CONTEXT-AWARE GRAPH-BASED SELF-SUPERVISED LEARNING OF WHOLE SLIDE IMAGES

Milan Aryal and Nasim Yahya Soltani

Department of Computer Science, Marquette University, Milwaukee, USA

ABSTRACT

The gigapixel resolution of a single whole slide image (WSI), and the lack of huge annotated datasets needed for computational pathology, makes cancer diagnosis and grading with WSIs a challenging task. Moreover, downsampling of WSIs might result in loss of information critical for cancer diagnosis. Motivated by the fact that context such as topological structures in the tumor environment may contain critical information in cancer grading and diagnosis, a novel two-stage learning approach is proposed. Self-supervised learning is applied to improve training through unlabled data and graph convolutional network (GCN) is deployed to incorporate context from tumor and surrounding tissues. More specifically, we represent the whole slide as a graph, where nodes are patches from the WSIs. The patches in the graph are represented as feature vectors obtained from pre-training the patches in self-supervised learning. The graph is trained using GCN which accounts for the context of each tissue for the cancer grading and classification. In this work, WSIs for prostrate cancer are validated and the model performance is evaluated based on diagnosis and grading of prostrate cancer and compared with ResNet50 as a traditional convolutional neural network (CNN) and multi-instance learning (MIL) as a leading approach in WSI diagnosis.

***Index Terms*—** Self-supervised learning, whole slide image, graph convolutional networks, computational pathology.

1. INTRODUCTION

Pathologists rely on tissue slides mostly stained with hematoxylin and eosin(H&E) and other tissue stains for the diagnosis and prognosis of the cancer [1]. The Whole Slide Images (WSIs) are the high resolution images produced by digitizing the histology slides of the tissues. Obviously, a preliminary diagnosis through machine can help pathologists to be more efficient in diagnosis and may help to avoid some errors. Recently, the use of WSIs in computational pathology has significantly increased. However, large volume of data due to gigapixel resolution of WSIs and lack of large annotated dataset have been posed as challenges for automating diagnosis/prognosis of cancer using WSIs.

There have been successful applications of deep learning in the area of medical imaging [2], [3], [4], [5]. These achievements mostly through supervised-learning rely on lots of annotated data. In medical domain, the availability of the annotated dataset is very limited. However, unsupervised-learning approaches can be advocated to learn the representations from the unlabeled dataset. Self-supervised learning, [6] a form of unsupervised learning, can be used to learn the meaningful representation from the unlabeled data and then be transferred to the downstream task. Despite powerful performance of CNN in image classification and advanced methods in processing high-resolution images, they cannot be well trained for high-resolution WSIs. This is mainly because single WSI contains more than billion pixels in highest resolution and reducing the resolution in WSIs through downsampling WSI or capturing regions of interest may lose the necessary information in the neighborhood containing the tumor required for the cancer diagnosis. Weakly supervised learning based on multi-instance leaning (MIL) or tile-based patches have been recently used to handle computational complexity of training by WSIs [7], [8]. However, none of these methods are able to capture all the tumor neighborhood information. In MIL, instead of labeling each instance they are bagged together and then given a label. When using MIL approaches in WSI, a WSI is divided into smaller tiles/patches and bagged together [9]. Using this approach, only a fixed number of patches is trained. In [10], graph-based structure of WSI is presented where the graph is constructed based on patch selection with only regions of interest. It has been further shown that graph-based methods outperform MIL approaches.

Graph neural networks (GNN) can be used to model the tumor environment neighborhood information where the information of all the neighbors can be globally aggregated for cancer diagnosis.

In a nutshell, the challenges with training the WSIs are as follows:

- Large dimensionality of the WSIs.
- Insufficient annotated data.

In this paper, we overcome those challenges by proposing the GCN based self-supervised learning for WSIs. Self-supervised learning allows learning the meaningful representation presented by the data without the need of data labelling

or annotation. Presenting WSIs as a graph, the tumor environment and neighborhood information are used in training. The contribution of this work can be summarized as:

- Introduction of context-aware self-supervised learning on patches and graph-based learning on WSIs.
- Learning the features in patch levels and representation of any arbitrary size WSIs as a graph in full resolution.

