_{j=1}^{N}$ with $\mathcal{S}_j$ for node j. Most graph convolutions follow the message passing [64] update scheme, which finds a central node's smooth representation by aggregating its 1-hop neighbor information. At layer ℓ, the propagation for the ith node reads

$$\mathcal{H}_i^{\ell} = \gamma\left(\mathcal{H}_i^{\ell-1}, \square_{j\in\mathcal{N}(i)}\phi\left(\mathcal{H}_i^{\ell-1}, \mathcal{H}_j^{\ell-1}, \mathcal{A}_{ij}\right)\right), \mathcal{H}^0 = \mathcal{S} \tag{2.3}$$

where $\square(\cdot)$ is a differentiable and permutation invariant aggregation function, such as summation, average, or maximization. The set $\mathcal{N}(i)$ includes $\mathcal{V}_i$ and its 1-hop neighbors. Both $\gamma(\cdot)$ and $\phi(\cdot)$ are differentiable aggregation functions, such as multilayer perceptrons (MLPs).

In our approach, we construct graph data using superpixels and their predefined relationships. GNNs learn features from the graph space to enhance the segmentation capability. We employ the graph isomorphism network (GIN) [32] as the backbone graph representation network. At each layer ℓ, the GIN model updates the ith node representation as follows:

$$\mathcal{H}_i^{(\ell)} = \mathrm{MLP}^{(\ell)}\left(\left(1 + \mathbf{w}^{(\ell)}\right)\cdot\mathcal{H}_i^{(\ell-1)} + \sum_{j\in\mathcal{N}(i)}\mathcal{H}_j^{(\ell-1)}\right), \tag{2.4}$$

where $\mathbf{w}$ is a learnable weight.

GIN has been proven to possess expressive power equivalent to the 2-Weisfeiler-Lehman test [65]. By utilizing GIN, we can effectively extract informative features from the superpixel graph, enabling accurate and efficient segmentation performance.

2.4. Memory efficient metric learning

Although pixel-wise contrast can learn the global context to form a good segmentation embedding space [63], computing the contrastive loss requires using training image pixels, which leads to a significant amount of computation and memory cost. In this study, we propose superpixel-based methods that can significantly reduce the number of data samples from N to K, and we introduce a memory-efficient distance-based metric loss function.

The fundamental concept of metric learning is to bring similar samples closer together in the embedding space while pushing dissimilar samples further apart. However, pixel-wise contrast methods that involve setting a large number of anchor pixels and using tensor multiplication to compute positive similarity incurs high memory costs. To address this, we define the superpixel metric loss using the mean square error (MSE) between the similarity and metric score of the embeddings, as follows:

$$\mathcal{L}_{\mathrm{SM}}(\mathcal{A}, \mathcal{R}) = MSE(SIM(\mathcal{A}, \mathcal{R}), Metric(\mathcal{A}, \mathcal{R})), \tag{2.5}$$

$$SIM_{i,j}(\mathcal{A}, \mathcal{R}) = a\cdot cos(\mathcal{A}_i, \mathcal{R}_j) + b, SIM \in \mathbb{R}^{n_1\times n_2}, \tag{2.6}$$

$$Metric_{i,j}(\mathcal{A},\mathcal{R}) = \frac{\mathcal{Y}(\mathcal{A}_i) + \mathcal{Y}(\mathcal{R}_j)}{2}, Metric \in \mathbb{R}^{n_1 \times n_2}, \tag{2.7}$$

where n_1 and n_2 represent the number of anchor samples $\mathcal{A} \in \mathbb{R}^{n_1 \times d}$ and reference samples $\mathcal{R} \in \mathbb{R}^{n_2 \times d}$, respectively. Here, d denotes the dimension of the embedding, and a and b are learnable parameters. Additionally, $\mathcal{Y}(x)$ represents the superpixel label of node x, which is defined by Eq (2.1). It's important to note that the anchor/reference samples are not restricted to being from the same image $\mathcal{A}$. The objective of Eq (2.5) is to bring the embeddings of similar superpixel samples closer together and push dissimilar ones apart.

3. The work-flow and architecture of SMGNN

We use the parameter-free methods to generate superpixels [47, 66], on which we construct graph structure with two alternative strategies. Each superpixel is treated as a node and the mean RGB values and mean postion value consist five dimension node features. Suppose the mean scale of superpixels $S = \frac{H \times W}{K}$. We can define the adjacency between nodes according to their positional relation as

$$\mathcal{E}_{i,j} = \begin{cases} 1, & |x_i - x_j| + |y_i - y_j| < \alpha \sqrt{S} \\ 0, & \text{otherwise} \end{cases},$$

where (x_i, y_i) and (x_j, y_j) are the positions of superpixel i and superpixel j respectively, and $\alpha \geq 2$ is a hyperparameter to control the number of neighboring nodes. The above definition will pose strong local connectivity to the graph data. For batched images, we can use parallel computing to accelerate the graph generation process and multiple subgraphs to combine as a large graph, where only connected nodes can perform message passing with the GNNs.

Regarding the model architecture, we utilize three layers of GIN to generate the embeddings of superpixels, upon which metric learning is performed. In addition to the transformed graph data derived from superpixels, we retain the original image data. We concatenate the features of superpixels and pixels and pass them through a lightweight CNN to smooth out small pixel groups in the output of the GNNs. This process helps enhance the segmentation accuracy by incorporating nondegraded image information. The overall architecture of our proposed model, named SMGNN, is illustrated in Figure 5.

To tackle the clinical image segmentation, we employ the Dice loss [67], which is a structure-aware and widely used loss function for medical image segmentation. This loss function is designed to measure the similarity between predicted and ground truth segmentation masks. We also use the Dice coefficient to evaluate the performance of different models [67], which a widely used metric on image segmentation. To be more specific, given a set G, we define its characteristic/label function by $\iota_G(i) = \begin{cases} 1, & i \in G \\ 0, & o.w. \end{cases}$. The Dice coefficient of two sets G and $\hat{G}$ is defined as

$$Dice(G, \hat{G}) = \frac{2 \sum_{i \in \Omega} \iota_G(i) \cdot \iota_{\hat{G}}(i)}{\sum_{i \in \Omega} \left(\iota_G(i) + \iota_{\hat{G}}(i)\right)}, \tag{3.1}$$

where Ω indicates the domain containing the two sets. The Dice metric is also directly used as a loss function to train a supervised segmentation task. The Dice loss function is formulated as

$$\mathcal{L}_{\text{Dice}}(G, \hat{G}) = -\frac{2 \sum_{i \in \Omega} \iota_G(i) \cdot \iota_{\hat{G}}(i)}{\sum_{i \in \Omega} \left(\iota_G(i) + \iota_{\hat{G}}(i)\right)}. \tag{3.2}$$

The joint loss function for both the superpixel metric learning and segmentation tasks is defined as a combination of the Dice loss and the superpixel metric loss. This joint loss function enables us to optimize the model parameters simultaneously for both tasks, effectively leveraging the benefits of both superpixel-based metric learning and pixel-wise segmentation. The joint loss function is formulated as

$$\mathcal{L}_{\text{Joint}} = \mathcal{L}_{\text{Dice}} + \lambda \mathcal{L}_{\text{SM}}, \tag{3.3}$$

where $\lambda \geq 0$ is a hyperparameter to trade off the learning on image space and graph space. Empirically, as shown on Figure 4, we set $\lambda = 0.1$ in the numerical experiments for better performance based on the extended experiments on the searching space $[0, 0.1, 0.5, 1, 10]$.

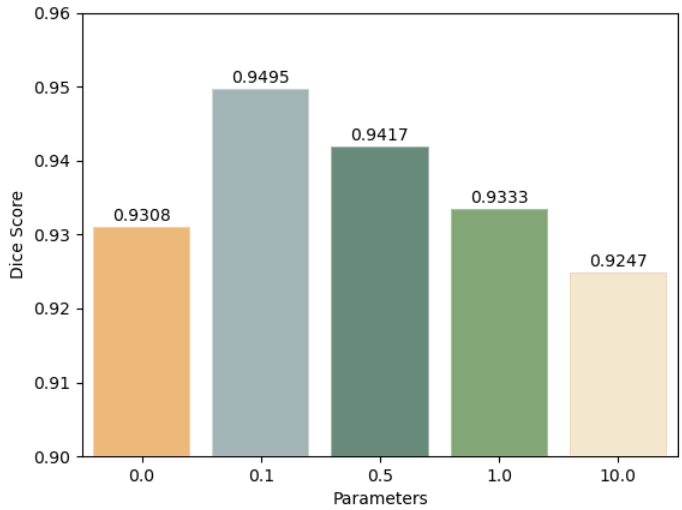


Figure 4. Comparison study of the parameter choice on $\lambda = 0.1$.

