Refining Self-Supervised Feature Embeddings Through Graph Convolutional Networks for Improved Cancer Detection in WSIs

Wu, Ian
University of Southern California
Los Angeles, California
ianwu@usc.edu

Subramaniam Janakiraman, Sudharshan
University of Southern California
Los Angeles, California
ss20785@usc.edu

***Abstract*—Histopathology tissue analysis is currently considered the "gold standard" in cancer diagnosis, Whole slide imaging (WSI) offers a promising approach that enables digital analysis of entire histology slides. However, classical convolutional neural network (CNN) algorithms are ineffective for cancer detection in WSIs due to the extremely high resolution, imbalanced cancerous and non-cancerous regions in the images, and the lack of local annotations in many datasets. Multiple instance learning (MIL) has emerged as a promising approach for cancer detection, where predictions are generated by aggregating features for a bag of patches from the original image. Currently, these features generated from self-supervised learning on un-annotated patches. However, the feature embedding generated from SSL typically do not take into account the spatial relationships between the patches. This means that information about the neighboring or nearby patches is not incorporated into the feature embeddings, which limits their usefulness for certain types of analysis tasks. Our approach utilizes a graph convolutional network (GCN) applied to a nearest neighbor graph to model spatial and contextual relationships in feature embeddings. This approach aims to transform the input embeddings such that they lead to increased accuracy in the cancer detection in WSI's. Additionally, we transform feature embeddings via a simple weighted aggregation of randomly chosen patches as a simple alternate baseline which improves sensitivity. Our model is evaluated against two WSI datasets - Camelyon16 and TCGA. We show that our methods achieve an improvement of approximately 4% in accuracy and 10% in sensitivity for binary cancer detection, although performance remains similar for the task of distinguishing between cancer subtypes.**

***Index Terms*—Graph Neural Networks, Multiple Instance Learning, Self-Supervised Learning, Whole Slide Imaging, Camelyon16, TCGA, Accuracy, Sensitivity, AUC Score**

I. Introduction

Recent progress in deep learning systems has led to the widespread use of artificial intelligence (AI) algorithms for analyzing medical images. With its reduced latency and high accuracy, digital histopathology analysis has emerged as a promising technique for the diagnosis and prognosis of cancer.

In digital histopathology, whole slide images (WSIs) are digital scans of traditional glass slides, taken at gigapixel (extremely-high) resolution. WSI analysis is currently the best method available for cancer detection and prognosis [1], [3]. However, manual analysis of WSIs is extremely time consuming and requires the knowledge of an expert for accurate cancer detection. Also because of their extremely large size/resolution, WSIs cannot be directly processed by conventional convolutional neural networks (CNNs). In addition to the extremely high computational cost of training/predicting, traditional CNN architectures may also encounter learning bias, artifacts issues, and resolution learning obstacles.

In current literature, the issue of gigapixel resolution in WSIs is typically addressed using a multiple instance learning (MIL) approach, in which the entire WSI image is divided into a large number of smaller-sized patches which can be analyzed individually and in parallel. In works applying the MIL pipeline to gigapixel image analysis, the process follows as such:

1) Patch extraction from original images
2) Feature extraction from patches - represented as a "bag"
3) Aggregation of the feature bag to generate a final prediction

One issue with applying this approach to WSIs is that direct feature extraction from patches necessitates local annotation of patches in order to perform supervised learning and identify cancerous regions in WSIs. The process of generating these labels is a highly time-consuming and expensive task because subject matter experts must manually annotate the patches.

Due to this issue, typical WSI datasets only contain global-level labels. To deal with this, self-supervised learning (SSL) is a commonly used approach to handle the lack of local annotations by generating embeddings based on similarity measure. The embeddings generated by SSL will be used by the mil framework for classifying WSI's. In particular, self-supervised contrastive learning (SSCL), where a network is trained to extract similar features (embeddings) for similar patches, has been shown to be an extremely effective approach [7].

A major drawback of the MIL approach is that it treats the set of patches or embeddings as a bag, neglecting

the spatial relationship between instances. This loss of contextual and structural information for the patches can disrupt the critical relationships in the image that are necessary for cancer detection. In our work, we improve on the approach proposed by Li et. al [7] in Dual-stream Multiple Instance Learning Network (DSMIL) by introducing the use of GCNs for embedding refinement and improved WSI classification. Additionally, we introduce a weighted aggregation transformation as a baseline with improved sensitivity compared the state-of-the-art method.

Our approach encodes spatial information from MIL patch embeddings in a graph by taking advantage of the embedding generation strategy used in DSMIL. This graph is then processed using a graph convolutional network (GCN) [22] to refine the embeddings, which we show to boost the performance of an attention-based aggregator which can be used to generate a final prediction from the GCN output.

Additionally, Our Weighted Aggregation Transforms aims to enhance positive class information within the bags by utilizing Weighted Aggregation Transformation (WAT) to address the imbalance in the count of positive class embeddings. The embeddings are modified to preserve a portion of their original information while incorporating some additional information from other patch embeddings within the same bag leading to improved in sensitivity

The process of generating accurate embeddings also increases the computational cost of current methods. By framing the MIL bag as a graph, we hope to allow for weaker embeddings to be initially generated using SSCL, which can then be refined and used to generate a final classification through a graph convolutional network (GCN). The graph will be constructed from individual patch embeddings, with connections determined by a combination of spatial distance. This ability to use weaker embeddings also enables the use of smaller networks which can be trained on fewer samples, which is particularly valuable in a WSI setting where datasets are typically quite small.

