DAS-MIL--: Exploring Graph Neural Networks for histopathological image classification

ELLIS Summer School

Gennaro Iannuzzo
University Parthenope of Naples, CVPRLab
gennaro.iannuzzo001@studenti.uniparthenope.it

Rutger Hendrix
Università Campus Bio-Medico di Roma
Rutger.Hendrix@unicampus.it

Andrea Favilli
ISTI-CNR, University of Pisa
andrea.favilli@isti.cnr.it

Weronika Hryniewska-Guzik
Warsaw University of Technology
weronika.hryniewska.dokt@pw.edu.pl

Luca Lumetti
University of Modena and Reggio Emilia
luca.lumetti@unimore.it

Mattia Paladino
E4 Computer Engineering
mattia.paladino-ext@e4company.com

Abstract

Histopathological image analysis is a critical area of research with the potential to aid pathologists in faster and more accurate diagnosis. However, Whole-Slide Images (WSIs) present challenges for deep learning frameworks due to their large size and lack of pixel-level annotations. Multi-Instance Learning (MIL) is a popular approach that can be employed for handling WSIs, treating each slide as a bag composed of multiple patches or instances. DAS-MIL is a recent solution that feeds a MIL model with augmented features from graph attention layers. In this paper, after a short introduction to the problem, we leverage Leonardo, the 4th supercomputer in the top500 leaderboard to explore different variations on DAS-MIL, as well as a depiction of the features extracted by the dataset to better understand the problem.

Keywords Graph Neural Networks · Multi-Instance Learning · Whole Slide Images

1 Introduction

Deep neural networks have undeniably made substantial advancements in the realm of medical image analysis, consistently pushing the boundaries of what is possible [1, 2]. These networks have achieved remarkable results in diagnosing diseases, uncovering anomalies, and providing invaluable support to clinicians in their decision-making processes. Nevertheless, a formidable challenge persists, particularly in the context of digital pathology: the precise detection of minute anomalies, such as small malignant tissues, within vast and intricate images known as Whole-Slide Images (WSIs) [3].

WSIs offer unprecedented opportunities for the preservation, sharing, and comprehensive examination of tissue specimens, ushering in a new era for pathology. However, their size makes them impractical to be fed in a GPU for standard deep learning architectures. Moreover, annotating WSIs at the pixel level demands extensive medical expertise and is a labor-intensive and time-consuming endeavor [4]. Labels for WSIs are often available at higher levels of granularity, such as the entire slide or patient level, making it challenging to pinpoint specific regions of interest within these expansive images.

To leverage the capabilities of deep learning for the analysis of WSIs, researchers have resorted to breaking down WSIs into smaller patches and utilizing them as input for neural networks [5]. While this patch-based approach is practical for

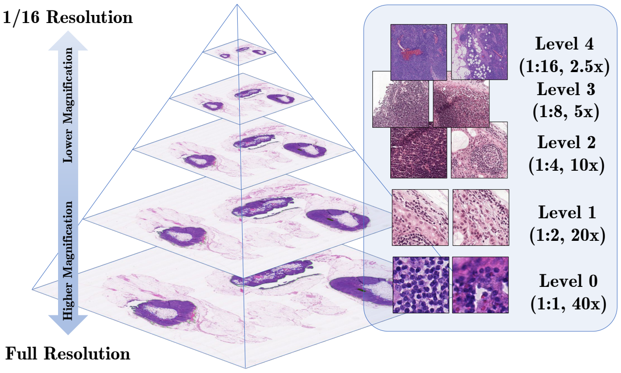

Figure 1: Visualization of the different kinds of resolutions available in a WSI.

handling a vast amount of data, having a single annotation for a whole set of patches of the same WSI makes training non-trivial and requires methods tailored for this specific task.