4. Numerical experiments

We employed a widely used WBCs dataset to evaluate the effectiveness of our proposed recognizing system. We employed the widely-acclaimed Adam optimization technique [68] for model training during the backpropagation phase. The implementations are programmed with PyTorch-Geometric (version 2.2.0) and PyTorch (version 1.12.1) and executed on an NVIDIA Tesla A100 GPU with 6, 912 CUDA cores and 80GB HBM2 installed on an HPC cluster.

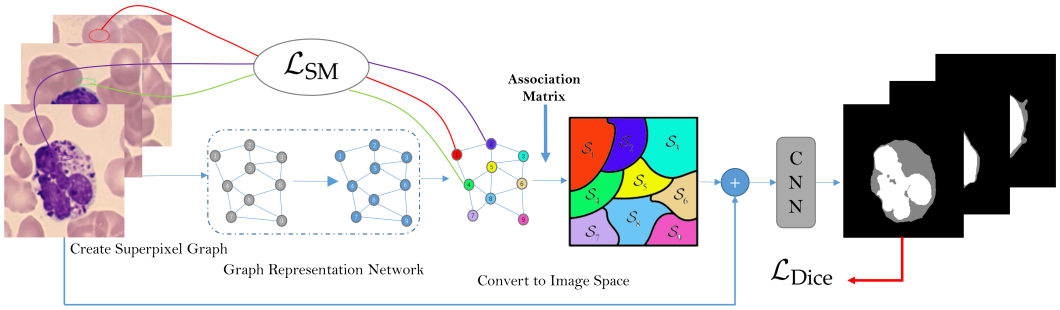


Figure 5. The framework of our proposed SMGNN for medical image segmentation consists of three main stages: 1) Create Superpixel Graph: The input images are initially over-segmented, generating multiple superpixels. Subsequently, a superpixel graph is constructed based on these segments. 2) Graph Representation Network: The superpixel embeddings are learned using a combination of GNN and metric learning techniques. This stage focuses on capturing the relationships and representations within the superpixel graph. 3) Convert to Image Space Feature: To facilitate segmentation, the superpixel graph is projected back to the image domain using the association matrix. The CNN layer is utilized to perform the segmentation task in the image space. The training process is supervised by superpixel metric loss function $\mathcal{L}_{SM}$ in the graph space and Dice loss function $\mathcal{L}_{Dice}$ in the image space.

4.1. Dataset description

We verify the robustness of our methods on two WBC image datasets. The first dataset (dataset-1) originates from Jiangxi Tecom Science corporation of China [69] , which contains 300 120×120 color images (176 neutrophils, 22 eosinophils, 1 basophils, 48 monocytes and 53 lymphocytes). Dataset-2 contains 100 300×300 color images (30 neutrophils, 12 eosinophils, 3 basophils, 18 monocytes and 37 lymphocytes). The second dataset (dataset-2) is publicly available on CellaVision blog * and widely used to conduct leukocyte research. These WBCs datasets leverage three-channel RGB images, which are processed via neural networks in an end-to-end training regimen. Each WBC image is manually-labeled, marking three primary regions: Nuclei (represented in white), cytoplasm (depicted in gray), and the surrounding peripheral blood (captured in black). The number of training/validation/testing data is 80/10/10% of the total numbers and the Dice loss function is applied to train the segmentation model. The dataset comprises five different cell types, staining effect, and illumination conditions which causes large variations in the sample distribution.

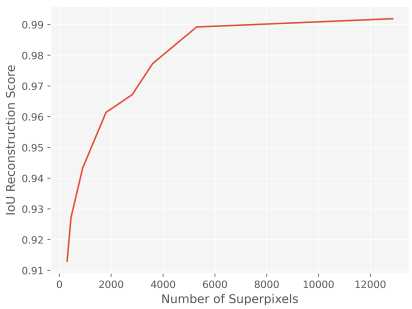
4.2. Evaluation of superpixel scale

To efficiently over-segment our input images, we adopted the simple non-iterative clustering (SNIC) superpixel generation methodology [66]. A distinctive feature of SNIC is its ability to visit each pixel just once—with the sole exception being those situated on superpixel boundaries. The computational traversal is characterized by the total number of pixels, N, augmented by a variable dictated by the desired superpixel count, K. Such a design renders SNIC more computationally nimble compared to alternatives like simple linear iterative clustering (SLIC) [47].

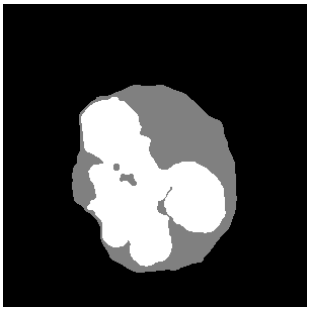
The superpixel quality, and its potential ramifications on segmentation, is an aspect we delve deeply

*http://blog.cellavision.com

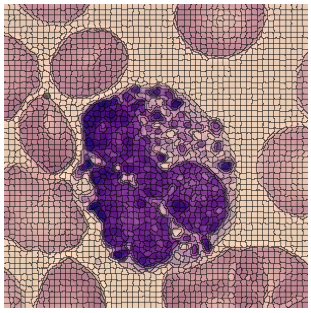
into. By modulating the number of superpixels, we could ascertain its influence. For instance, the WBCs dataset results (illustrated by the red trajectory in Figure 6) signify that as the granularity of the superpixel method amplifies, there's a corresponding upswing in the mIoU (mean Intersection over Union) score. Balancing optimal segmentation outcomes with computational practicality, we've pegged the mean scale of superpixels (S) at 16 for all ensuing experiments.