We evaluate our approaches on both cancer detection and cancer subtype classification using the Camelyon 16 and TCGA Lung Cancer datasets using accuracy, sensitivity and AUC scores respectively. We show that our GCN approach improves results in cancer metastasis detection, with an improvement of approximately 4% in accuracy and 10% in sensitivity. However, we do not see the same improvement when evaluating on subtype classification, with scores remaining similar to the baseline.

A. Contributions

- We introduce a method of feature embedding refinement using Graph Convolutional Networks for improved WSI classification
- We introduce a simple weighted aggregation transform to improve model sensitivity
- We demonstrate improvement of 10% in sensitivity with our GCN approach over the current state-of-the-art work on cancer metastasis detection

II. Related Work

Our work further develops the approach to WSI classification used in DSMIL [7] by applying a GCN based approach to feature augmentation. MIL is the current state-of-the-art approach for gigapixel image classification [3], and the application of graph-based networks in histopathology has also become highly popular in recent works for their ability to preserve structure [13]. Graph machine learning has also previously been applied to classify MIL bags, showing strong results [11], [12], [14].

A. Background

1) Multiple Instance Learning for WSI classification: As previously mentioned, MIL is highly effective for computer vision tasks involving gigapixel images and is currently considered the state-of-the-art approach for processing WSIs. In early approaches to MIL, a simple aggregator, typically mean or max-pooling, was applied to the bag of instances. One of the primary weaknesses to this is that it does not consider the relationships between patches, as the aggregator considers the bag to be permutation invariant, meaning the instances in the bag are considered to be unordered [2]. However, recent works have used more complicated aggregators to consider the relationships between patches within the permutation invariant bag when generating a classification. Recent studies have developed more sophisticated aggregators and embedding generation techniques that consider the relationships between patches to improve classification accuracy. These include attention-based aggregation [6], [7], recurrent neural network (RNN)-based approaches [8], and using locally supervised learning for embedding generation [3]. Multi-stream approaches to bag aggregation have also been attempted, in which the outputs of multiple aggregators are used to strengthen the final classification results [7].

2) DSMIL: DSMIL was introduced in 2022 by Li et. al. [7] and improved on MIL for WSI classification in 2 primary ways. Here we provide an overview of the methods used in the paper.

Firstly, the paper introduces a multi-scale feature extraction method, where patches are generated at multiple levels of resolution for each WSI. In DSMIL, patches are generated specifically at 5x and 20x resolution, although their approach is extensible to more than 2 levels of resolution and different resolution levels. Features are then learned and extracted from each patch using SimCLR [18], a contrastive learning approach that trains a CNN to generate embeddings for

images such that similar images have similar embeddings. To train this model, images first transformed/augmented, then fed into 2 CNN feature extractor branches. By trying to maximize the "agreement" between the extracted features for differently augmented versions of the same image, SimCLR is able to extract features without them being explicitly defined. The CNN backbone for SimCLR used in the DSMIL paper is ResNet-18 [29].

The extracted features for the multi-scale patches are then combined using a "pyramidal concatenation strategy" introduced in the paper. Using this strategy, features for patches at the highest resolution are concatenated with the features of patches at lower resolutions which contain their patches as shown in figure 1. Because of this, features for lower resolution patches are concatenated to multiple patch features at higher resolution, causing the final embeddings in the MIL bag to become closer to each other within the embedding space. This captures some spatial aspects of each patch and is an effect which we capitalize upon in our GCN-based approach.

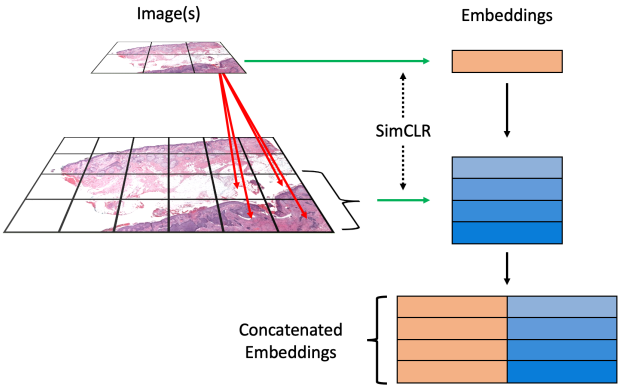

Fig. 1. Pyramidal concatenation of features in DSMIL

Secondly, DSMIL introduces a dual-stream, attention-based approach to patch aggregation. A more detailed description of the operations used in this block can be found in the original DSMIL paper [7], but the general concepts are presented here.

The first branch (stream) of the aggregator is a single linear classifier which is applied to each vector in the MIL bag to create a set of instance level predictions. Max pooling is then applied to this set of predictions to identify the "critical instance." This critical instance corresponds to the feature vector/patch which the model classifies as most likely to be of the positive class, and is used to compute attention weights in the second branch. The value of the instance level prediction for the critical instance is also averaged with the output of the second branch to generate the model's final prediction.

After the critical instance is identified, the second branch first applies a weight matrix to each embedding h_i to produce query and information vectors, denoted q_i and v_i respectively 1.

$$q_i = W_q h_i, \quad v_i = W_v h_i, \quad i = 0, ..., N-1 \tag{1}$$

A bag embedding is then calculated as a weighted sum of all information vectors, where the weight of each information vector v_i is calculated based on the (normalized) distance between the corresponding query vector q_i and the query vector q_m for the critical instance found in the first branch. The bag embedding is then multiplied by a weight vector for binary classification W_b to generate a bag level prediction. Finally, this bag level prediction is averaged with the instance level prediction of the critical instance from the first branch to generate the final prediction for the model.