Multi-instance Learning (MIL) has emerged as a promising solution [6]. In MIL treats a WSI as a "bag" composed of numerous image patches, or "instances" and leverages attention mechanisms to weigh the importance of these instances in the overall classification decision. Multi-scale approaches appear to be a promising strategy to exploit the hierarchical pyramidal structure inherent in WSIs. [7] proposed a Graph-based multi-scale MIL framework named DAS-MIL. With the graph structures their work aims to improve information flow across multiple scales while enforcing latent space alignment with a Knowledge distillation scheme.

The primary scope of this project is to employ the supercomputer Leonardo [8], thus we dedicated much effort to improving the efficiency of the original pipeline and making it suitable to be scaled across multiple GPUs and multiple nodes.

Finally, this work proposes improvements to the information flow and encoding of DAS-MIL [7]. Specifically, the investigated variations concern Graph Attention Networks V2 (GATv2), which recently demonstrated superior performance w.r.t. its previous version, and graph connectivity based on K-nearest neighbors(k-NN) in the feature space, to emphasize different long-range relationships in the graph.

While this paper does not focus on achieving state-of-the-art results, it is dedicated to refining and optimizing existing techniques to better address the complex problem of analyzing WSIs and detecting subtle anomalies within them.

2 Datasets

Two datasets have been employed in all the experiments: Camelyon16 and TCGA Lung. The former has been created with the purpose of automatic detection of metastases in Hematoxylin and Eosin (H&E) stained whole-slide images of lymph node sections, as part of the homonymous challenge held at the International Symposium on Biomedical Imaging (ISBI) in 2016 [9]. The dataset comprises a total of 398 WSIs, out of which 128 are designated as "official test set". The images were acquired through two slide scanners, namely RUMC and UMCU, respectively equipped with $\times 20$ and $\times 40$ objective lenses. The specimen-level pixel sizes are comparable, i.e., $0.243\mu m \times 0.243\mu\mathrm{m}$ for RUMC and $0.226\mu\mathrm{m} \times 0.226\mu\mathrm{m}$ for UMCU. From the official training set, a subset of 22 WSIs has been extracted to serve as a validation set.

The second dataset, publicly available on the GDC Data Transfer Portal, comprises two sub-types of cancer: Lung Adenocarcinoma (LUAD) and Lung Squamous Cell Carcinoma (LUSC), counting 541 and 513 WSIs respectively. In this case, 43 WSIs of the training set have been extracted to serve as a validation set.

Both datasets had already been pre-processed using DINO [10] to embed each patch of each resolution to a lower-dimensional vector of size 384 for TCGA Lung and 256 for Camelyon16.

2.1 Pre-processing

Each slide has been cropped using the CLAM framework [11], a state-of-the-art tool for selecting tissue patches and removing the WSI background. In particular, each slide has been processed at thumbnail level through a combination of Otsu thresholding [12] and connected components analysis [13], to obtain the tissue contours. After that, each