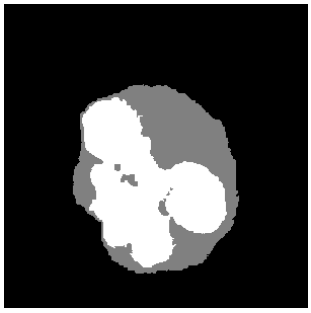
(a) IoU Reconstruction Score



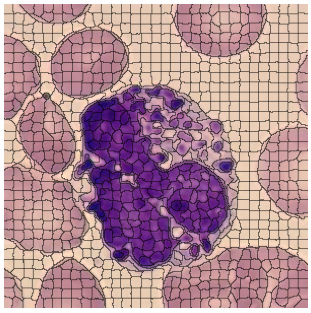
(b) Ground Truth Label



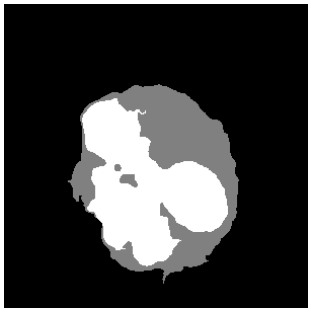
(c) Superpixels with Number 3000



(d) RLI with Number 3000

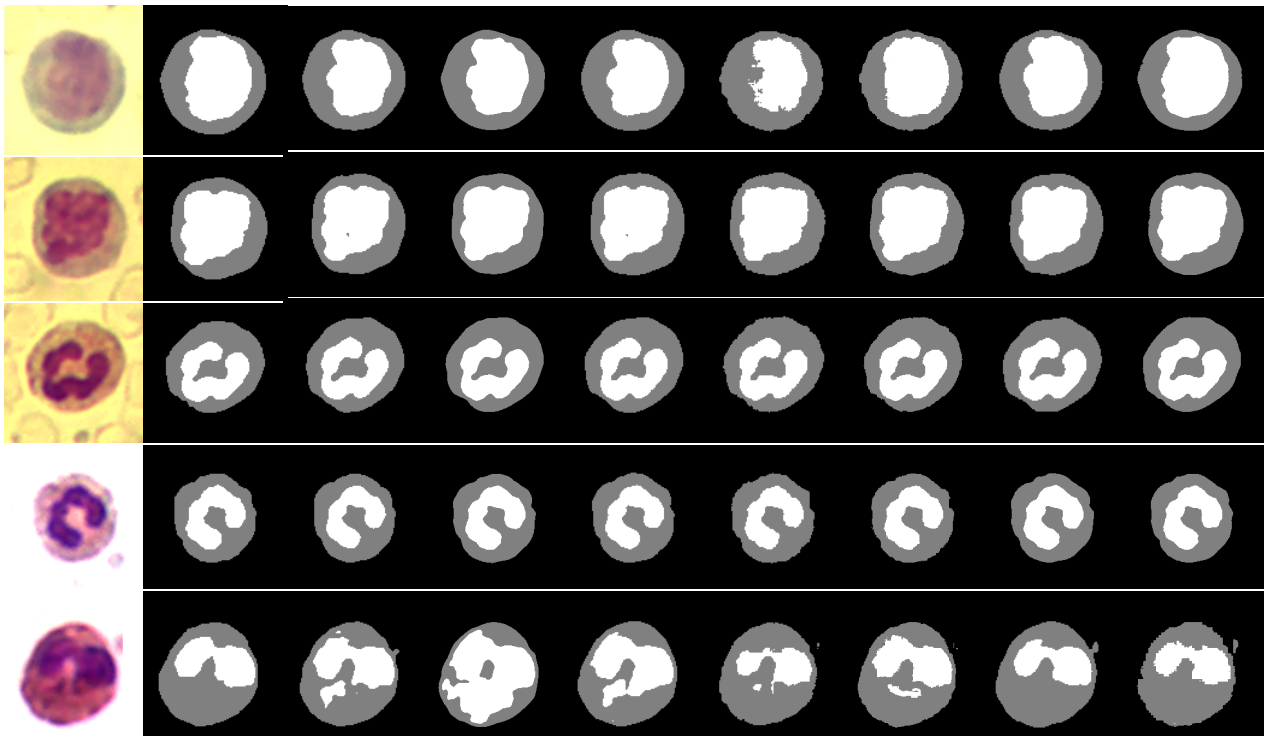


(e) Superpixels with Number 900



(f) RLI with Number 900

Figure 6. (a) The IoU reconstruction score versus the number of superpixels. (b) Ground truth label image. (c)(d)(e)(f) Superpixel images and reconstructed label images (RLIs) of two different superpixel numbers. With an increase in the number of superpixels, the RLI tends to converge toward the ground truth label image. This trend indicates that the boundaries of the superpixels become more consistent with the boundaries of the cells, leading to improved quality of the superpixel segmentation.



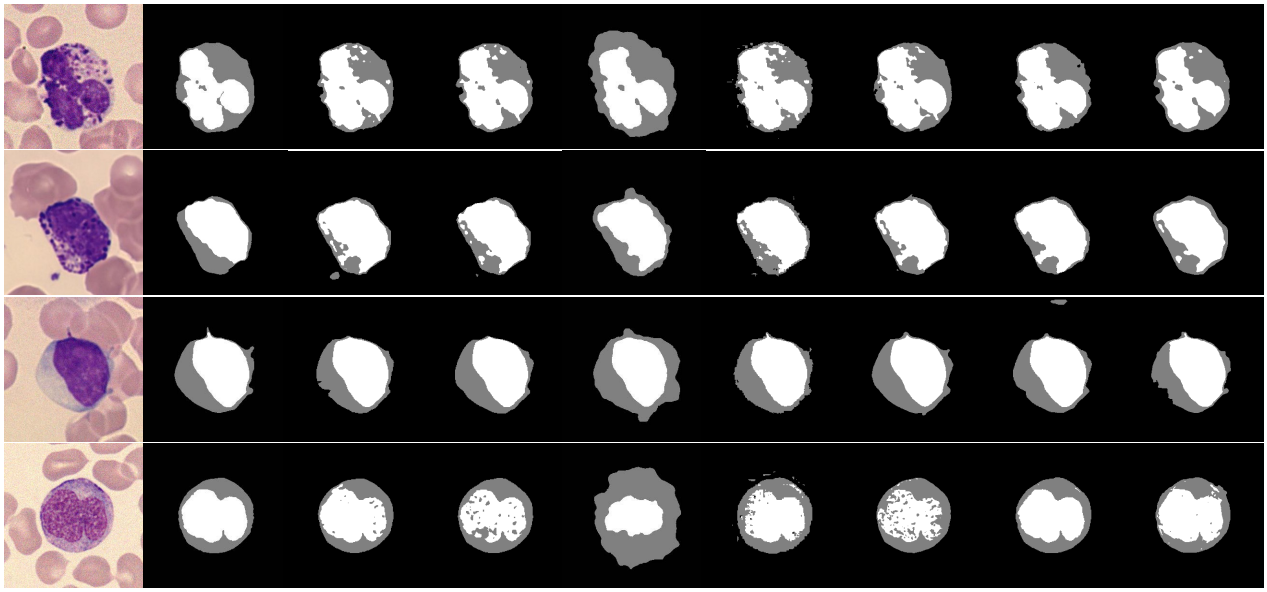
(a) Input (b) GT (c) UNet (d) Att U-Net (e) Runet (f) Swin-Unet (g) MIF-Net (h) VGU-Net (i) SMGNN

Figure 7. The segmentation results on dataset-1. Some CNN-based models result in over-segmentation issue. The proposed SMGNN model exhibits good segmentation performance.

4.3. Comparison with mainstream deep learning segmentation methods

A particularly intricate aspect of the WBCs dataset segmentation is the differentiation between the cytoplasm and the nuclei. Several images from dataset-2 present nuclei that are suboptimally stained, leading to a coloration reminiscent of the cytoplasm. Ideally, accurate segmentation demands that the representation of the nucleus be a cohesive, uninterrupted region. This color overlap often ensnares traditional CNN-based segmentation models, like U-Net, Attention U-Net and MIF-Net, resulting in predicted segmentations marred by interruptions or holes. The RU-Net demonstrates the ability to preserve piecewise constant nuclei regions; however, it tends to over-segment the cytoplasm region, which hinders overall segmentation performance. On the other hand, the Swin U-Net leverages the self-attention mechanism of the Transformer model and shows promising segmentation results. Nevertheless, the Swin U-Net's large number of parameters hampers training efficiency and requires significant computational resources. In comparison, the VGU-Net achieves a balance between efficiency and effectiveness, although it still utilizes a considerable number of parameters compared to the SMGNN. The proposed SMGNN model only utilizes about 7,000 parameters and takes an innovative approach by bolstering the connectivity of adjacent superpixels. Therefore, it is efficient in learnable parameters and proficient in preserving the integrity of the nucleus region, where the capability is vividly showcased in Figure 8. Compared to those end-to-end segmentation models, the SMGNN model takes a preprocessing step to cluster homogeneous pixels into superpixels and constructs graph structured data. We take the SNIC algorithm, which is non-iterative, requires less memory, is faster, and yet is a simpler superpixel

segmentation algorithm. This step might cost some computational time, but the SMGNN segmentation model is very efficient. We compare the computational time of different methods in Table 3.



(a) Input (b) GT (c) UNet (d) Att U-Net (e) Runet (f) Swin-Unet (g) MIF-Net (h) VGU-Net (i) SMGNN

Figure 8. The segmentation results on dataset-2. The ground truth annotation image contains a connected nuclei region without holes inside. Most CNN-based methods tend to over-segment the nuclei region induced by the model bias. The SMGNN model can well preserve local connectivity and achieve comparable performance.

Table 1. WBC segmentation results on dataset-1. SMGNN has the least number of trainable parameters (million) and achieves good performance in terms of the following metrics. Higher value means a better performance for metric with ↑ and vice versa.

Model	Parameters↓	Dice↑	Hausdorff ↓	PPV↑	Accuracy↑	Sensitivity↑
Unet	$5.43M$	0.9405	4.214	0.9468	0.9913	0.9325
RUnet	$31.03M$	0.9452	4.137	0.9528	0.9915	0.9377
Att-Unet	$34.88M$	0.9410	4.287	0.9413	0.9908	0.9301
Swin-Unet	$73.46M$	0.9501	4.056	0.9578	0.9918	0.9385
MIF-Net	$2.67M$	0.9523	4.035	0.9556	0.9911	0.9373
VGU-Net	$4.99M$	0.9604	3.987	0.9643	0.9923	0.9414
SMGNN (ours)	$7e^{-3}M$	0.9572	4.017	0.9545	0.9914	0.9401

In Figure 9, we show the Dice performance and number of learnable parameters of different deep learning models. Our SMGNN model can reach the SOTA performance while using far fewer parameters. In Tables 1 and 2, we show the quantitative comparison of these mainstream baseline segmentation models using the Dice coefficient, Hausdroff distance, positive predicted value (PPV), accuracy and sensitivity as the metric. The proposed SMGNN model can achieve SOTA segmentation performance while using remarkablely less parameters.