2. METHOD

In this work, first the patches of the WSIs are pre-trained using self-supervised learning and then the whole graph structure is trained for cancer grade classification using GCN. In Fig. 1, the whole process of graph generation [11] for the WSI is presented. The patch generation, graph generation and learning from them is explained in the following sections.

2.1. Patch generation

The WSI in the datset are of multiscale resolution. The white background in WSIs is of no use in cancer diagnosis. First step in the patch generation is to remove the white background from WSIs. The white background can be removed from any resolution of the WSI. The segmentation network based on Unet [12] style encoder and decoder is used to separate the tissue from the background. The smallest resolution of WSI was rescaled to an image of size 512×512 to train the segmentation network. Given the output of the segmentation network, OpenCV is used to generate the contour separating tissue region from background in the WSI. Based on the contours the patches of size 256×256 are generated from the WSI at the highest resolution. From these patches the network can learn about the cancerous cells.

2.2. Learning from patches

2.2.1. Self-supervised learning

Self-supervised learning is a form of unsupervised learning that is used for pre-training known as pretext task. Then, these pretext tasks are fine-tuned for downstream task [13], [6]. Contrastive learning [14], is one of the most popular variants of the self-supervised learning. In this framework, positive samples stay close together and negative samples remain far apart.

We use self-supervised learning approach to learn from the patches. We apply augmentation randomly to patches to generate the query patch and key patch. The key and query can be seen as dictionary lookup. The positive pair in contrastive learning are when the query and key patches are from the same sample and different sample in negative pair. For

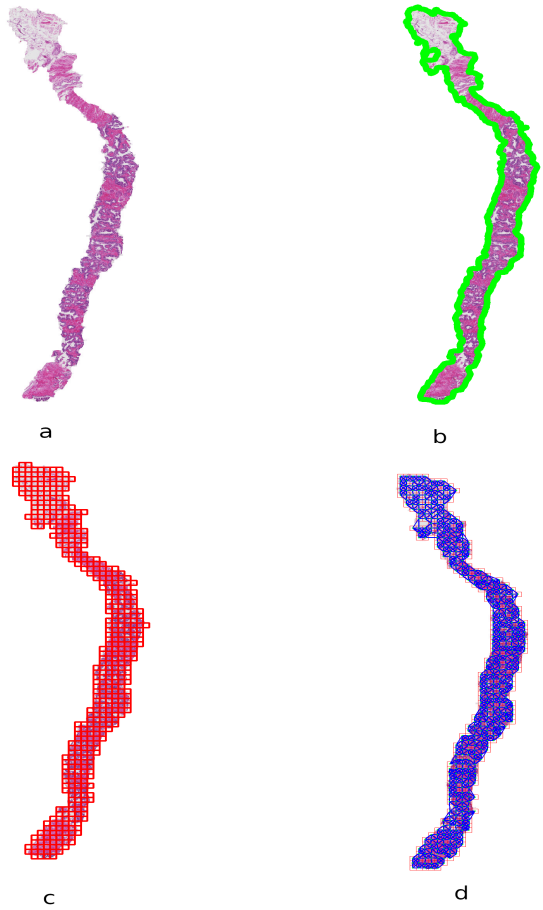

Fig. 1. a) WSI b) Separating tissue from the background. c) Generation of patches. d) Graph visualised in WSI.

training the self-supervised learning we use contrastive learning loss in the form of InfoNCE given by [15] :

$$L_q = -log \frac{exp(q \cdot k^+/\tau)}{exp(q \cdot k^+/\tau) + \sum_{k^-} exp(q \cdot k^-/\tau)} \quad (1)$$

where k^+ is the positive pair for the patch q and the k^- is the negative pair for the patch q, and τ is the temperature hyper-parameter.

The features extracted from these patches trained through contrastive learning are used as features of nodes in the GCN. This allows training the patches of WSIs for extracting useful features without the need of further annotation by the pathologists. Then, there is no need to find regions of interest for cancer grading or concatenate tile patches based on pixel intensity for feature extraction as commonly used approaches.

