Relation-Aware Multi Channel Attention Based Graph Convolutional Network for Breast Cancer Image Classification

Yaqi Chen 30920211154125,[1] Xinjian Chen 30920211154150,[1] Hefu Xie 30920211154170,[1] Faquan Chen 30920211154144,[1] Rongcan Chen 30920211154147 [1]

[1] School of informatics, Xiamen University, Xiamen

Abstract

Breast cancer is a life-threatening human disease. Timely diagnosis of breast cancer is related to disease prognosis, and correct classification of benign and malignant pathology images is of great importance for precision treatment. Although convolutional neural network (CNN) based deep learning can capture detailed local features in images, the weight sharing and local perception of the convolutional kernel itself limits the ability of CNN-based models to perceive global semantics. Not only that, the relevance and consistency of high-dimensional image semantics are lost because the correlation that exists in the high-dimensional image semantic space cannot be captured in the deep CNN architecture. To mitigate these constraints, we introduce a multi-channel attention mechanism to address the shortcomings of ignoring relevance and consistency in the high-dimensional semantic space based on the CNN model itself. Furthermore, we introduce the graph convolutional network (GCN) model which utilizes a pixel-based graph structure representation to solve the problem of global feature perception of images. Experiments demonstrate that our method achieves 82.8% accuracy on the Databiox dataset, improves the performance of classification tasks, and has wide generalizability and excellent application prospects.

Introduction

Globally, breast cancer is the second most commonly diagnosed cancer among women. In 2018, there were approximately 18.1 million new cancer cases worldwide, of which breast cancer constituted 2.089 million, accounting for 11.5% of new cancer cases; there were about 9.6 million cancer deaths, of which breast cancer constituted 627,000, accounting for 6.5% of deaths, with the mortality rate of breast cancer ranking first among female malignant tumors.

When evaluating histopathological images of breast cancer, experts will consider the morphology of histological structures to determine malignancy. Diagnosis is based on visual inspection of complex structures, considers a large number of dimensions, and is both time consuming and subjective. The development of deep learning methods for automated recognition of breast cancer can aid enhanced diagnosis of this disease and help reduce the subjectivity associated with clinical diagnostic procedures.

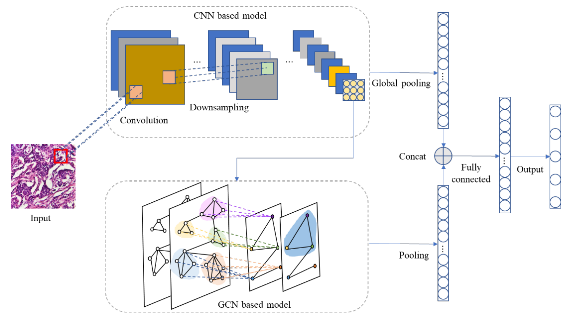

Figure 1: Flowchart of the proposed framework.

Classification based on manual feature extraction usually consists of image segmentation, feature extraction and selection, and classifier training. Texture based analysis has considered Grey-Level Co-occurrence Matrices (CLCM) (GadKari 2021) and Local Binary Patterns (LBP) (Spanhol et al. 2015). Spanhol et al. (Spanhol et al. 2015) proposed to discriminate between benign and malignant tumors using multiple texture features and classifiers. Spatial features are quantitative representations of the arrangement and distribution between cells. Doyle et al. (Doyle et al. 2008) proposed an SVM classification approach based on a combination of texture and spatial structure features in grayscale images of cell nuclei with an accuracy of 95.8% for the classification of breast and non-breast cancers. Manual feature extraction methods, despite achieving promising classification accuracy, are limited in that they cannot extract the deep semantic features and content relevance information of the image itself. Manual methods also have the disadvantage of low efficiency and high cost.