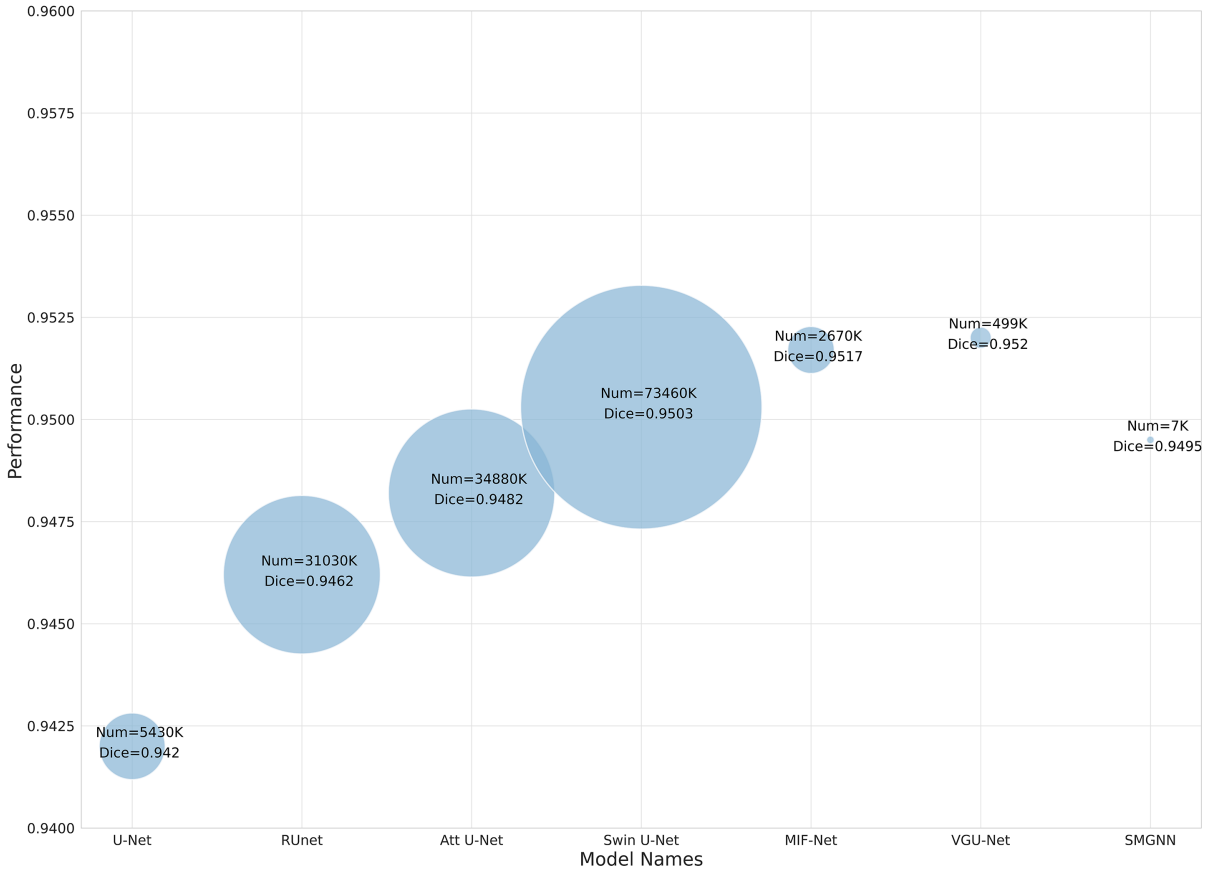


Figure 9. The comparison of model performance and network parameter size across different models on the WBCs dataset. The center of the circle indicates the Dice score of the model. The radius of the circle indicates the number of learnable parameters. The SMGNN model, utilizing approximately 7,000 parameters, achieves comparable performance to models with millions of parameters.

Table 2. WBC segmentation results on dataset-2. SMGNN has the least number of trainable parameters (million) and achieves good performance in terms of the following metrics. Higher value means a better performance for metric with ↑ and vice versa.

Model	Parameters↓	Dice↑	Hausdorff ↓	PPV↑	Accuracy↑	Sensitivity↑
Unet	$5.43M$	0.9420	4.5018	0.9419	0.9910	0.9204
RUnet	$31.03M$	0.9462	4.8409	0.9303	0.9876	0.9116
Att-Unet	$34.88M$	0.9482	4.7425	0.9333	0.9914	0.9307
Swin-Unet	$73.46M$	0.9503	4.7886	0.9438	0.9919	0.9232
MIF-Net	$2.67M$	0.9517	4.224	0.9428	0.9909	0.9311
VGU-Net	$4.99M$	0.9520	4.209	0.9493	0.9913	0.9378
SMGNN (ours)	$7e^{-3}M$	0.9495	4.4531	0.9431	0.9903	0.9373

Table 3. Comparison of the time cost. We sample eight WBC images from dataset-2 and count the inference time of different segmentation models.

Model	Time Cost
Unet	0.1094
RUnet	0.1145
Att-Unet	0.2681
Swin U-Net	0.7794
MIF-Net	0.0986
VGU-Net	0.1756
SMGNN (ours)	0.8173

4.4. Ablation study

To optimize node embeddings within our methodology, we selected the GIN model for the GNN module due to its superior discriminative capacity. Comparative tests with popular GNN models like GCN and graph attention network (GAT) revealed that a configuration using three layers of GIN demonstrated an enhanced performance, significantly improving node classification accuracy.

Beyond the conventional setup detailed in Figure 5, which incorporates pixel-level embedding, we ventured into an approach that solely leverages a GNN, transitioning from image-level segmentation components to a strictly node-based classification method, as illustrated in Figure 10. The training for this node classification is driven by a cross-entropy loss function, delineated as follows:

$$\mathcal{L}_{\mathrm{CE}}(\mathcal{S}, \hat{\mathcal{S}}) = -\frac{1}{K}\sum_{i=1}^{K}\sum_{c=1}^{C}\mathcal{S}_{ic}\log(\hat{\mathcal{S}}_{ic}), \tag{4.1}$$

where we use the majority voting rule to define the supepixel label as

$$\mathcal{S} = \lfloor \mathcal{Y} + 0.5 \rfloor, \tag{4.2}$$

guiding the supervised learning in graph space. In Eq (4.2), $\lfloor\cdot\rfloor$ is the round down function, $\hat{\mathcal{S}}$ is the predicted probability of superpixel, and C is the number of semantic classes. Though such a rule may group mistaken pixels whose pixel label is not the majority, the pure GNN methods may achieve good performance when the scale of the superpixel is small.

Our ablation studies, as depicted in Figure 11, offer keen insights into the performance nuances of pure GNN-based segmentation models. These models manifest commendable segmentation outcomes when oriented to small-scale superpixel environments. However, as we scale the superpixels, the model's performance is inversely impacted by its heightened sensitivity to superpixel quality, leading to notable performance drops. Introducing convolutional filters via CNN feature embedding for image-level segmentation does augment the model with additional parameters. Nevertheless, the significance of these filters is evident in the stability they confer upon the model, especially when navigating varying superpixel scales.

This investigation underpins a critical takeaway: The scale of superpixels and a model's sensitivity to their quality must be harmoniously calibrated. Relying exclusively on GNN-driven segmentation models may prove suboptimal when maneuvering larger superpixel frameworks.

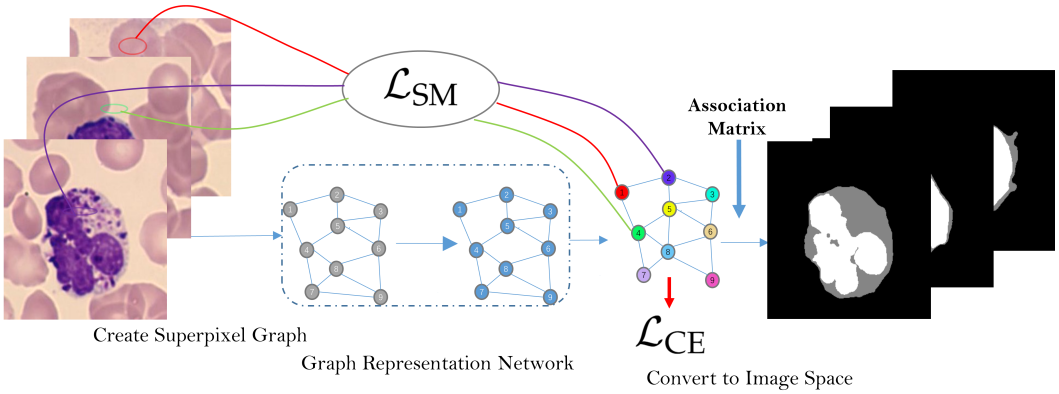


Figure 10. Pure GNN method converts segmentation task as the superpixel classification task, without involving learning in image space. The classification task is trained with cross entropy loss function $\mathcal{L}_{CE}$. The classified superpixels are projected back to the image domain through an association matrix.