2.3. Graph generation

To train the WSIs using graph based learning, WSIs have to be first converted to graph based structure. A graph is a data

structure represented by tuple $G = (V, E)$ where V is the set of nodes and E is the set of edges representing connectivity between nodes. We discussed the construction of patch in section 2.1. Each patch in WSI is represented as a node in graph structure. The edges between the patches are formed using the fast approximate k-nearest neighbor (K-NN) [16].

The WSI is a multi resolution file. In this work, we have extracted the patches of size 256×256 at the highest available resolution. The dimension of the WSI is not the same for all the WSIs present in the dataset. Therefore, the number of patches extracted per WSI is different which results in different number of nodes in the graphs. Also, every node has a feature matrix, that is obtained by passing a patch through a pre-trained self-supervised model.

When WSI is structured as a graph with each patch as a node and the adjacency matrix and feature vectors for each node extracted from self-supervised trained model, any WSI with an arbitrary size can be represented in this form. The graph can then be used in GCN to learn meaningful representation from the data where local patch features can then be aggregated with the neighboring patch features using the graph-based training.

2.4. Learning on graphs

2.4.1. GCN

As discussed in previous section, the WSIs are represented as a graph. The diagnosis of cancer depends on the tissue and the context of the neighboring cells in the tissue. In graph neural network a node aggregates messages from its local neighborhood and is able to learn surrounding context. The input graph $G = (V, E)$ with node features $X \in \mathbb{R}^{d \times |V|}$, d is the dimension of node feature vector, learns from u's graph neighborhood $N(u), \forall u \in V$ through message passing. The message passing update over kth iteration is given by [17]

$$h_u^{(k)} = \sigma\left(W_{self}^{(k)} h_u^{(k-1)} + W_{neigh}^{(k)} \sum_{v \in N(u)} h_v^{(k-1)} + b^{(k)}\right) \quad (2)$$

where $W_{self}^{(k)}$ and $W_{neigh}^{(k)}$ are trainable parameters, σ accounts for non-linearity and $b^{(k)}$ is a bias term. The embeddings h_u are updated over the iterations and at $k = 0$, $h_u^{(0)} = X_u, \forall u \in V$. Equation (2) represents the graph in node level. It can also be presented in graph level as follows:

$$H^{(k)} = \sigma\left(AH^{(k-1)} W_{neigh}^{(k)} + H^{(k-1)} W_{self}^{(k)}\right) \quad (3)$$

where $H^{(k)} \in \mathbb{R}^{|V| \times d}$ is the matrix of node representations, A is the graph adjacency matrix. In this work, GCN [18] is used to learn the graph level representation from WSIs. The message passing in GCN can be expressed as following

$$h_u^{(k)} = \sigma\left(W^{(k)} \sum_{v \in N(u)} \frac{h_v}{\sqrt{|N(u)||N(v)|}}\right) \quad (4)$$

The GCN layer is implemented for 6 layers and from the penultimate node feature matrix H, global average pooling on all nodes followed by MLP head is applied. The term MLP stands for multi-layer perceptron. Cross-entropy loss function is used to grade the cancer.

3. IMPLEMENTATION DETAILS

3.1. Dataset

The dataset used in this paper has been collected from the Kaggle PANDA challenge [19]. The challenge provides the WSIs for grading the cancer in prostrate based on the Gleason score. The primary Gleason score ranges from 1-5 to most cancerous tissue and pathologists also assign secondary score for other surrounding tissue. So, the total Gleason score ranges from 2 to 10. Table 1 shows the Gleason scores and the International Society of Urological Pathology (ISUP) grades. The challenge consists of grading the WSIs into 6 ISUP grades. Fig. 2 shows primary and secondary Gleason scores and ISUP grades in different patches in the dataset. As shown in Fig. 2, the primary Gleason score increases as the glandular or white holes characteristics in the tissues are lost. The dataset consists of around 10500 WSIs. Each WSI has resolution in the scale of 1, 4, and 16. All the experiments were carried out at the highest resolution. For our experiment, 9500 WSIs were included in the train set and the rest were included in the validation set.