This aggregator architecture essentially functions as a self-attention block for the MIL embedding bag. However, instead of matching each query vector with a set of key vectors for each instance, queries are matched with the query vector for the critical instance, so instances are only attended to based on their similarity to the critical instance. In our approach, we attempt to capitalize on inter-patch relationships to augment this attention block through refined embeddings generated by a GCN.

B. Graph Neural Networks for MIL in Histopathology

Graph-based techniques have become popular in WSI tasks recently, particularly in the context of MIL. Tu et. al. [11] were the first to apply a GNN approach to MIL, but did not focus on a large image or histopathology context. Their approach also utilizes a fully connected graph, which cannot scale to the bag size for MIL in the WSI context. In contrast, our approach utilizes a partially connected graph, such that it can handle the size of WSIs. However, the approach achieved state-of-the-art performance on multiple benchmark datasets, showing that this could potentially be used to improve results in gigapixel image/WSI classification.

Within medical machine learning literature, other approaches have utilized a variety of different GNN types to boost performance. In MS-RGCN, Bazargani et. al. [14] applied a specialized GCN to aggregate features extracted from multiple resolution levels and achieve high performance in predicting prostate cancer. However, their approach necessitates patch-level labeling, so the approach does not extend to the weakly supervised context. In another approach to graph-based bag aggregation, Adnan et. al. [12] achieved high performance by constructing a fully connected graph from MIL bag instances, then refining it using adjacency layer learning. The use of a fully connected graph in their work is enabled by the use of random sampling of patches to decrease the size of the graph. However, in

WSIs with small tumor regions, the sampling approach can lead to misclassification. Both of these methods used deep feature extractors, which we also utilize in our approach.

Several approaches using individual cells (as opposed to patch embeddings) as nodes in a graph have also been tested [13]. Among these, both Gadiya et. al. [15] and Wang. et. al. [16] use GCN's to process the resultant graphs, further motivating our approach. In MS-GWNN, Li and Zhang [10] used graph wavelet neural networks at multiple image resolutions to achieve state-of-the-art performance on breast cancer diagnosis. However, these approaches all required specialized WSI datasets, typically with specialized staining, to perform, reducing their practical value.

We focus on improving the accuracy of the DSMIL method while maintaining its ability to be trained on weakly labeled data such that it can be broadly applied in histopathology tasks. The abundance of graph-based approaches in recent literature shows importance of structural aspects of WSIs, which we capitalize on by using a GCN to refine initial patch embeddings and improve overall bag aggregation.

III. Dataset

We utilize pre-generated feature embeddings for the Camelyon 16 and The Cancer Genome Atlas (TCGA) Lung Cancer datasets created with the DSMIL pyramidal concatenation strategy and SimCLR features for evaluation of our methods. The WSIs from these datasets provide an accurate depiction of the problem and are suitable for evaluating the capabilities of algorithms utilized in the detection of cancer in WSIs.

The Camelyon 16 dataset [19] is a public grand challenge dataset aimed at detecting breast cancer metastases through the use of machine learning classification algorithms in whole-slide images. The dataset consists of 399 WSIs, with 239 and 160 normal and tumor images, respectively. As previously mentioned, DSMIL samples these images at 20x and 5x resolution, yielding approximately 3.2 million feature vectors in total, with an average of 8,000 patches/vectors per slide. However, there is high variance in the sizes of the images, ranging from roughly 430 MB to 3.7 GB and 3500 to 57,000 patches per image. The Camelyon dataset also contains pixel-level annotation of tumor regions although these were ignored, as they are unnecessary for the self-supervised approach used in DSMIL.

The TCGA datasets are a set of WSI datasets focusing on cancers with typically poor prognosis and high public health impact. The datasets are sourced and made available by the National Cancer Institute Center for Cancer Genomics. The datasets for two subtypes of lung cancer, Lung Adenocarcinoma (LUAD) and Lung Squamous Cell Carcinoma (LUSC), are used specifically. LUAD and LUSC are two of the most common subtypes of lung cancer but manifest in different biomarkers [20]. The LUAD and LUSC Datasets contain 535 and 513 WSIs, respectively, but contain slides that are significantly smaller than Camelyon 16's, with an average of roughly 5000 patches per image, but ranging from 24 to 15,000 patches per image. These datasets only contain global level labels, although this is irrelevant as pixel-wise annotations are unnecessary for our approach.

In contrast to the Camelyon 16 dataset, which allows us to evaluate our methods capability to distinguish between WSIs with and without cancer, the TCGA dataset allows us to evaluate our methods' effectiveness on differentiating between different types of the same cancer. In addition the proportions of each image containing cancerous regions differ greatly, with TCGA having 80% of each slide containing a tumor region, while Camelyon 16 samples having 10% [7].

IV. Methodology

As shown in the t-SNE plot 2, a positive class image presents a challenge in distinguishing between positive and negative patches in the pre-trained embeddings, as the embeddings are not clearly clustered. To address this challenge, DSMIL utilizes the MIL Network, but this approach falls short in integrating spatial and contextual information. This section introduces our graph method, which incorporates global context and provides a solution to this limitation. Furthermore, we present a baseline model with a simple weight aggregation transform that improves sensitivity compared to current state-of-the-art models.