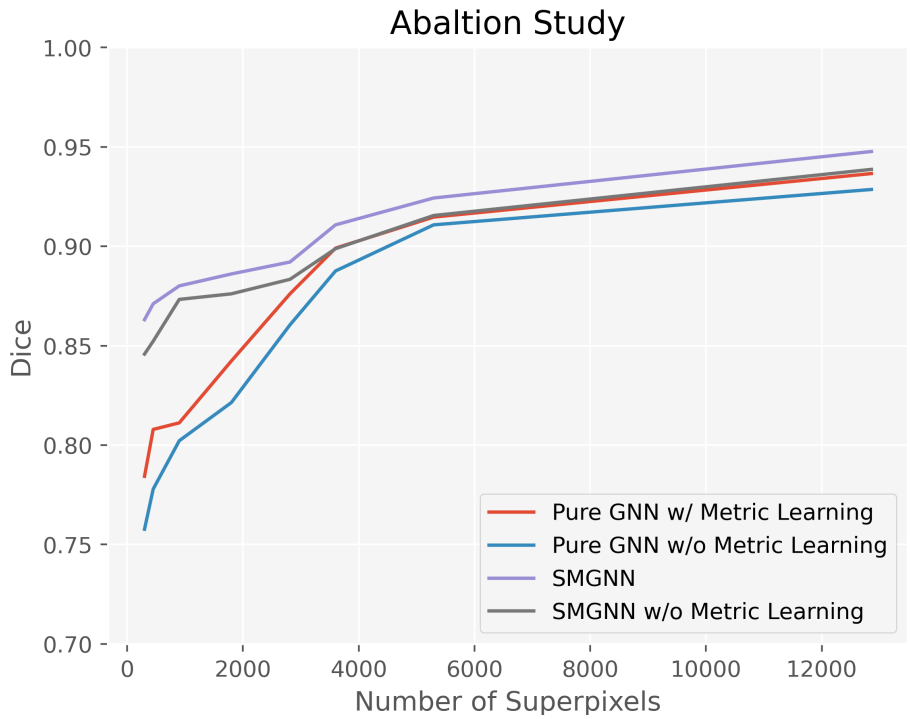


Figure 11. This ablation study delves into the impact of metric learning and convolutional filtering within the image domain. Segmentation trials were undertaken on WBCs datasets.

4.5. Effectiveness of metric learning on embedding space

To understand the impact of metric learning on the embedding space, we visually represent the spatial relationships of superpixels. We assign different colors to these superpixels based on their labels, as determined by Eq (4.2). Figure 12, created using uniform manifold approximation and projection (UMAP) [70], demonstrates the distances among embeddings derived from the GNN. A notable observation from this visualization is the pronounced separation between superpixels with distinct

labels—a testament to the efficacy of incorporating metric learning. Furthermore, there's a heightened cosine similarity between samples that are alike, while distinct samples exhibit reduced similarity. This distinction underscores the model's ability to effectively differentiate and group superpixels in the embedding space.

4.6. WBCs type classification results

Though the segmentation of cell salient regions, such as the nucleus and cytoplasm, is fundamental and challenging, there are various off-the-shelf methods available for subsequent cell type classification [71–73]. Segmentation may provide a direct means to obtain distinguishing characteristics for cell type classification, as cell morphology is closely related to cell type.

Algorithm 1 Training SMGNN Segmentation Model and ResNet Classification Model

Input: White blood cell image dataset $D = \{\mathbf{X}_i, \mathbf{Y}_i, \mathbf{C}_i\}_{i=1}^{n}$

Ensure: Optimal $\boldsymbol{\theta}_{GNN}$, $\boldsymbol{\theta}_{CNN}$ and $\boldsymbol{\theta}_{ResNet}$

//$\mathbf{X}, \mathbf{Y}, \mathbf{C}$ denotes the image, segmentation label and cell type label respectively;

1). Preprocessing:

$$\mathbf{X}, \mathbf{Y}, \mathbf{C} \leftarrow \text{ random mini-batch from } D$$

//Using SNIC algorithm [66] to generate superrpixels S and association matrix Q;

$$\mathcal{S}, \mathbf{Q} = SNIC(X)$$

//Using Eq (1) of the manuscript as the generation of superpixel metric labels (GSML);

$$\mathcal{Y} = GSML(\mathbf{Y}, \mathbf{Q})$$

//Construction of Superpixel Graph (CSG) to get the adjacency matrix of graph data with position relationship;

$$\mathcal{A} = CSG(\mathcal{S})$$

2). Training SMGNN Segmentation Model:

$$\boldsymbol{\theta}_{GNN}, \boldsymbol{\theta}_{CNN} \leftarrow \text{ initialize network parameters}$$

Repeat :

$$\mathcal{H} = GNN(\mathcal{S}, \mathcal{A})$$

//Convert embedding of graph space to image space with association matrix $\mathbf{Q}$;

$$h = \mathbf{Q}\mathcal{H}$$
$$h_{out} = CNN(h, X)$$
$$h_{label} = Softmax(h_{out})$$

//Compute the loss function according to Equation (10);

$$\mathcal{L}_{\text{Joint}} = \mathcal{L}_{\text{Dice}}(h_{label}, \mathbf{Y}) + \lambda \mathcal{L}_{\text{SM}}(H, \mathcal{Y})$$

// Update parameters according to gradients;

$$\boldsymbol{\theta}_{GNN} \overset{+}{\leftarrow} -\nabla\boldsymbol{\theta}_{GNN}\mathcal{L}_{\text{Joint}};$$
$$\boldsymbol{\theta}_{CNN} \overset{+}{\leftarrow} -\nabla\boldsymbol{\theta}_{CNN}\mathcal{L}_{\text{Joint}};$$

Until deadline

3). Training Lightweight ResNet:

$$\boldsymbol{\theta}_{ResNet} \leftarrow \text{ initialize network parameters}$$

Repeat :

// Get the segmentation result with trained SMGNN model;

$$h_{label} = SMGNN(\mathbf{X})$$
$$\hat{\mathbf{Y}} = argmaxh_{label}$$

// Training the ResNet Model;

$$\hat{\mathbf{C}} = ResNet(\hat{\mathbf{Y}}, \mathbf{X})$$

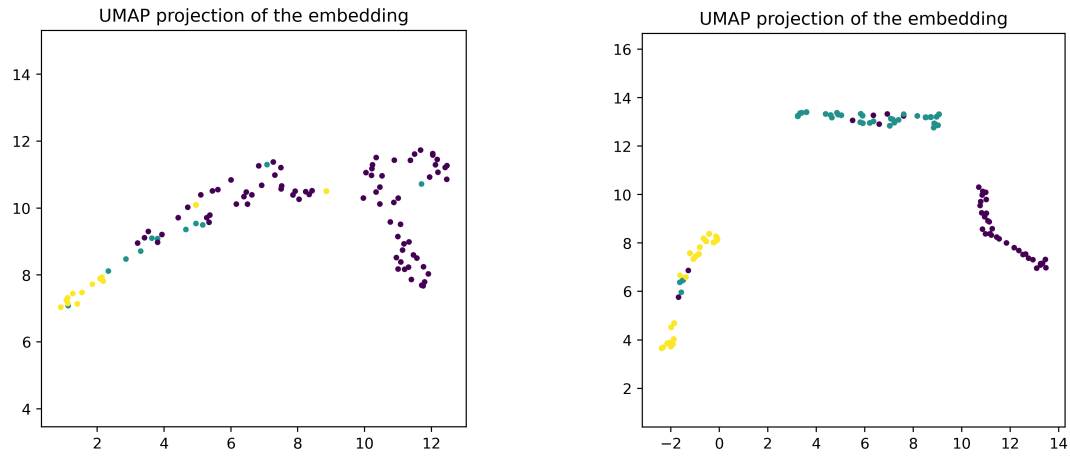
Compute cross-entropy loss;

$$\mathcal{L}_{\text{cross-entropy}} = \mathcal{L}_{\text{cross-entropy}}(\hat{\mathbf{C}}, \mathbf{C})$$

$$\boldsymbol{\theta}_{ResNet} \overset{+}{\leftarrow} -\nabla\boldsymbol{\theta}_{ResNet}\mathcal{L}_{\text{cross-entropy}};$$