Table 1. Prostrate cancer grading

Gleason Score	ISUP Grade
6	1
7 (3+4)	2
7 (4+3)	3
8	4
9-10	5

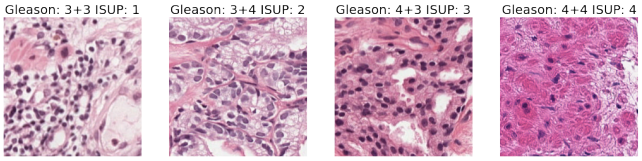

Fig. 2. Patches from WSI showing Gleason and ISUP score for prostrate cancer.

3.2. Training the self-supervised model

The self-supervised model was trained for 30 epochs. The learning rate was 3×10^{-3}, the weight decay was 1×10^{-6} with the Adam optimizer. The cosine scheduler was used as scheduler to adjust the learning rate. The patch size and the batch size were 256×256 and 256, respectively. The parameter τ was set to be 0.2. As a backbone ResNet50 was used for training the self-supervised network. Data augmentations applied during the training include random Gaussian Blur, random contrast adjustment, random horizontal and vertical flip.

3.3. Training the graph network

As regards the graph network, GCN was trained for 30 epochs using the Adam optimizer with the learning rate of 1×10^{-4} and weight decay of 10^{-6}. The cosine scheduler was used to adjust the learning rate. The batch size was chosen 1.

4. RESULTS AND DISCUSSION

We trained our model on PANDA dataset. Each of the WSIs were to be predicted into ISUP grade of 1-5 scale if the slide has cancer and grade of 0 for non cancerous slide. The performance of the model was evaluated based on the quadratic weighted kappa score [20].

Before deploying the proposed approach, ResNet50 [21] as a traditional CNN was used to train the model for the PANDA dataset. The WSIs were trained using concatenated tile pooling. From each WSI 36 tiles are selected, concatenated and then trained in the ResNet50. The validation dataset obtained a kappa score of 0.764 with this model. We further evaluated our dataset with MIL-based approach using Efficient Net [22]. The 36 patches were bagged together and then trained in the Efficient net. The kappa score improved to 0.79 with this model.

The proposed model was evaluated using 4-fold cross validation. We trained two GCN models with different feature sizes for nodes. The features for each node were obtained by passing the patch represented by node through pretrained self-supervised model. Then, there were 248 and 2048 features for nodes in the graph for each GCN model, respectively. The two GCN models were then ensembeled to predict the final grade of the cancer. Using the proposed model the kappa score of 0.899 was achieved. This is a great improvement compared to the simple tile-based approach. The kappa score for each model is summarized in Table 2.

Table 2. Kappa Score for different methods

Method	Kappa Score
ResNet50 [21]	0.764
MIL with Efficinet Net [23]	0.79
GCN with 248 Features for each Node	0.871
GCN with 2048 Features for each Node	0.891
GCN with 248 + GCN with 2048	0.899

5. CONCLUSION

In this work, we proposed a novel method for learning features from the patches in WSIs using self-supervised learning. The learned features for the patches from the self-supervised model were used as node features for the WSI graph. Then, this graph was trained using GCN to incorporate the context of each cell and its neighborhood for cancer diagnosis and grading. Our approach allows to learn the features from the patches without further annotation from the pathologists. In addition, the use of GCN enables the learning of WSIs in full resolution.

6. ACKNOWLEDGMENT

We would like to thank Raj High Performance Computing at Marquette University for providing the resources for training the model.

7. REFERENCES

[1] M. N. Gurcan, L. E. Boucheron, A. Can, A. Madabhushi, N. M. Rajpoot, and B. Yener, "Histopathological image analysis: A review," *IEEE Reviews in Biomedical Engineering*, vol. 2, pp. 147–171, 2009.

[2] H. Chen, L. Wu, Q. Dou, J. Qin, S. Li, J. Cheng, D. Ni, and P. Heng, "Ultrasound standard plane detection using a composite neural network framework," *IEEE Transactions on Cybernetics*, vol. 47, no. 6, pp. 1576–1586, 2017.

[3] X. Yang, L. Yu, L. Wu, Y. Wang, D. Ni, J. Qin, and P. Heng, "Fine-grained recurrent neural networks for automatic prostate segmentation in ultrasound images," *AAAI Conference on Artificial Intelligence*, p. 1633–1639, 2017.