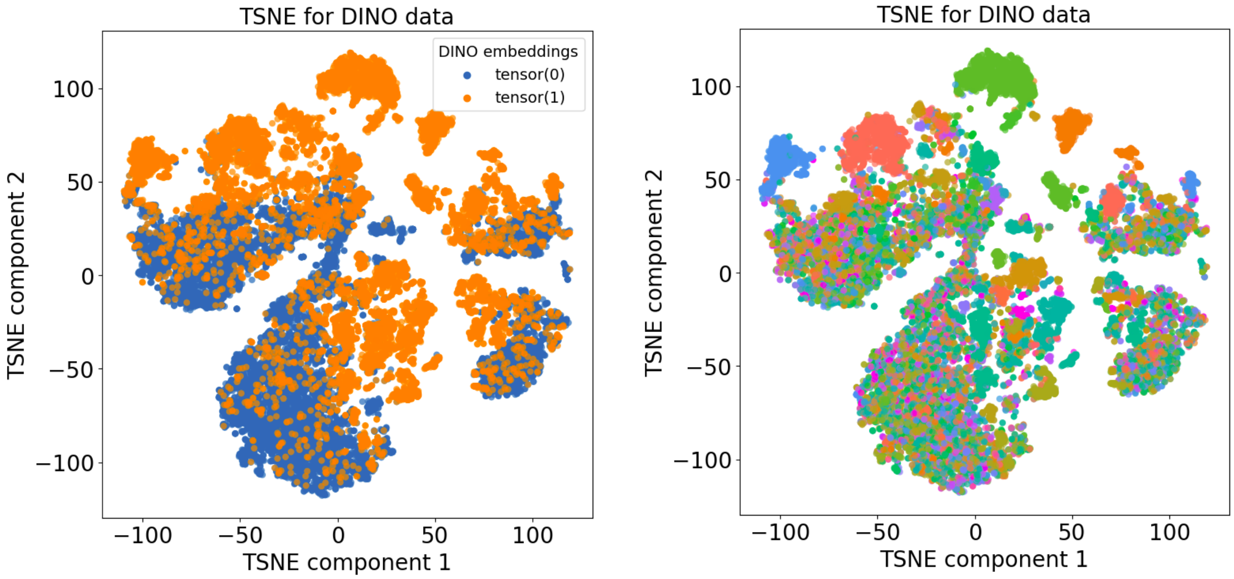

Figure 2: t-SNE visualizations of DINO embeddings, suggest the distinguishability of lesions from healthy tissue. The leftmost visualization represents the original dataset distribution, where orange dots correspond to patches that are part of a positive slide, and in blue instances which are part of a negative slide. Clusters with a mixture of blue and orange dots are due to the presence of negative patches within positive slides. In the second image, the dots with the same colors are obtained by the same slide. This highlights the effective presence of some clusters which hints a good separation between healthy and pathological patches.

256×256 patch within the selected contours is extracted without overlapping at $20\times$ scale resolution ($5\times$ and $20\times$ in the multi-scale setting).

Finally, instance embeddings are obtained through a ViT model trained in a self-supervised fashion by means of the DINO paradigm [14]. The training is performed separately on each dataset/resolution. The model has been trained for a week with two NVIDIA GeForce GTX 2080 Ti GPUs using the default parameters proposed by the authors.

3 Feature Visualization

We embarked on the creation of T-SNE visualizations for DINO embeddings with the intention of evaluating the distinguishability between lesions and healthy tissue. To maintain a manageable dataset, we opted to work with a representative 1% subset of the original training data.

To gain deeper insights into the separability of healthy and pathological patches, we conducted a random sampling procedure, ensuring an equal representation of both categories. The results, as illustrated in Figure 2, demonstrate that the DINO embeddings facilitate a clear separation between healthy tissue and those featuring tumors. This finding underscores the effectiveness of the embeddings in distinguishing between these two classes. Furthermore, we extended our analysis to WSI, as depicted on the right of Fig. 2. Here, we aimed to ascertain the extent to which the DINO embeddings enable the discrimination of slides on a broader scale. These visualizations collectively offer valuable insights into the differentiability of various tissue types within the dataset.

4 Proposed Modifications

4.1 Graph Attention Network

Graph Attention Networks (GATs) stand out as one of the most widely embraced Graph Neural Network (GNN) architectures, acclaimed as the forefront choice for graph-based representation learning. Within the framework of GAT, each node conducts an attention process toward its neighboring nodes, employing its own representation as the query. Brody et al. [15] demonstrated that GAT employs a rather restricted form of attention. Specifically, the ranking of attention scores remains unaltered regardless of the query node. To overcome this constraint, they presented a straightforward remedy by altering the sequence of operations thus introducing GATv2. They demonstrated that this dynamic graph attention variant surpasses GAT in terms of expressive power, enabling it to address a broader range of graph problems. For this reason, we decided to replace GAT with GATv2 in DAS-MIL, to understand if it could be suitable also for our specific scenario.