We propose a framework for breast cancer image classification combining a convolutional neural network and graph convolutional network (Figure 1). To address the problem that the deep CNN network model cannot reflect inter-channel correlation in the high-dimensional semantic space, we introduce a cross-channel attention mechanism to fuse the local feature representations in the high-dimensional semantic space for correlation. Specifically, for the weakly labeled data with complex histological morphology and similar differentiation process of breast cancer medical images, we perform data augmentation on the scarce data to reduce the disadvantages of the weakly labeled data. Second, by introducing a multi-channel attention mechanism, we address the disadvantage of ignoring correlation and consistency in the high-dimensional semantic space. Finally, to address the drawback that the CNN model cannot complete the global feature perception of the image, we introduce the GCN-based network model to convert the traditional image analysis method into a pixel-based graph structure representation analysis method. This approach not only introduces the topology of the image itself but also enhances the semantic information of each pixel point through global semantic fusion and propagation to better represent the original image.

Related Work

Deep learning using Convolution Neural Networks (CNNs) are widely used in pathological image analysis (Spanhol et al. 2015). CNNs learn features through hierarchical deep structure analysis and has a powerful ability to learn and represent abstract features (Doyle et al. 2008). CNNs have also been successfully applied to image classification. Araujo et al. (Araujo et al. 2017) apply CNN learning to breast biopsy classification, achieving two-class accuracy (carcinoma / non-carcinoma) of 83.3%. Vesal et al. (Vesal et al. 2018) proposed a method based on transfer learning for classification of breast histological images. Wang et al. (Wang et al. 2017) proposed a new histopathological image classification method based on bilinear CNN, and obtained experimental results superior to those of traditional approaches.

Graph Cut is an image segmentation algorithm (Boykov et al. 2001; Boykov 2001; Boykov and Veksler 2006; Boykov and Funka-Lea 2006) proposed by Boykov in 2001. Since there are many similarities between images and graphs, it is convenient to apply graph theory to image segmentation. Alternative to representing images and speech as a Euclidean structure data, with images regarded as 2-dimensional grid data (Araújo et al. 2017) and speech regarded as 1-dimensional data (Vesal et al. 2018), (Yan et al. 2018), graph-based methods can be applied to represent data as a non-Euclidean structure. However, for different nodes in the graph, the number of neighbor nodes of each node is not fixed, and it is difficult to define its neighbor nodes directly. At the same time, because the order of nodes in each graph structure may be different, it is difficult to define the Euclidean distance, so it is necessary to embed the graph data into the appropriate Euclidean vector space for measurement. Graph neural network can achieve the purpose of vectorization of graph data (Wang et al. 2017).

At present, most of the Graph neural network models have a general framework, called Graph Convolutional Networks (GCN). The goal of graph convolutional networks is to learn a mapping of features or signals on the graph $G = (V, E)$. Suppose there are n nodes on an input graph, and each node corresponds to d features, then these node features can form an $n \times d$ node feature matrix X, and the relationship between each node can be expressed by the adjacency matrix $A \in R^{(n \times n)}$. The node eigenmatrix X and the adjacency matrix A of the graph are taken as the input of the graph convolutional network, and then the node-level representation $Z \in R^{(n \times m)}$ is obtained, where m is the number of output features of each node (Kipf and Welling 2017).

In recent years, graph convolutional networks have been applied in the field of medicine. Parisot et al. (Parisot et al. 2017) applied frequence-domain graph convolutional network to the classification of autism spectrum disorder (ASD), modeling the subject group as a graph, where the node represents the subject and is associated with the imaging features of the subject, and the edge weight represents the similarity of the imaging and non- imaging features of the subject. Then the figure is fed into GCN for ASD classification learning. Anirudh et al. (Anirudh et al. 2019) used bootstrapping based on (Parisot et al. 2017) to enhance the ASD classification performance of GCN. Kazi et al. (Kazi et al. 2018) designed a parallel GCN framework to merge different non-imaging features of subjects and connect to a ranking layer to automatically learn the weight of correlation of different features for ASD classification.