A. GCN-Based Approach

We introduce a GCN module into the original DSMIL pipeline as can be seen in 3. This can be broken down into 3 primary operations: graph construction, embedding refinement with a GCN, and final classification with the

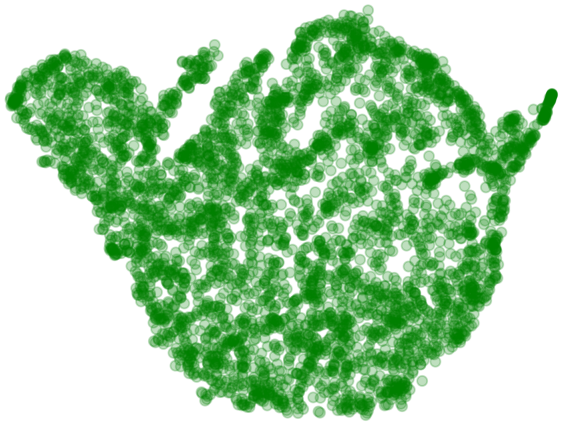

Fig. 2. t-SNE Plot of a Tumor Positive Image from Camelyon16

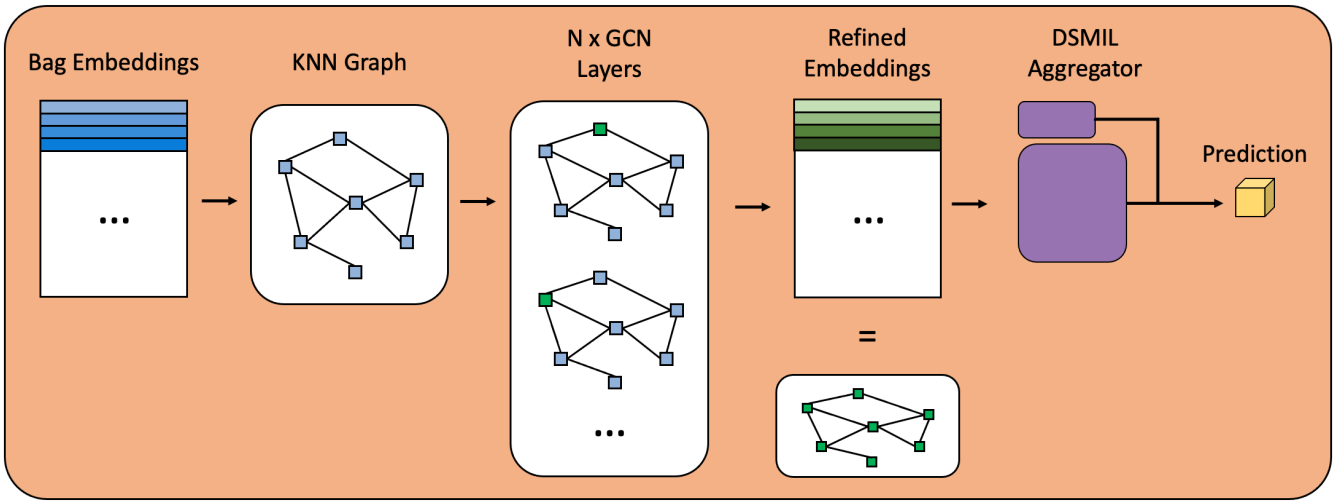

Fig. 3. Pipeline for our GCN-Based Approach

DSMIL aggregator. We first construct a graph from the set of patch embeddings, with which the values in the embedding can be further refined based on their connectivity to other patches using a GCN. We add this to the original DSMIL pipeline to allow the model to learn inter-patch relationships which are not modeled in the attention stream of DSMIL which focuses only on the critical instance.

The code for this approach is available on GitHub at https://github.com/ianwu13/EE-638-WSI-Project

1) KNN Graph Construction: Before applying a graph convolutional network, a graph must first be constructed from the patches. Keeping in line with the MIL paradigm, this must also be done in a permutation invariant manner.

Due to the large size of WSIs and the resulting high number of image patches, utilizing a fully connected graph is not an effective method. The calculation of edge weights for N^2 edges in a bag with N patches would be necessary, and applying a GCN to this graph would be computationally infeasible. As such, we construct a directional graph using a K-nearest neighbors graph (K-NNG) approach [23]. For each node, an edge is created pointing from it to each of its K nearest neighbors, where K is a hyperparameter of our model. In our evaluations, we determine the set of nearest neighbors of a node using euclidean distance between feature vectors, although other similarity measures such as cosine distance can also be used [23]. A self-edge or self-loop is also added to each node. This is referred to as the "renormalization trick" by Kipf and Welling in their paper "Semi-Supervised Classification with Graph Convolutional Networks" [22], in which they empirically show its use in GCNs improves accuracy.

The pyramidal concatenation strategy employed to construct the set of feature vectors in the bag creates embeddings for patches at higher resolutions that share many values in their feature vector with neighboring patches, making it highly probable that they share a connection in our K-NNG. This effectively results in spatial information about the patch locations being represented in the graph. The resolution levels used to generate patch embeddings is highly interrelated with the optimal hyperparameter values for K and the number of GCN layers, as this will affect the number of high resolution patches which share a lower resolution parent patch.