4.2 k-NN feature connectivity

Bontempo et al. [7] employed a Pyramidal Graph Neural Network (PGNN) structure in the original work, Figure 1. Instead of using all the available resolution levels, only the last two were employed by them, and we also followed this strategy to reduce the number of samples in the training set. At each one of the two image scales, they establish an 8-connectivity of patches based on adjacency. Two Graph ATtention layers (GAT), $\mathcal{G}_1$ and $\mathcal{G}_2$, one for each scale, process DINO features on the 8-connectivity graphs. In the end, a new GAT layer $\mathcal{G}_3$ processes the features produced by $\mathcal{G}_1$ and $\mathcal{G}_2$ on a global graph that merges the nodes from the two scales according to a relation "part of", Fig. 3 (a).

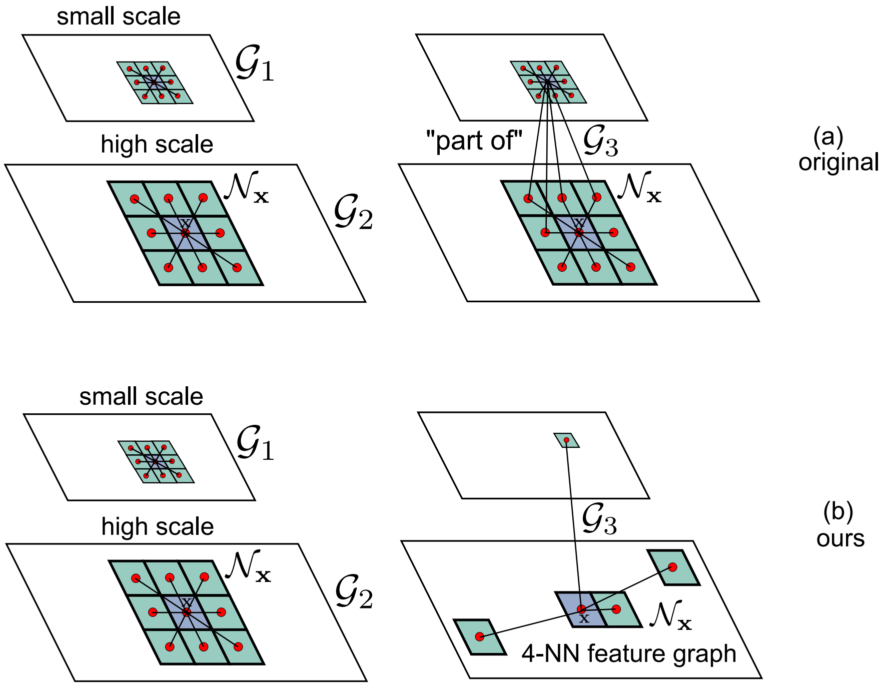

Figure 3: (a): The graph structure of the original work. Layers $\mathcal{G}_1$ and $\mathcal{G}_2$ work on the 8-connectivity graphs of the two scales, $\mathcal{G}_3$ matches the two scales according to the relation "part of". (b): Our local-to-global graph structure. $\mathcal{G}_1$ and $\mathcal{G}_2$ are the same, and $\mathcal{G}_3$ works on a scale-agnostic 4-nearest-neighbor graph on the space of the features.

Cancer is sometimes *multifocal*, meaning that tumors can grow non-locally as separated spots. In this work, we consider graph connections in K-Nearest-Neighbors of feature vectors along with the original 8-connectivity graph. A KNN graph edge links each patch node with the k nodes having the lower Euclidean distance of feature vectors. This approach makes the learning model aware both of local patch adjacency and non-local patch similarity. We do not change the attention layers $\mathcal{G}_1$ and $\mathcal{G}_2$, while we rethink $\mathcal{G}_3$ by removing the “part of” relation and adopting a scale-agnostic and global KNN graph, as depicted in Figure 3 (b). In the experiments, the number of neighbors k has been set to 4.