Method

We propose a Relation-Aware Multi Channel Attention based Graph Convolutional Network (RMCG) for breast cancer image classification (Figure 2). The model consists of three main modules: multi-channel attention based on Resnet18; image topological structure construction module based on mutual information; and image features and spatial information fusion module, based on attention mechanism of stacked map convolutional network and map pooling network.

In general, our proposed method was initially based on the complex spatial relationships and texture structures of the data in Databiox dataset, with semantic features learned by using Resnet18. Since the multi-channel information has semantic relevance and aggregability, we introduce the multi-channel attention mechanism to enhance the high-level semantic information of the pixel nodes. The reason for adding the attention mechanism to the original stacked graph convolutional network structure is that the stacked graph convolutional network can fuse the features of the farthest pixel representation in the picture structure in the pixel-level topological relationship, but there is a differentiated association between the pixel content relationships at different distances, so this differentiation can be captured by the attention mechanism and brought into the whole picture content and spatial representation. So, this differentiation can be captured by the attention mechanism and brought into the content and spatial representation of the whole image.

Resnet18 Framework

The input of the network structure is the enhanced and normalized breast cancer data block. The network structure consists of a convolution block layer and four BasicBlocks. The size of convolution kernel is $7 * 7$, the stride size is 2, the padding is 3, and output channel is 64. From shallow to deep, each BasicBlock contains a convolution layer with a convolution kernel size of $3 * 3$, and a residual block of four $[64, 128, 256, 512]$ channel numbers. With the residual connections between different residual blocks and down-sampling operations, the model will learn the spatial and

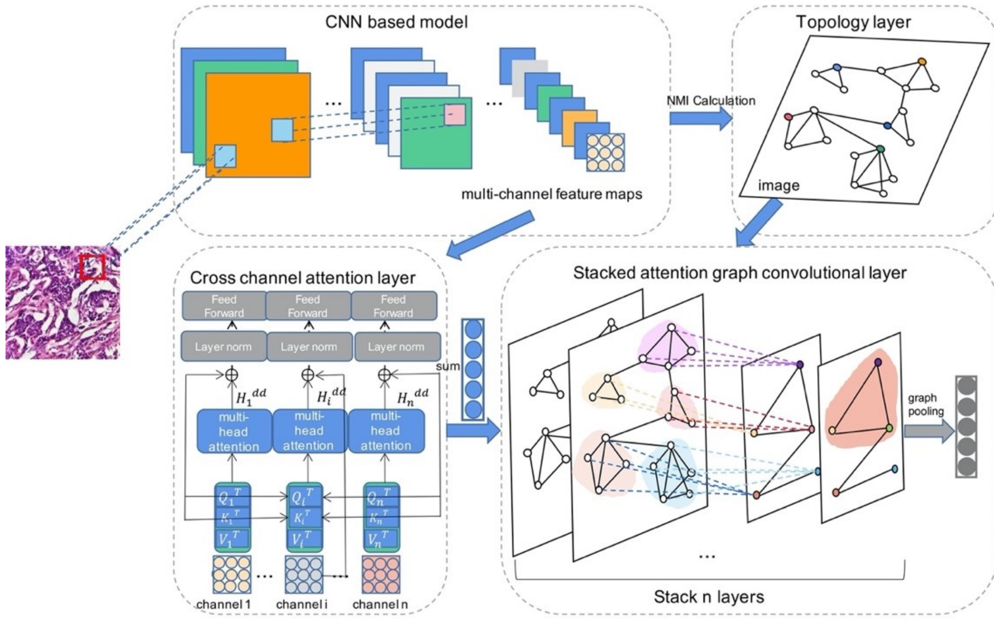

Figure 2: The overall structure of our proposed Relation-aware multi-channel attention based graph convolutional network.

content information of the image from shallow to deep, which can be shown as the deep semantic and spatial position relationship of pixel representation.Figure 3 illustrates how a building block works in the ResNet.