Until deadline

In this part, we employ a lightweight ResNet neural network [74, 75] to train a classifier based on the outputs of segmentation networks, as shown in Figure 13. The overall recognition algorithm is shown in Algorithm 1. We sample about 1/3 training images of dataset-2 from each class and leave the remaining for testing. We train the segmentation and classification model separately. The classification result is shown in Table 4, and our classification method can achieve about 96.72% overall accuracy. In addition to our proposed segmentation-based cell type recognition system, we also implemented a baseline method to predict cell types without utilizing segmented cell regions. The corresponding results are presented in Table 5, and such methods without extraction on salient cell regions can barely achieve 72.13% overall accuracy. There are also traditional methods that employ handcrafted features extracted from segmented regions, combined with machine learning classifiers like support vector machine (SVM) [71], and such methods can get overall accuracy ranging from 89.69 to 96%. Our proposed deep learning-based automatic recognition system demonstrates high efficiency, and its accuracy can be further improved with an increase in the number of available training samples. To compare the overall accuracy of cell type classification task, we take the segmented regions of different models as input and compare the classification accuracy as shown in Table 6. Our proposed method achieves SOTA performance in the WBC recognition workflow.



(a) w/o ML (b) w/ ML

Figure 12. (a) Without superpixel metric learning, the embeddings are hardly able to separate. (b) With superpixel metric learning, the well-learned embeddings will form into three groups corresponding to nuclei, cytoplasm and background.

Table 4. Confusion matrix, accuracy, and overall accuracy with ResNet classification network using segmenteation results of SMGNN.

	Recognized Basophil	Recognized Eosinophil	Recognized Lymphocyte	Recognized Monocyte	Recognized Neutrophil	Accuracy
Basophil	2	0	0	0	0	100%
Eosinophil	0	8	0	0	0	100%
Lymphocyte	0	0	23	1	0	92.00%
Monocyte	0	0	0	10	0	100%
Neutrophil	1	0	0	0	16	94.12%
Overall Accuracy	-	-	-	-	-	96.72%

Table 5. Confusion matrix, acuracy, and overall accur acy with ResNet classification network without segmenteation methods.

	Recognized Basophil	Recognized Eosinophil	Recognized Lymphocyte	Recognized Monocyte	Recognized Neutrophil	Accuracy
Basophil	1	1	0	0	0	50%
Eosinophil	2	5	0	0	1	62.50%
Lymphocyte	0	1	20	2	1	83.33%
Monocyte	0	0	3	6	1	60%
Neutrophil	2	1	2	0	12	70.58%
Overall Accuracy	-	-	-	-	-	72.13%

Table 6. Segmentaion-based cell type classification experiments on dataset-2.

Model	Parameters↓	Dice↑	CA↑
Unet	$5.43M$	0.9420	0.9344
RUnet	$31.03M$	0.9462	0.9672
Att-Unet	$34.88M$	0.9482	0.9508
Swin-Unet	$73.46M$	0.9503	0.9508
MIF-Net	$2.67M$	0.9517	0.9508
VGU-Net	$4.99M$	0.9520	0.9672
SMGNN(ours)	$7e^{-3}M$	0.9495	0.9672

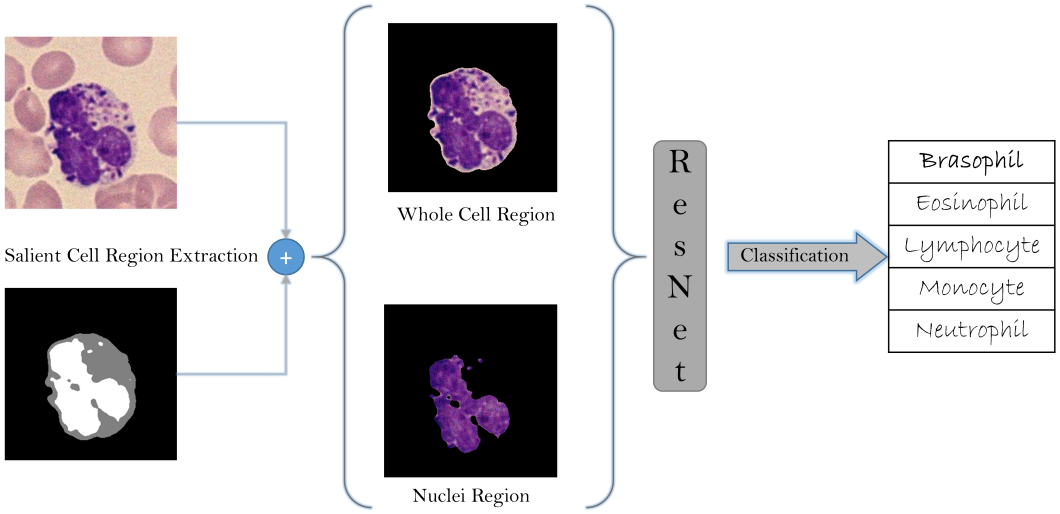


Figure 13. The segmentation-based cell type classification workflow. The segmented salient region implicitly provides important cell features, such as shape, perimeter, mean and variance of the nucleus boundaries. The lightweight ResNet extracts region-level embeddings to classify five cell types.

5. Conclusions

In this research paper, we proposed a deep learning based automatic recognizing system for the challenging WBC image recognizing task. In the first part, we proposed the SMGNN segmentation model, which combines superpixel methods and a lightweight GIN to significantly reduce memory usage while preserving segmentation capabilities. We innovatively proposed superpixel metric learning to capture cross-image global context information, making it highly suitable for medical images with limited training samples. Comparing our model to other mainstream deep learning models, we achieved comparable segmentation performance with a remarkable reduction of at most 10000 times fewer parameters. Through extended numerical experiments, we further investigated the effectiveness of metric learning and the quality of superpixels In the second part, the segmentation-based cell type classification processes exhibited satisfactory results, indicating that the overall automatic recognition algorithms are accurate and efficient for execution in hematological laboratories. We have made our code publicly available at https://github.com/jyh6681/SPXL-GNN, and we encourage its widespread implementation in portable devices of hematologists and remote rural areas.

Use of AI tools declaration

The authors declare they have not used Artificial Intelligence (AI) tools in the creation of this article.

Conflict of interest

The authors declare there is no conflict of interest.

References

1. H. Mohamed, R. Omar, N. Saeed, A. Essam, N. Ayman, T. Mohiy, et al., Automated detection of white blood cells cancer diseases, in *2018 First International Workshop on Deep and Representation Learning (IWDRL)*, (2018), 48–54. https://doi.org/10.1109/IWDRL.2018.8358214
2. M. S. Kruskall, T. H. Lee, S. F. Assmann, M. Laycock, L. A. Kalish, M. M. Lederman, et al., Survival of transfused donor white blood cells in hiv-infected recipients, *Blood J. Am. Soc. Hematol.*, **98** (2001), 272–279. https://doi.org/10.1182/blood.V98.2.272
3. F. Xing, L. Yang, Robust nucleus/cell detection and segmentation in digital pathology and microscopy images: a comprehensive review, *IEEE Rev. Biomed. Eng.*, **9** (2016), 234–263. https://doi.org/10.1109/RBME.2016.2515127
4. X. Zheng, Y. Wang, G. Wang, J. Liu, Fast and robust segmentation of white blood cell images by self-supervised learning, *Micron*, **107** (2018), 55–71. https://doi.org/10.1016/j.micron.2018.01.010
5. Z. Zhu, S. H. Wang, Y. D. Zhang, Rernet: A deep learning network for classifying blood cells, *Technol. Cancer Res. Treatment*, **22** (2023), 15330338231165856. https://doi.org/10.1177/15330338231165856
6. Z. Zhu, Z. Ren, S. Lu, S. Wang, Y. Zhang, Dlbcnet: A deep learning network for classifying blood cells, *Big Data Cognit. Comput.*, **7** (2023), 75. https://doi.org/10.3390/bdcc7020075
7. C. Cheuque, M. Querales, R. León, R. Salas, R. Torres, An efficient multi-level convolutional neural network approach for white blood cells classification, *Diagnostics*, **12** (2022), 248. https://doi.org/10.3390/diagnostics12020248
8. Y. Zhou, Y. Wu, Z. Wang, B. Wei, M. Lai, J. Shou, et al., Cyclic learning: Bridging image-level labels and nuclei instance segmentation, *IEEE Trans. Med. Imaging*, **42** (2023), 3104–3116. https://doi.org/10.1109/TMI.2023.3275609
9. Z. Gao, J. Shi, X. Zhang, Y. Li, H. Zhang, J. Wu, et al., Nuclei grading of clear cell renal cell carcinoma in histopathological image by composite high-resolution network, in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, (2021), 132–142. https://doi.org/10.1007/978-3-030-87237-3_13
10. X. Liu, L. Song, S. Liu, Y. Zhang, A review of deep-learning-based medical image segmentation methods, *Sustainability*, **13** (2021), 1224. https://doi.org/10.3390/su13031224
11. O. Ronneberger, P. Fischer, T. Brox, U-net: Convolutional networks for biomedical image segmentation, in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, (2015), 234–241. https://doi.org/10.1007/978-3-319-24574-4_28