5 Experimental Evaluation

5.1 Method

As the main topic of the school, as well as the scope of the project, regards “Large-Scale AI”, the first effort was about optimization and parallelization of the whole pipeline. The original code supported a only batch size of 1, a single GPU, and a single node. After our additions, we managed to increase the batch size, the GPU number, and also the number of concurrent nodes, thus reducing by a lot the time required to train the model. The original work employed the test set also as a validation set. This approach could lead to biases, and most importantly unfairness when we compare the metrics obtained as a high value might be just due to randomness and not due to a real improvement of the model. Moreover, we decided to perform 3 runs for each experiment, instead of 1 as in the original paper, to report the mean and the standard deviation of the results, in order to have more robust statistical support to our statements. To gain insight into the distinguishability between malignant tissue and healthy tissue features extracted by DINO, exploratory t-SNE visualizations for patch feature embeddings have been performed and depicted in Fig. 2.

The validation accuracy has been used to identify the best-performing model. This model is then evaluated on the test set and the accuracies and AUC are reported. In the Tab. 1 all the metrics obtained using the proposed methodologies, as well as a baseline value, are reported.

5.2 Results and Discussions

Table 1: Performance comparison on Camelyon16 and TCGA Lung dataset.

	Camelyon16		**TCGA Lung**	
Model	**Accuracy**	**AUC**	**Accuracy**	**AUC**
Baseline	$\mathbf{0.906 \pm 0.030}$	0.907 ± 0.033	$\mathbf{0.874 \pm 0.027}$	0.950 ± 0.010
GATv2	0.896 ± 0.007	$\mathbf{0.937 \pm 0.040}$	0.849 ± 0.050	$\mathbf{0.955 \pm 0.009}$
KNN Graph	0.870 ± 0.010	0.933 ± 0.002	0.837 ± 0.010	0.918 ± 0.007

Regarding the findings reported in the original paper [7], our baseline experiments scored lower on Accuracy and AUC. However, our study employs a more robust methodology for model evaluation and training, which could partially explain the observed differences. Although parallel computing can reduce performance, we don't expect this to be the main contributor. Interestingly, all models have better accuracy performance on Camelyon16 than TCGA Lung. With an AUC of 0.937 ± 0004 and 0.955 ± 0009, GATv2 outperforms the other models for Camelyon16 and TCGA Lung, respectively.

The k-NN graph approach does not show any improvement. The reason for this can be attributed to the low dimensional encoding with 4 connectivity. Intuitively, this approach suits the DAS framework, especially for the high scale graph: for k-NN-based feature connectivity, similar features tend to originate from the same region in the high-scale image. Consequently, the structural differences between a spatial connectivity graph (low-scale) and a feature k-NN connectivity graph (high-scale) are likely to be relatively minimal. This similarity translates into a comparable information flow behavior when dealing with 2 different connectivity-based graphs. Therefore, the use of KD loss is still applicable to encourage the agreement between information in different scales. In conclusion, we recommend further investigation on higher connective graphs, at different places within the network.

6 Conclusions

In summary, as deep neural networks continue to revolutionize medical image analysis, our work focuses on fine-tuning and enhancing existing methods rather than claiming state-of-the-art results. Our proposed modifications, including the use of graph layers, the creation of KNN edges, and the scaling of the whole pipeline aim to make strides in improving the efficiency of WSI classification models, ultimately contributing to the ongoing advancement of digital pathology.

References

[1] Salome Kazeminia, Christoph Baur, Arjan Kuijper, Bram van Ginneken, Nassir Navab, Shadi Albarqouni, and Anirban Mukhopadhyay. GANs for medical image analysis. *Artificial Intelligence in Medicine*, 109(August): 101938, 2020. doi:10.1016/j.artmed.2020.101938.

[2] Weronika Hryniewska, Przemysław Bombiński, Patryk Szatkowski, Paulina Tomaszewska, Artur Przelaskowski, and Przemysław Biecek. Checklist for responsible deep learning modeling of medical images based on covid-19 detection studies. *Pattern Recognition*, 118:108035, 2021. doi:https://doi.org/10.1016/j.patcog.2021.108035.