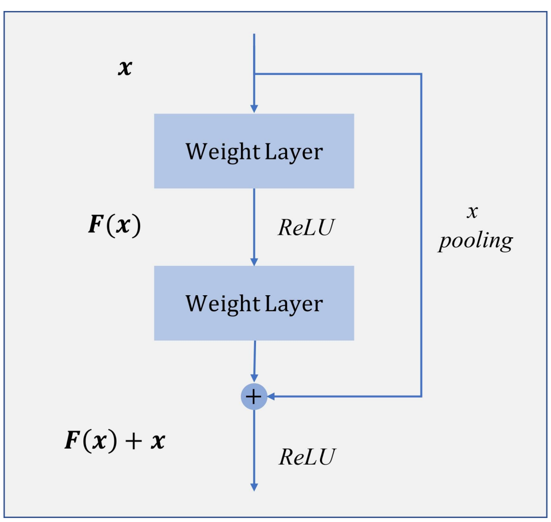

Figure 3: The basic block of ResNet.

Multi-channel Attention Mechanism

The convolutional operation of images can be regarded as the way of semantic fusion of images in each channel. Since we need to find out the high-level semantic association and information existing at the pixel level, we introduce the multi-channel attention mechanism to learn the semantic association and content features of pixels in different channels.

As shown in Figure 4, the multi-channel attention mechanism integrates the feature map $\{H_c^d\}_{(c=1)}^C$ of each channel and fuses the semantic information across channels. we propose to create Q_c^T,K_c^T,V_c^T as follows:

$$Q_c{}^T = \sigma\left(H_c{}^d W_c{}^q + 1\left(b_c{}^q\right)^T\right) \tag{1}$$

$$K_c^T = \sigma\left(H_{cc}{}^d W_c^k + 1\left(b_c{}^k\right)^T\right) \tag{2}$$

$$V_c^T = \sigma\left(H_{cc}{}^d W_c{}^v + 1\left(b_c{}^v\right)^T\right) \tag{3}$$

$$H_{cc}^d = \sum_{j,j\neq c} K_j \odot H_j^d \tag{4}$$

Where H_c^d is the input that generates cross-channel semantic fusion. It can be seen that both the keys and values are the outcome of the cross-channel query fusion. Note that, κ_j,$W_c^\star$,$b_c^\star$ all represent learnable parameters, $\odot$ represents element-wise multiplication.

After multi-channel fusion learning, the feature representation H^d can be represented as:

$$H_c{}^d = Softmax\left(\frac{Q_c{}^T\left(K_c{}^T\right)^T}{\sqrt{d_m}}\right)V_c{}^T \tag{5}$$

$$H^d = sum\left(H_c{}^d\right)_{c=1}^c \tag{6}$$

Where d_m represents the feature dimension of the image after the convolutional operation. We send the results of cross-channel content perception to the feedforward neural network to generate H_c^d, which is the same as the original structure of attention mechanism, and then we use the weighted summation method to obtain the pixel representation H_c^d with advanced semantic features.

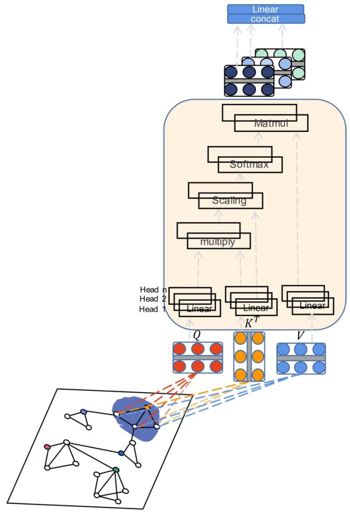

Figure 4: Multi-channel attention mechanism, which can find out the high-level semantic association and information existing at the pixel level.

The Objective Function

We introduce cross-entropy to learn the learnable parameters in our proposed model, the loss function $\mathcal{L}$ is defined as:

$$\mathcal{L} = -\sum_{i=1}^{\mathbf{K}} y_i \log(\mathbf{p}_i) + \frac{\lambda}{2}\|\mathbf{w}\|^2 \tag{7}$$

where K represents the number of categories in the data, y_i represents the category label of the sample, p_i represents the prediction probability of category i generated by the prediction layer, and w represents the learning parameters of the neural network in the final prediction layer. λ represents the super parameter of the regularization term, which is used to prevent the over-fitting phenomenon of the model.