12. F. Falck, C. Williams, D. Danks, G. Deligiannidis, C. Yau, C. Holmes, et al., A multi-resolution framework for U-Nets with applications to hierarchical VAEs, *Adv. Neural Inf. Process. Syst.*, **35** (2022), 15529–15544.

13. Z. Zhou, M. Rahman Siddiquee, N. Tajbakhsh, J. Liang, Unet++: A nested u-net architecture for medical image segmentation, in *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support*, (2018), 3–11. https://doi.org/10.1007/978-3-030-00889-5_1

14. H. Huang, L. Lin, R. Tong, H. Hu, Q. Zhang, Y. Iwamoto, et al., Unet 3+: A full-scale connected unet for medical image segmentation, in *IEEE International Conference on Acoustics, Speech and Signal Processing*, (2020), 1055–1059. https://doi.org/10.1109/icassp40776.2020.9053405

15. F. Jia, J. Liu, X. Tai, A regularized convolutional neural network for semantic image segmentation, *Anal. Appl.*, **19** (01), 147–165. https://doi.org/10.1142/s0219530519410148

16. N. Akram, S. Adnan, M. Asif, S. Imran, M. Yasir, R. Naqvi, et al., Exploiting the multiscale information fusion capabilities for aiding the leukemia diagnosis through white blood cells segmentation, *IEEE Access*, **10** (2022), 48747–48760. https://doi.org/10.1109/access.2022.3171916

17. A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, et al., An image is worth 16x16 words: Transformers for image recognition at scale, in *International Conference on Learning Representations*, 2021.

18. Z. Liu, Y. Lin, Y. Cao, H. Hu, Y. Wei, Z. Zhang, et al., Swin transformer: Hierarchical vision transformer using shifted windows, in *Proceedings of the IEEE International Conference on Computer Vision*, (2021), 10012–10022. https://doi.org/10.1109/iccv48922.2021.00986

19. C. Nicolas, M. Francisco, G. Synnaeve, N. Usunier, A. Kirillov, S. Zagoruyko, End-to-end object detection with transformers, in *European Conference on Computer Vision*, (2020), 213–229.

20. O. Petit, N. Thome, C. Rambour, L. Themyr, T. Collins, L. Soler, U-net transformer: Self and cross attention for medical image segmentation, in *International Workshop on Machine Learning in Medical Imaging*, (2021), 267–276. https://doi.org/10.1007/978-3-030-87589-3_28

21. J. Valanarasu, P. Oza, I. Hacihaliloglu, V. Patel, Medical transformer: Gated axial-attention for medical image segmentation, in *Medical Image Computing and Computer Assisted Intervention*, (2021), 36–46. https://doi.org/10.1007/978-3-030-87193-2_4

22. H. Cao, Y. Wang, J. Chen, D. Jiang, X. Zhang, Q. Tian, et al., Swin-unet: Unet-like pure transformer for medical image segmentation, in *Proceedings of European Conference on Computer Vision Workshops*, **3** (2023), 205–218. https://doi.org/10.1007/978-3-031-25066-8_9

23. Z. Chi, Z. Wang, M. Yang, D. Li, W. Du, Learning to capture the query distribution for few-shot learning, *IEEE Trans. Circuits Syst. Video Technol.*, **32** (2021), 4163–4173. https://doi.org/10.1109/tcsvt.2021.3125129

24. G. Li, S. Masuda, D. Yamaguchi, M. Nagai, The optimal gnn-pid control system using particle swarm optimization algorithm, *Int. J. Innovative Comput. Inf. Control*, **5** (2009), 3457–3469. https://doi.org/10.1109/GSIS.2009.5408225

25. Y. Wang, K. Yi, X. Liu, Y. Wang, S. Jin, Acmp: Allen-cahn message passing with attractive and repulsive forces for graph neural networks, in *International Conference on Learning Representations*, 2022.

26. S. Min, Z. Gao, J. Peng, L. Wang, K. Qin, B. Fang, Stgsn—a spatial–temporal graph neural network framework for time-evolving social networks, *Knowl. Based Syst.*, **214** (2021), 106746. https://doi.org/10.1016/j.knosys.2021.106746

27. B. Bumgardner, F. Tanvir, K. Saifuddin, E. Akbas, Drug-drug interaction prediction: a purely smiles based approach, in *IEEE International Conference on Big Data*, (2021), 5571–5579. https://doi.org/10.1109/bigdata52589.2021.9671766

28. J. Bruna, W. Zaremba, A. Szlam, Y. LeCun, Spectral networks and locally connected networks on graphs, in *International Conference on Learning Representations*, 2014.

29. M. Defferrard, X. Bresson, P. Vandergheynst, Convolutional neural networks on graphs with fast localized spectral filtering, *Adv. Neural Inf. Process. Syst.*, **29** (2016).

30. T. Kipf, M. Welling, Semi-supervised classification with graph convolutional networks, *International Conference on Learning Representations*, 2017.

31. P. Veličković, G. Cucurull, A. Casanova, A. Romero, P. Liò, Y. Bengio, Graph attention networks, in *International Conference on Learning Representations*, 2018.

32. K. Xu, W. Hu, J. Leskovec, S. Jegelka, How powerful are graph neural networks?, preprint, arXiv:1810.00826.

33. Y. Lu, Y. Chen, D. Zhao, J. Chen, Graph-FCN for image semantic segmentation, in *International Symposium on Neural Networks*, 2019.

34. L. Zhang, X. Li, A. Arnab, K. Yang, Y. Tong, P. Torr, et al., Dual graph convolutional network for semantic segmentation, in *British Machine Vision Conference*, 2019.

35. Y. Jiang, Q. Ding, Y. G. Wang, P. Liò, X.Zhang, Vision graph u-net: Geometric learning enhanced encoder for medical image segmentation and restoration, *Inverse Prob. Imaging*, **2023** (2023). https://doi.org/10.3934/ipi.2023049

36. Z. Tian, L. Liu, Z. Zhang, B. Fei, Superpixel-based segmentation for 3d prostate mr images, *IEEE Trans. Med. Imaging*, **35** (2015), 791–801. https://doi.org/10.1109/tmi.2015.2496296

37. F. Monti, D. Boscaini, J. Masci, E. Rodola, J. Svoboda, M. Bronstein, Geometric deep learning on graphs and manifolds using mixture model cnns, in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, (2017), 5115–5124. https://doi.org/10.1109/cvpr.2017.576

38. R. Gadde, V. Jampani, M. Kiefel, D. Kappler, P. Gehler, Superpixel convolutional networks using bilateral inceptions, in *Proceedings of the European Conference on Computer Vision*, (2016). https://doi.org/10.1007/978-3-319-46448-0_36

39. P. Avelar, A. Tavares, T. Silveira, C. Jung, L. Lamb, Superpixel image classification with graph attention networks, in *SIBGRAPI Conference on Graphics, Patterns and Images (SIBGRAPI)*, (2020), 203–209. https://doi.org/10.1109/sibgrapi51738.2020.00035

40. W. Zhao, L. Jiao, W. Ma, J. Zhao, J. Zhao, H. Liu, et al,. Superpixel-based multiple local cnn for panchromatic and multispectral image classification, *IEEE Trans. Geosci. Remote Sens.*, **55** (2017), 4141–4156. https://doi.org/10.1109/tgrs.2017.2689018