2) Feature Refinement with Graph Convolutional Network: The constructed graph is then passed through a GCN to "refine" the feature vectors for each node by allowing the model to learn inter-patch/node relationships for feature vectors in the MIL bag. As the name implies, GCNs essentially apply a convolution operation to their input graphs. By passing their input through some number of hidden layers, where the output of for each node n_i at each layer is a learnable combination of those nodes adjacent to n_i, the value of the node features becomes a function of itself and the features of adjacent nodes. The mathematical formulation is shown below [22], where σ is a non-linear activation function, $H^{(l)}$ is the l^{th} layer of network, and W^l is the weight matrix for the corresponding layer. $\tilde{A}$ and $\tilde{D}$ represent the adjacency and degree matrices, respectively, where the tilde represents the addition of a self loop to each node, although this is redundant for our case, as we explicitly add self-loops in our graph construction phase.

$$h^{(l+1)} = \sigma(\tilde{D}^{-\frac{1}{2}}\tilde{A}\tilde{D}^{-\frac{1}{2}}H^{(l)}W^{(l)}) \tag{2}$$

3) Classification with DSMIL aggregator: Following the application of the GCN the node features of the graph have been convolved, and now represent our "refined" embeddings. Although the output of the GCN is still a graph, the edges

are discarded, bringing the set of node features back to a bag representation. This set of node features is then passed through the DSMIL aggregator, the structure of which is described in our "related works" section, to generate the models final prediction for the bag.
.

B. Weighted Aggregation Transform (WAT)

We introduce a simple weighted aggregation transform to transform the pretrained feature embedding of patches to incorporate information from other patches through the weighted sum of their embeddings. For some pair of patches x_i and x_j the transformation is given by

$$x_i^{t+1} \leftarrow (1-\theta)x_i^t + (\theta)x_j^t \tag{3}$$

$$x_j^{t+1} \leftarrow (1-\theta)x_j^t + (\theta)x_i^t \tag{4}$$

Repeating the transformation update for N iterations will ensure enough information about tumor patches is propogated to other patches, such that it will be easier to fit the DSMIL aggregator for better classification of WSI's.

Algorithm 1 Weighted Aggregation Transform (WAT)

$t \leftarrow 0$
$N \leftarrow 100$
while $t < N$ **do**
  for some x_i^t, x_j^t in X **do**
    $x_i^{t+1} \leftarrow (1-\theta)x_i^t + (\theta)x_j^t$
    $x_j^{t+1} \leftarrow (1-\theta)x_j^t + (\theta)x_i^t$
  end for
  $t \leftarrow t+1$

end while

V. Results and Discussion

A. GCN Approach

We evaluate our GCN approach and compare it with the baseline DSMIL model using the aforementioned Camelyon 16 and TCGA Lung Cancer datasets and an 80/20 split for training and testing segments. For each of these datasets, we trained our model using $k \in \{2, 4, 8, 16, 32\}$ for graph construction and n_GCN_Layers ranging from 1 to 3 for each value of k.For each set of hyperparameters, the model was trained for 100 epochs, with the weights and results for the epoch with the highest accuracy score being saved.

1) Implementation: During training, we used the Adam optimizer with an initial learning rate of 0.0001 and a batch size of 1. To reduce computational load for graph construction, the K-NNG edges for each bag was computed using the "Ball-Tree" method [24], which has a worst-case time complexity of $O(d * n)$, as opposed to $O(n^2)$ for the brute force algorithm, where d and n are the dimensionality and number of points, respectively. Similar to the patch embeddings, the graph was pre-computed offline before the training phase.

2) Results on Camelyon16: The results of our evaluation using the Camelyon 16 dataset are shown in Figure 5. It can be seen that with the addition of our GCN module there is an increase in the model's accuracy across all configurations, with the best performance coming from the 2 GCN layer, k=8 variant and the 3 GCN layer, k=8 variant. An even more significant increase can be seen in the sensitivity scores of the model, with a 10% increase over the baseline DSMIL performance in our best case. The AUC score remains relatively similar across models. It is notable that across all values for the number of GCN layers, the highest accuracy and sensitivity scores come from the variant with k=8.

3) Results on TCGA: The results of our evaluation using the TCGA dataset are shown in Figure 6. Sensitivity and AUC are reported for both classes, as the task for the TCGA dataset is differentiating between the two subtypes of lung cancer as opposed to detection. Notably, we observe little to no improvement in accuracy with the addition of a GCN module, differing from our Camelyon 16 results. Although there is more variation in the sensitivity scores, it is more of a trade-off between the LUAD and LUSC classes, with the average sensitivity (not shown) remaining relatively constant. Similarly to our Camelyon 16 evaluation results, the AUC score remains relatively constant across models.

4) Discussion: We hypothesize that the reason for the increased performance on the Camelyon 16 dataset is that the addition of our GCN module functions to increase attention weights for important/cancerous patches in the DSMIL aggregator for cancer positive WSI samples, while in negative sample it serves to spread the attention weights more evenly across all patches. In the case that the image contains a cancerous or abnormal region, we can expect this region to be more interconnected with itself in the graph, whereas in a normal image we would expect more uniform connectivity across entire the graph. The MIL bag will also contain embeddings for background patches that do not contain tissue sample regions, which we can expect to be clustered in the graph for both positive and negative WSIs. Because of this clustering, embeddings for patches most similar to (connected to) the critical instance will have been convolved with the embedding of the critical instance by the GCN. This will result in amplified similarity between the critical instance and its connected vectors in the positive case, as their nodes will be more clustered in the graph. The degree and reach of this convolution can be tuned by adjusting the number of layers in the GCN, as the output for node n in an m layer GCN is a function of neighbors within m steps of n. The result of this increased similarity is similar to case amplification in the second stream of the DSMIL aggregator, where these patches' information vectors $\{v_i, ..., v_j\}$ will have higher attention weights in the construction of the bag embedding.