[3] Kun Fan, Shibo Wen, and Zhuofu Deng. Deep learning for detecting breast cancer metastases on wsi. In Yen-Wei Chen, Alfred Zimmermann, Robert J. Howlett, and Lakhmi C. Jain, editors, *Innovation in Medicine and Healthcare Systems, and Multimedia*, pages 137–145, 2019.

[4] Sara P. Oliveira, Pedro C. Neto, João Fraga, Diana Montezuma, Ana Monteiro, João Monteiro, Liliana Ribeiro, Sofia Gonçalves, Isabel M. Pinto, and Jaime S. Cardoso. Cad systems for colorectal cancer from wsi are still not ready for clinical acceptance. *Scientific Reports*, 11:14358, 2021. doi:10.1038/s41598-021-93746-z.

[5] Jiandong Ye, Yihao Luo, Chuang Zhu, Fang Liu, and Yue Zhang. Breast cancer image classification on wsi with spatial correlations. In *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 1219–1223, 2019. doi:10.1109/ICASSP.2019.8682560.

[6] Neofytos Dimitriou, Ognjen Arandjelović, and Peter D Caie. Deep learning for whole slide image analysis: an overview. *Frontiers in medicine*, 6:264, 2019.

[7] Gianpaolo Bontempo, Angelo Porrello, Federico Bolelli, Simone Calderara, and Elisa Ficarra. DAS-MIL: Distilling Across Scales for MIL Classification of Histological WSIs. In *Medical Image Computing and Computer Assisted Intervention*, 2023.

[8] Matteo Turisini, Giorgio Amati, and Mirko Cestari. Leonardo: A pan-european pre-exascale supercomputer for hpc and ai applications. *arXiv preprint arXiv:2307.16885*, 2023.

[9] Babak Ehteshami Bejnordi, Mitko Veta, Paul Johannes Van Diest, Bram Van Ginneken, Nico Karssemeijer, Geert Litjens, Jeroen AWM Van Der Laak, Meyke Hermsen, Quirine F Manson, Maschenka Balkenhol, et al. Diagnostic Assessment of Deep Learning Algorithms for Detection of Lymph Node Metastases in Women With Breast Cancer. *Jama*, 318(22):2199–2210, 2017.

[10] Mathilde Caron, Hugo Touvron, Ishan Misra, Hervé Jégou, Julien Mairal, Piotr Bojanowski, and Armand Joulin. Emerging properties in self-supervised vision transformers. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 9650–9660, 2021.

[11] Ming Y Lu, Drew FK Williamson, Tiffany Y Chen, Richard J Chen, Matteo Barbieri, and Faisal Mahmood. Data-efficient and weakly supervised computational pathology on whole-slide images. *Nature Biomedical Engineering*, 5(6):555–570, 2021.

[12] Jun Zhang and Jinglu Hu. Image Segmentation Based on 2D Otsu Method with Histogram Analysis. In *International Conference on Computer Science and Software Engineering*, volume 6, pages 105–108. IEEE, 2008.

[13] Stefano Allegretti, Federico Bolelli, Michele Cancilla, Federico Pollastri, Laura Canalini, and Costantino Grana. How does Connected Components Labeling with Decision Trees perform on GPUs? In *Computer Analysis of Images and Patterns*, pages 39–51. Springer, 2019. ISBN 978-3-030-29887-6.

[14] Mathilde Caron, Hugo Touvron, Ishan Misra, Hervé Jégou, Julien Mairal, Piotr Bojanowski, and Armand Joulin. Emerging Properties in Self-Supervised Vision Transformers. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 9650–9660, 2021.

[15] Shaked Brody, Uri Alon, and Eran Yahav. How attentive are graph attention networks? *arXiv preprint arXiv:2105.14491*, 2021.