41. B. Cui, X. Xie, X. Ma, G. Ren, Y. Ma, Superpixel-based extended random walker for hyperspectral image classification, *IEEE Trans. Geosci. Remote Sens.*, **56** (2018), 3233–3243. https://doi.org/10.1109/tgrs.2018.2796069

42. S. Zhang, S. Li, W. Fu, L. Fang, Multiscale superpixel-based sparse representation for hyperspectral image classification, *Remote Sens.*, **9** (2017), 139. https://doi.org/10.3390/rs9020139

43. Q. Liu, L. Xiao, J. Yang, Z. Wei, Cnn-enhanced graph convolutional network with pixel-and superpixel-level feature fusion for hyperspectral image classification, *IEEE Trans. Geosci. Remote Sens.*, **59** (2020), 8657–8671. https://doi.org/10.1109/tgrs.2020.3037361

44. P. Felzenszwalb, D. Huttenlocher, Efficient graph-based image segmentation, *Int. J. Comput. Vision*, **59** (2010), 167–181. https://doi.org/10.1109/icip.2010.5653963

45. X. Ren, J. Malik, Learning a classification model for segmentation, *Proceedings of the IEEE International Conference on Computer Vision*, **2** (2003), 10–10. https://doi.org/10.1109/iccv.2003.1238308

46. M. Liu, O. Tuzel, S. Ramalingam, R. Chellappa, Entropy rate superpixel segmentation, in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, (2011), 2097–2104. https://doi.org/10.1109/cvpr.2011.5995323

47. R. Achanta, A. Shaji, K. Smith, A. Lucchi, P. FuaSü, S. Sstrunk, Slic superpixels compared to state-of-the-art superpixel methods, *IEEE Trans. Pattern Anal. Mach. Intell.*, **34** (11), 2274–2282. https://doi.org/10.1109/tpami.2012.120

48. Z. Li, J. Chen, Superpixel segmentation using linear spectral clustering, in *Proceedings of the IEEE conference on computer vision and pattern recognition*, (2015), 1356–1363. https://doi.org/10.1109/cvpr.2015.7298741

49. Y. Liu, C. Yu, M. Yu, Y. He, Manifold slic: a fast method to compute content-sensitive superpixels, in *Proceedings of the IEEE conference on Computer Vision and Pattern Recognition*, (2016), 651–659. https://doi.org/10.1109/cvpr.2016.77

50. R. Achanta, S. Susstrunk, Superpixels and polygons using simple non-iterative clustering, in *Proceedings of the IEEE conference on computer vision and pattern recognition*, (2017), 4651–4660. https://doi.org/10.1109/cvpr.2017.520

51. W. Tu, M. Liu, V. Jampani, D. Sun, S. Chien, M. Yang, et al., Learning superpixels with segmentation-aware affinity loss, in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, (2018), 568–576. https://doi.org/10.1109/cvpr.2018.00066

52. V. Jampani, D. Sun, M. Liu, M. Yang, J. Kautz, Superpixel sampling networks, in *Proceedings of the European Conference on Computer Vision*, (2018), 352–368. https://doi.org/10.1007/978-3-030-01234-2_22

53. F. Yang, Q. Sun, H. Jin, Z. Zhou, Superpixel segmentation with fully convolutional networks, in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, (2020), 13964–13973. https://doi.org/10.1109/cvpr42600.2020.01398

54. T. Suzuki, Superpixel segmentation via convolutional neural networks with regularized information maximization, in *IEEE International Conference on Acoustics, Speech and Signal Processing*, (2020), 2573–2577. https://doi.org/10.1109/icassp40776.2020.9054140
55. L. Zhu, Q. She, B. Zhang, Y. Lu, Z. Lu, D. Li, J. Hu, Learning the superpixel in a non-iterative and lifelong manner, in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, (2021), 1225–1234. https://doi.org/10.1109/cvpr46437.2021.00128
56. C. Saueressig, A. Berkley, R. Munbodh, R. Singh, A joint graph and image convolution network for automatic brain tumor segmentation, in *International Conference on Medical Image Computing and Computer-Assisted Intervention Workshop*, (2021), 356–365. https://doi.org/10.1007/978-3-031-08999-2_30
57. V. Kulikov, V. Lempitsky, Instance segmentation of biological images using harmonic embeddings, in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, (2020), 3843–3851. https://doi.org/10.1109/cvpr42600.2020.00390
58. J. Kim, T. Kim, S. Kim, C. Yoo, Edge-labeling graph neural network for few-shot learning, in *Proceedings of the IEEE conference on computer vision and pattern recognition*, (2019), 11–20. https://doi.org/10.1109/cvpr.2019.00010
59. T. Chen, S. Kornblith, M. Norouzi, G. Hinton, A simple framework for contrastive learning of visual representations, in *International conference on machine learning*, (2020), 1597–1607.
60. X. Chen, H. Fan, R. Girshick, K. He, Improved baselines with momentum contrastive learning, preprint, arXiv:2003.04297.
61. K. He, H. Fan, Y. Wu, S. Xie, R. Girshick, Momentum contrast for unsupervised visual representation learning, in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, (2020), 9729–9738. https://doi.org/10.1109/cvpr42600.2020.00975
62. M. Caron, I. Misra, J. Mairal, P. Goyal, P. Bojanowski, A. Joulin, Unsupervised learning of visual features by contrasting cluster assignments, *Adv. Neural Inf. Process. Syst.*, **33** (2020), 9912–9924.
63. W. Wang, T. Zhou, F. Yu, J. Dai, E. KonukogluVan, L. Gool, Exploring cross-image pixel contrast for semantic segmentation, in *Proceedings of the IEEE International Conference on Computer Vision*, (2021), 7303–7313. https://doi.org/10.1109/iccv48922.2021.00721
64. J. Gilmer, S. Schoenholz, P. Riley, O. Vinyals, G. Dahl, Neural message passing for quantum chemistry, in *International Conference on Machine Learning*, 2017.
65. B. Weisfeiler, A. Leman, A reduction of a graph to a canonical form and an algebra arising during this reduction, *Nauchno Tech. Inf.*, **2** (1968), 12–16.
66. R. Achanta, S. Susstrunk, Superpixels and polygons using simple non-iterative clustering, in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2017. https://doi.org/10.1109/cvpr.2017.520
67. F. Milletari, N. Navab, S. Ahmadi, V-net: Fully convolutional neural networks for volumetric medical image segmentation, in *International Conference on 3D Vision*, (2016), 565–571. https://doi.org/10.1109/3dv.2016.79
68. D. Kingma, J. Ba, Adam: A method for stochastic optimization, in *International Conference on Learning Representations*, 2015.

69. X. Zheng, Y. Wang, G. Wang, Z. Chen, A novel algorithm based on visual saliency attention for localization and segmentation in rapidly-stained leukocyte images, *Micron*, **56** (2014), 17–28. https://doi.org/10.1016/j.micron.2013.09.006

70. L. McInnes, J. Healy, J. Melville, Umap: Uniform manifold approximation and projection for dimension reduction, preprint, arXiv:1802.03426.

71. A. Acevedo, A. Merino, S. Alférez, A. Molina, L. Boldú, J. Rodellar, A dataset of microscopic peripheral blood cell images for development of automatic recognition systems, *Data Brief*, **30** (2020), 105474. https://doi.org/10.1016/j.dib.2020.105474

72. P. Yampri, C. Pintavirooj, S. Daochai, S. Teartulakarn, White blood cell classification based on the combination of eigen cell and parametric feature detection, in *IEEE Conference on Industrial Electronics and Applications*, (2006), 1–4. https://doi.org/10.1109/iciea.2006.257341

73. I. Livieris, E. Pintelas, A. Kanavos, P. Pintelas, Identification of blood cell subtypes from images using an improved ssl algorithm, *Biomed. J. Sci. Tech. Res.*, **9** (2018), 6923–6929. https://doi.org/10.26717/bjstr.2018.09.001755

74. R. Banerjee, A. Ghose, A light-weight deep residual network for classification of abnormal heart rhythms on tiny devices, in *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*, (2022), 317–331. https://doi.org/10.1007/978-3-031-23633-4_22

75. K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition, in *Proceedings of the IEEE conference on Computer Vision and Pattern Recognition*, (2016), 770–778. https://doi.org/10.1109/cvpr.2016.90

AIMS Press