[4] H. Chen, C. Shen, J. Qin, D. Ni, L. Shi, J. C. Y. Cheng, and P. Heng, "Automatic localization and identification of vertebrae in spine ct via a joint learning model with deep neural networks," *Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, pp. 515–522, 2015.

[5] D. C. Cireşan, A. Giusti, L. M. Gambardella, and J. Schmidhuber, "Mitosis detection in breast cancer histology images with deep neural networks," *Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, pp. 411–418, 2013.

[6] I. Misra and L. van der Maaten, "Self-supervised learning of pretext-invariant representations," *Computer Vision and Pattern Recognition (CVPR)*, pp. 6706–6716, June 2020.

[7] G. Ampanella, M.G. Hanna, Geneslaw, A. Miraflor, W. Krauss, V. Silva, K.J. Busam, E. Brogi, V.E. Reuter, D.S. Klimstra, and T.J. Fuchs, "Clinical-grade computational pathology using weakly supervised deep learning on whole slide images," *Nature Medicine*, 2019.

[8] M.Y. Lu, R. J. Chen, J. Wang, D. Dillon, and F. Mahmood, "Semi-supervised histology classification using deep multiple instance learning and contrastive predictive coding," *Advances in Neural Information Processing Systems (NeurIPS) Workshop in Machine Learning for Health*, 2019.

[9] M. Ilse, J. M. Tomczak, and M. Welling, "Attention-based deep multiple instance learning," *In Proc. of the 35th International Conference on Machine Learning (ICML)*, vol. 80, pp. 2127–2136, 2018.

[10] M. Adnan, S. Kalra, and H. R. Tizhoosh, "Representation learning of histopathology images using graph neural networks," *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2020.

[11] R. J. Chen, M.Y. Lu, M. Shaban, C. Chen, T.Y. Chen, D.FK. Williamson, and F. Mahmood, "Whole slide images are 2d point clouds: Context-aware survival prediction using patch-based graph convolutional networks," *Medical Image Computing and Computer Assisted Intervention (MICCAI)*, pp. 339–349, 2021.

[12] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," *Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, pp. 234–241, 2015.

[13] K. He, H. Fan, Y. Wu, S. Xie, and R. Girshick, "Momentum contrast for unsupervised visual representation learning," *Computer Vision and Pattern Recognition (CVPR)*, pp. 9726–9735, 2020.

[14] T. Chen, S. Kornblith, M. Norouzi, and G. Hinton, "A simple framework for contrastive learning of visual representations," *Proceedings of the 37th International Conference on Machine Learning*, vol. 119, pp. 1597–1607, 13–18 Jul 2020.

[15] A. Oord, Y. Li, and O. Vinyals, "Representation learning with contrastive predictive coding," *arXiv preprint*, vol. abs/1807.03748, 2018.

[16] M. Muja and D. G. Lowe, "Fast approximate nearest neighbors with automatic algorithm configuration," *International Conference on Computer Vision Theory and Applications*, pp. 331–340, 2009.

[17] W. L. Hamilton, "Graph representation learning," *Synthesis Lectures on Artificial Intelligence and Machine Learning*, vol. 14, no. 3, pp. 1–159.

[18] D. K. Duvenaud, D. Maclaurin, J. Iparraguirre, Rafael B., Timothy H., Alan A., and R. P. Adams, "Convolutional networks on graphs for learning molecular fingerprints," *Advances in Neural Information Processing Systems*, vol. 28, 2015.

[19] "Prostate cancer grade assessment (panda) challenge," https://www.kaggle.com/c/prostate-cancer-grade-assessment, 2021.

[20] J. Cohen, "Weighted kappa: Nominal scale agreement with provision for scaled disagreement or partial credit," *Psychological Bulletin 70 (4)*, pp. 213–220, 1968.

[21] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016.

[22] M. Tan and Q. Le, "EfficientNet: Rethinking model scaling for convolutional neural networks," *Proceedings of the 36th International Conference on Machine Learning*, pp. 6105–6114, 2019.

[23] X. Wang, Y. Yan, P. Tang, X. Bai, and W. Liu, "Revisiting multiple instance neural networks," *Pattern Recognition*, vol. 74, pp. 15–24, Feb 2018.