Experimental Overview And Results

In this section, we validate the effectiveness and usability of our method using the Databiox breast cancer dataset.

Datasets

Our dataset consists of 922 medical images of 124 invasive ductal carcinomas (IDC) with different magnifications (4X, 10X, 20X, 40X) of breast tumor tissue. The tumors were classified into three classes according to the degree of differentiation. In this study, 40X image data were selected for the experiment. Due to the rare and special nature of medical image data, overfitting is more likely to occur in the training process of network structure than natural images. Since the pathological images are rotationally invariant under the premise of classification problem, the pathologist can diagnose the images by different angles and orientations without causing large errors in the diagnostic results, so we use translation and rotation (90, 180, 270 degrees) to enhance the datastrengthen the diverse representation of the data.

Baseline Methods

We compare our model with the following baseline methods:

- Alexnet (Krizhevsky et al. 2017) uses deep convolutional neural networks to encode the deep semantics of images to classify medical images.
- Resnet18 (He et al. 2016) constructs a deep convolutional network using residual linking, thus having better convergence speed and more efficient feature extraction.
- ECC (Simonovsky and Komodakis 2017) maps the convolution operation to the graph structure, and the weights in the convolution are constructed from the weights of the edges in the graph, instead of using the original weighted average convolution method in the image.
- Resnet-GCN (Kipf and Welling 2017) aggregates the convolution operation of 2D images with the pixel-level graph convolution operation to learn the texture and structure information of the whole image using the convolution method and then uses the graph convolution kernel to learn the correlation between individual pixels and fuse the pixel features to represent the spatial and content information of the whole image more comprehensively.
- Resnet-Graph Attention network (GAT) (Veličković et al. 2018) fuses the convolution operation of two-dimensional images with graph attention network, and since only the content information of nodes is considered in graph attention network without considering the topology, the graph attention network is used to illustrate the interconnection of spatial and content information of topology in images.
- Resnet-GCNII (Chen et al. 2020) extends the shallow graph convolutional network into the deep graph convolutional structure and solves the over smoothing problem of the deep graph convolutional structure utilizing residual linking so that the deep graph convolutional structure can be used to learn the high-level semantic information of pixel points.

Evaluation Metrics

We utilize the following evaluation metrics to evaluate our image classification model, specifically accuracy (Acc), precision, recall, and f1-score. Accuracy indicates the proportion of positive and negative classes that are correctly predicted by the current prediction model:

$$\mathrm{Acc} = \frac{\mathrm{TP} + \mathrm{TN}}{\mathrm{TP} + \mathrm{TN} + \mathrm{FP} + \mathrm{FN}} \tag{8}$$

where TP denotes True Positive, FP denotes False positives, TN denotes True Negatives, and FN denotes False Negatives. The precision rate denotes the weight of the number of correctly predicted positive classes in the prediction results of the model to the number of all predicted positive classes:

$$Precision = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FP}} \tag{9}$$

Recall can also be referred to as the percentage of correct predictions of positive classes to the total number of actual positive classes, i.e.

$$Recall = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FN}} \tag{10}$$

The f1-score is a weighted summation process of Precision and Recall, and the larger the value of f1-score (F1) is larger, the more accurate our model prediction is:

$$F_1 = \frac{2TP}{2TP + FP + FN} \tag{11}$$

Performance Comparison

Table I and Figure 5 summarize baseline results compared to our proposed methodology. It is apparent from Table I that our proposed framework outperforms the other six approaches across all metrics. We achieve classification accuracy of 82.87%, F1-score of 84.12%, Precision of 82.15% and Recall of 82.09%. Furthermore, our model improves by 1.5-2 percentage points across all metrics compared to the optimal Resnet18-GCNII model and by 2-3 percentage points compared to the normal graph convolutional network and graph attention network.