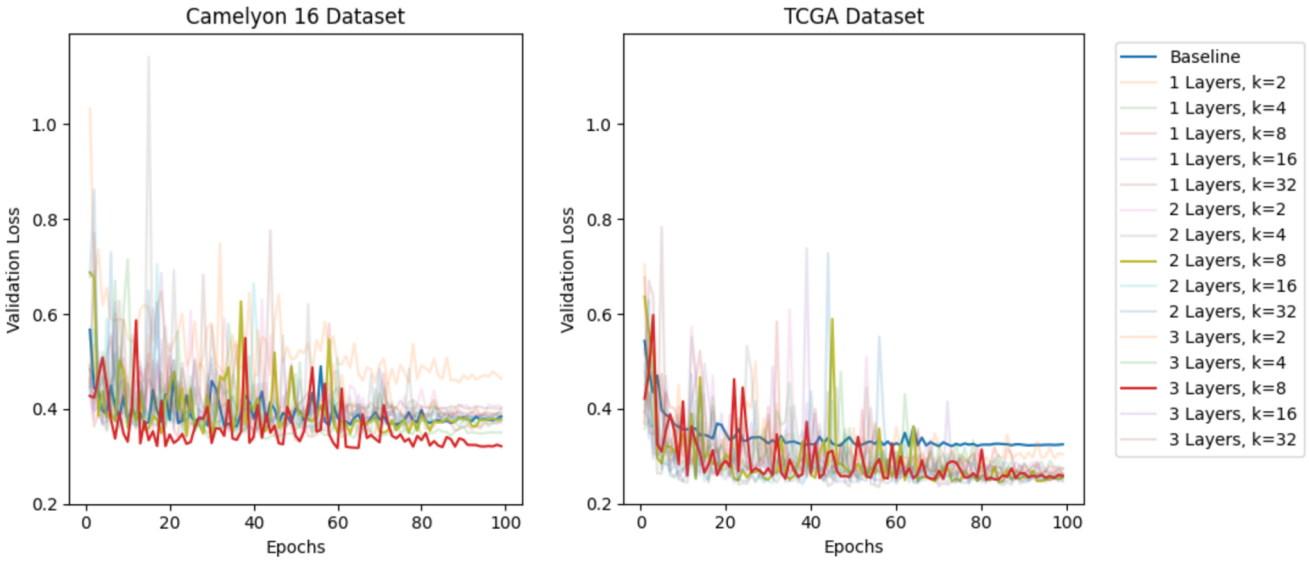

Fig. 4. Validation Loss Plot

Model	**Accuracy**	**Sensitivity**	**AUC**
DSMIL Baseline	0.9125	0.8667	0.962
WAT	0.8625	0.9	0.88
1 Layers, k=2	0.925	0.9	0.948
1 Layers, k=4	0.925	0.9667	0.9713
1 Layers, k=8	0.9375	0.9667	0.9713
1 Layers, k=16	0.925	0.9333	0.954
1 Layers, k=32	0.925	0.9333	0.9673
2 Layers, k=2	0.9125	0.9	0.9653
2 Layers, k=4	0.9375	0.9667	0.9733
2 Layers, k=8	**0.95**	0.9667	0.966
2 Layers, k=16	0.9375	0.9667	0.9633
2 Layers, k=32	0.925	0.9	0.9667
3 Layers, k=2	0.925	0.9667	0.9733
3 Layers, k=4	0.9375	0.9667	0.9733
3 Layers, k=8	**0.95**	0.9667	0.968
3 Layers, k=16	0.9375	0.9667	0.9653
3 Layers, k=32	0.925	0.9333	0.94

Fig. 5. Table of Results for Camelyon 16

Notably, we do not see the same performance increase on the TCGA dataset that we do for Camelyon 16. However, this is still compatible with our hypothesis, as the task for TCGA involves distinguishing between two subtypes of lung cancer, rather than detecting the presence of cancer metastases in Camelyon 16. In the context of differentiating between subtypes of cancer, micro features (i.e., smaller portions of the image such as individual cells) become more important for accurate classification [15]. The impact of GCN smoothing on the feature vectors can result in increased similarity but at the expense of a loss in precision of the individual values in the vector. Our results suggest that these individual values may play a more significant role in classification for subtype differentiation, rendering the GCN less effective in refining features for this task.

B. WAT Approach

We also assessed the effectiveness of our Weighted Aggregation Transform methodology, and conduct a comparative analysis against the baseline DSMIL model for the Camelyon16 and TCGA Lung Cancer datasets. The same 80/20 split used in our GCN approach was also used to evaluate this method. We transformed the embeddings using a value of θ of 0.3 and N=100. The modified embedding bags were subsequently trained using the DSMIL patch aggregator for a duration of 100 epochs.

1) Implementation: We employ a randomized sampling approach to select pairs of embeddings from the embedding bag. We then apply equations (3) and (4) to transform the pairs, and repeat the process for 100 iterations and with a constant theta value of 0.3 for each iteration. This methodology allows us to evaluate the impact of the proposed modifications on the overall performance of the system.

2) Results: Our assessment on the Camelyon 16 and TCGA dataset, as illustrated in Figure 5, 6, reveals that the proposed transformation technique underperforms relative to the baseline model for both datasets. However, we did observe a minor improvement in sensitivity by 3%, indicating that the transformation does improve the model's ability to detect positive classes.