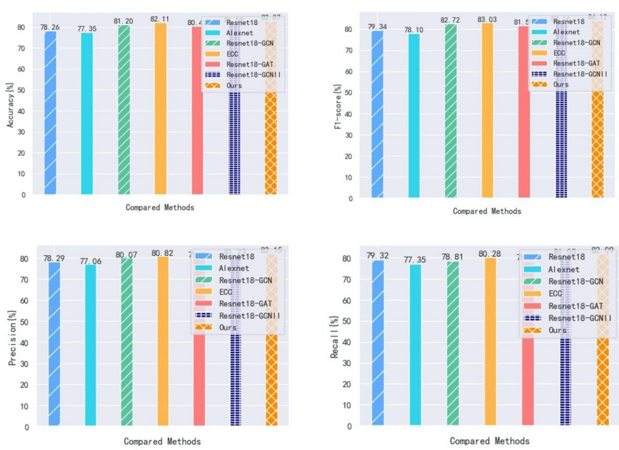

Figure 5: Comparison of experimental results of different models on the Databiox dataset. Our model improved by 4-5 percentage points in Acc, 4-6 percentage points in Precision, 3-5 percentage points in Recall, and 4-5 percentage points in F1-score compared to the traditional CNN-based model architecture. In addition, our model improved by 1.5-2 percentage points in four metrics compared to Resnet18-GCNII model and by 2-3 percentage points compared to the normal graph convolutional network and graph attention network.

Given the above results, we hold the following observations and conclusions. We build upon traditional CNN model architectures (Resnet18, Alexnet) by focusing on the association relationship between pixels in different channels. Traditional CNNs generate rich in semantic features, but the semantic association between different channels is not shown to be aggregated. Therefore, we introduce a cross-channel attention mechanism to fuse the semantic information between different channels and fuse the more effective channel features according to the attention score. This ensures that higher-order semantic and content information can be learned at each pixel location of the image.

Table 1: Results of Different Contrast Models.

Method	Acc.	FI-score	Precision	Recall
ResNet18	78.26	79.34	78.29	79.32
AlexNet	77.35	78.10	77.06	77.35
ResNet18-GCN	81.20	82.72	80.07	78.81
ECC	82.11	83.03	80.82	80.28
ResNet18-GAT	80.42	81.59	79.91	78.09
AlexNet-GCNII	81.59	82.16	81.23	81.05
Ours	82.87	84.12	82.15	82.09

In contrast to traditional network structures such as ECC, Resnet18-GCN, which rely on pixel points to model the topology of the image itself, we find that the image itself is a ”significant pixel” matter in terms of association. Therefore, the traditional network structure can learn its neighboring or second-order association features to a certain extent, but the ”importance” feature cannot be explicitly highlighted in the traditional network structure, which is the point that we use the attention-based stacked graph convolutional network structure to target. We hope to aggregate more effective semantic representations without causing over-smoothed semantic features and reduce the increase of noise.

We use the attention mechanism to amplify the content and location information of important pixel points by considering the sparse spatial structure of the image itself so that the important pixel points in the image can amplify their semantic representation and therefore increase the semantic representation of the image. However, as the stacked graph neural network becomes larger with the stacking of layers, we also find that the model has the disadvantage of feature over smoothing. In our future research work, we will consider how to better structure the attention mechanism and the deep layer convolutional network structure, to ensure that the features do not have the problem of over smoothing based on enhancing the fusion of important pixel features.

Conclusion

In this paper, we proposed a relation-aware multi-channel attention based graph convolutional network for breast cancer image classification. We use the convolutional network structure of stacked graph based on attention mechanism to aggregate more efficient semantic representation without causing the appearance of smooth semantic features and reduce the increase of noise. The experiments were conducted on the Databiox dataset. The accuracy and the Macro-F1 score of our RMCG model were 82.87% and 84.12%, respectively, which is better than other comparative approaches. The result can prove the effectiveness of our framework.The experimental results show that the attention-based deep convolutional network can help extract the fusion of spatial features and global features of pathological images through the image convolutional network, which has clinical research and application value.