Model	Accuracy	Sensitivity LUAD/ Specificity LUSC	Sensitivity LUSC/ Specificity LUAD	AUC LUAD	AUC LUSC
DSMIL Baseline	0.9333	0.9252	0.9515	0.9754	0.977
WAT	0.9380	0.9510	0.9252	0.9754	0.977
1 Layers, k=2	0.9381	0.9813	0.9126	0.9745	0.972
1 Layers, k=4	0.9333	0.972	0.9223	0.9752	0.9725
1 Layers, k=8	0.9143	0.9252	0.9515	0.9682	0.9746
1 Layers, k=16	0.919	0.9346	0.9515	0.9737	0.9746
1 Layers, k=32	0.9286	0.9159	0.9515	0.9721	0.9775
2 Layers, k=2	0.9333	0.972	0.9223	0.9731	0.9716
2 Layers, k=4	0.9286	0.972	0.9223	0.9747	0.9749
2 Layers, k=8	0.9238	0.9626	0.9126	0.974	0.9788
2 Layers, k=16	0.9238	0.9626	0.9223	0.9734	0.9772
2 Layers, k=32	0.9286	0.972	0.9223	0.9756	0.9702
3 Layers, k=2	0.9238	0.9533	0.932	0.9729	0.9667
3 Layers, k=4	0.9333	0.9346	0.9417	0.9735	0.9769
3 Layers, k=8	0.919	0.9346	0.9515	0.9755	0.9715
3 Layers, k=16	0.9286	0.9813	0.8932	0.9742	0.9726
3 Layers, k=32	0.9286	0.9252	0.9417	0.9629	0.9802

Fig. 6. Table of Results for TCGA

3) Discussion: The hyper-parameter θ has a significant impact on the performance, as it determines the proportion of new information incorporated into the embedding in each iteration. However, tuning this parameter can improve performance, as we observed increased sensitivity with lower values of θ. Therefore, we recommend reducing the value of θ and conducting further experiments with values less than 0.1.

VI. Future Work

In future work, it would be beneficial to expand our evaluations beyond the original GCN formulation proposed by Kipf and Welling [22]. For instance, other types of graph neural networks, such as graphSAGE [28] convolutions or graph attention networks [26], could be explored. Moreover, the varying performance results observed between the Camelyon 16 and TCGA Lung Cancer Datasets suggest that our method is more effective at detecting cancer rather than distinguishing between different subtypes or types. Further research could involve testing the model with different datasets to further evaluate its effectiveness in WSI classification tasks.

Although we have demonstrated the effectiveness of our approach using precomputed embeddings, it remains to be seen how it will perform with lower dimensional or weaker embeddings that have been generated using less powerful SimCLR backbones. Further investigation into the capabilities of our approach with weaker embeddings is necessary in future work.

Our future research endeavors involve exploring novel methods to enhance the effectiveness of our proposed approach. Specifically, we aim to incorporate weaker embeddings by leveraging self-supervised contrastive learning with lower batch size settings. Additionally, we plan to evaluate the effectiveness of our approach in the context of more complex datasets, and introduce advanced embedding transformations utilizing random processes. These future research efforts will enable us to refine our techniques and push the boundaries of WSI classification even further.

VII. Conclusion

We showcase a unique methodology for enhancing the precision and sensitivity in Whole Slide Image classification by refining pretrained embeddings. The key technical innovation in our approach is incorporating contextual information by refining the embeddings through a graph convolutional network. This new technique has shown to enhance the classification results in the cancer metastasis detection on the Camelyon 16 dataset.

Moreover, we have also proposed a straightforward weighted Aggregation transform as a baseline method, which has demonstrated improved sensitivity when compared to the current state-of-the-art, although we still see increased improvement from our GCN approach.

VIII. Acknowledgements

The results shown here are in part based upon data generated by the TCGA Research Network: https://www.cancer.gov/tcga.

Embeddings generated by the original authors of the DSMIL paper [7] were also used in producing these results.

REFERENCES

[1] Khened, M., Kori, A., Rajkumar, H., Srinivasan, B. & Krishnamurthi, G. A Generalized Deep Learning Framework for Whole-Slide Image Segmentation and Analysis. (arXiv,2020), https://arxiv.org/abs/2001.00258

[2] Carbonneau, M., Cheplygina, V., Granger, E. & Gagnon, G. Multiple instance learning: A survey of problem characteristics and applications. *Pattern Recognition*. **77** pp. 329-353 (2018,5), https://doi.org/10.1016

[3] Zhang, J., Zhang, X., Ma, K., Gupta, R., Saltz, J., Vakalopoulou, M. & Samaras, D. "Gigapixel Whole-Slide Images Classification Using Locally Supervised Learning." *Lecture Notes In Computer Science*. pp. 192-201 (2022), https://doi.org/10.1007

[4] Jia, Z., Huang, X., Chang, E. & Xu, Y. "Constrained Deep Weak Supervision for Histopathology Image Segmentation." *IEEE Transactions On Medical Imaging*. **36**, 2376-2388 (2017,11), https://doi.org/10.1109

[5] Hou, L., Samaras, D., Kurç, T., Gao, Y., Davis, J. & Saltz, J. "Efficient Multiple Instance Convolutional Neural Networks for Gigapixel Resolution Image Classification." *ArXiv*. **abs/1504.07947** (2015)

[6] Ilse, M., Tomczak, J. & Welling, M. Attention-based Deep Multiple Instance Learning. (arXiv,2018), https://arxiv.org/abs/1802.04712

[7] Li, B., Li, Y. & Eliceiri, K. Dual-stream Multiple Instance Learning Network for Whole Slide Image Classification with Self-supervised Contrastive Learning. (arXiv,2020), https://arxiv.org/abs/2011.08939