References

D. Gadkari.*IMAGE QUALITY ANALYSIS USING GLCM*,May 2021.

V. Ojansivu and J. Heikkilä.*Blur Insensitive Texture Classification Using Local Phase Quantization*.vol. 5099. 2008.

F. Spanhol, L. Soares de Oliveira, C. Petitjean, and L. Heutte.*F. Spanhol, L. Soares de Oliveira, C. Petitjean, and L. Heutte*, IEEE Trans. Biomed. Eng.vol. 63, Nov. 2015, doi: 10.1109/TBME.2015.2496264.

S. Doyle, S. Agner, A. Madabhushi, M. Feldman, and J. Tomaszewski. *Automated Grading of Breast Cancer Histopathology Using Spectral Clustering with Textural and Architectural Image Features*, vol. 29. 2008.

T. Araújo et al., *Classification of breast cancer histology images using Convolutional Neural Networks,*. PLoS One, vol. 12, p. e0177544, Jun. 2017, doi: 10.1371/journal.pone.0177544.

S. Vesal, N. Ravikumar, A. Davari, S. Ellmann, and A. Maier, *Classification of Breast Cancer Histology Images Using Transfer Learning,*. 2018, pp. 812–819.

R. Yan et al.,*A Hybrid Convolutional and Recurrent Deep Neural Network for Breast Cancer Pathological Image Classification.* 2018.

C. Wang, J. Shi, Q. Zhang, and S. Ying, *Histopathological image classification with bilinear convolutional neural networks,* vol. 2017. 2017.

Y. Boykov, O. Veksler, and R. Zabih, *Fast approximate energy minimization via graph cuts,* IEEE Trans. Pattern Anal. Mach. Intell., vol. 23, no. 11, pp. 1222–1239, 2001, doi: 10.1109/34.969114.

Y. Boykov, *An Experimental Comparison of Min-Cut/Max-Flow Algorithms for Energy Minimization in Vision,*. IEEE Trans. Pattern Anal. Mach. Intell., vol. 26, Aug. 2001.

Y. Boykov and O. Veksler, *Graph Cuts in Vision and Graphics: Theories and Applications,*, Handb. Math. Model. Comput. Vis., Jan. 2006, doi: 10.1007/0-387-28831-7-5.

Y. Boykov and G. Funka-Lea, Graph Cuts and Efficient ND Image Segmentation,, Int. J. Comput. Vis. - IJCV, vol. 70, pp. 109–131, Nov. 2006, doi: 10.1007/s11263-006-7934-5.

T. N. Kipf and M. Welling, *Semi-Supervised Classification with Graph Convolutional Networks,*, International Conference on Learning Representations. 2017.

S. Parisot et al., *Spectral Graph Convolutions for Population-Based Disease Prediction.* 2017.

R. Anirudh and J. J. Thiagarajan,*Bootstrapping Graph Convolutional Neural Networks for Autism Spectrum Disorder Classification.*2019.

A. Kazi, S. Albarqouni, K. Kortüm, and N. Navab,*Multi Layered-Parallel Graph Convolutional Network (ML-PGCN) for Disease Prediction.* 2018

A. Krizhevsky, I. Sutskever, and G. E. Hinton, *ImageNet classification with deep convolutional neural networks,* Commun. ACM, vol. 60, no. 6, pp. 84–90, Jun. 2017, doi: 10.1145/3065386.

K. He, X. Zhang, S. Ren, and J. Sun, *Deep residual learning for image recognition,* 2016, doi: 10.1109/CVPR.2016.90.

M. Simonovsky and N. Komodakis, *Dynamic edge-conditioned filters in convolutional neural networks on graphs,* 2017, doi: 10.1109/CVPR.2017.11.

P. Veličković, A. Casanova,*Graph attention networks,* 2018.

M. Chen, Z. Wei, Z. Huang, B. Ding, and Y. Li, *Simple and deep graph convolutional networks,* 2020.