[8] Campanella, G., Hanna, M., Geneslaw, L., Miraflor, A., Silva, V., Busam, K., Brogi, E., Reuter, V., Klimstra, D. & Fuchs, T. Clinical-grade computational pathology using weakly supervised deep learning on whole slide images. *Nature Medicine*. **25** pp. 1 (2019,8)

[9] Chen, R., Chen, C., Li, Y., Chen, T., Trister, A., Krishnan, R. & Mahmood, F. Scaling Vision Transformers to Gigapixel Images via Hierarchical Self-Supervised Learning. (arXiv,2022), https://arxiv.org/abs/2206.02647

[10] Zhang, M. & Li, Q. MS-GWNN:multi-scale graph wavelet neural network for breast cancer diagnosis. (arXiv,2020), https://arxiv.org/abs/2012.14619

[11] Tu, M., Huang, J., He, X. & Zhou, B. Multiple instance learning with graph neural networks. (arXiv,2019), https://arxiv.org/abs/1906.04881

[12] Adnan, M., Kalra, S. & Tizhoosh, H. Representation Learning of Histopathology Images using Graph Neural Networks. (arXiv,2020), https://arxiv.org/abs/2004.07399

[13] Nakhli, R., Moghadam, P., Mi, H., Farahani, H., Baras, A., Gilks, B. & Bashashati, A. AMIGO: Sparse Multi-Modal Graph Transformer with Shared-Context Processing for Representation Learning of Giga-pixel Images. (arXiv,2023), https://arxiv.org/abs/2303.00865

[14] Bazargani, R., Fazli, L., Goldenberg, L., Gleave, M., Bashashati, A. & Salcudean, S. Multi-Scale Relational Graph Convolutional Network for Multiple Instance Learning in Histopathology Images. (arXiv,2022), https://arxiv.org/abs/2212.08781

[15] Gadiya, S., Anand, D. & Sethi, A. Histographs: Graphs in Histopathology. (arXiv,2019), https://arxiv.org/abs/1908.05020

[16] Wang, J., Chen, R., Lu, M., Baras, A. & Mahmood, F. Weakly Supervised Prostate TMA Classification via Graph Convolutional Networks. (arXiv,2019), https://arxiv.org/abs/1910.13328

[17] Thandiackal, K., Chen, B., Pati, P., Jaume, G., Williamson, D., Gabrani, M. & Goksel, O. Differentiable Zooming for Multiple Instance Learning on Whole-Slide Images. (arXiv,2022), https://arxiv.org/abs/2204.12454

[18] Chen, T., Kornblith, S., Norouzi, M. & Hinton, G. A Simple Framework for Contrastive Learning of Visual Representations. (2020)

[19] Ehteshami Bejnordi, B., Veta, M., Diest, P., Ginneken, B., Karssemeijer, N., Litjens, G., Laak, J. & CAMELYON16 Consortium Diagnostic Assessment of Deep Learning Algorithms for Detection of Lymph Node Metastases in Women With Breast Cancer. *JAMA*. **318**, 2199-2210 (2017,12), https://doi.org/10.1001/jama.2017.14585

[20] Chen, J. & Dhahbi, J. Lung adenocarcinoma and lung squamous cell carcinoma cancer classification, biomarker identification, and gene expression analysis using overlapping feature selection methods. *Scientific Reports*. **11** (2021)

[21] Khader, A., Braschi-Amirfarzan, M., McIntosh, L., Gosangi, B., Wortman, J., Wald, C. & Thomas, R. Importance of tumor subtypes in cancer imaging. *European Journal Of Radiology Open*. **9** pp. 100433 (2022), https://www.sciencedirect.com/science/article/pii/S2352047722000405

[22] Kipf, T. & Welling, M. Semi-Supervised Classification with Graph Convolutional Networks. (2017)

[23] Dong, W., Moses, C. & Li, K. Efficient K-Nearest Neighbor Graph Construction for Generic Similarity Measures. *Proceedings Of The 20th International Conference On World Wide Web*. pp. 577-586 (2011), https://doi.org/10.1145/1963405.1963487

[24] Shakhnarovich, G., Darrell, T. & Indyk, P. New Algorithms for Efficient High-Dimensional Nonparametric Classification. *Nearest-Neighbor Methods In Learning And Vision: Theory And Practice*. pp. 75-101 (2006)

[25] Dolatshah, M., Hadian, A. & Minaei-Bidgoli, B. Ball*-tree: Efficient spatial indexing for constrained nearest-neighbor search in metric spaces. (2015)

[26] Veličković, P., Cucurull, G., Casanova, A., Romero, A., Liò, P. & Bengio, Y. Graph Attention Networks. (2018)

[27] Liang, M., Chen, Q., Li, B., Wang, L., Wang, Y., Zhang, Y., Wang, R., Jiang, X. & Zhang, C. Interpretable classification of pathology whole-slide images using attention based context-aware graph convolutional neural network. *Computer Methods And Programs In Biomedicine*. **229** pp. 107268 (2023), https://www.sciencedirect.com/science/article/pii/S0169260722006496

[28] Hamilton, W., Ying, R. & Leskovec, J. Inductive Representation Learning on Large Graphs. *CoRR*. **abs/1706.02216** (2017), http://arxiv.org/abs/1706.02216

[29] He, K., Zhang, X., Ren, S. & Sun, J. Deep Residual Learning for Image Recognition. (2015)