Multiple instance learning with pre-contextual knowledge

Andrea Grandi, Daniele Vellani

{275074,196186}@studenti.unimore.it

Data

Abstract

The visual examination of histopathological images is a cornerstone of cancer diagnosis, requiring pathologists to analyze tissue sections across multiple magnifications to identify tumor cells and subtypes. However, existing attention-based Multiple Instance Learning (MIL) models for Whole Slide Image (WSI) analysis often neglect contextual and numerical features, resulting in limited interpretability and potential misclassifications. Furthermore, the original MIL formulation incorrectly assumes the patches of the same image to be independent, leading to a loss of spatial context as information flows through the network. Incorporating contextual knowledge into predictions is particularly important given the inclination for cancerous cells to form clusters and the presence of spatial indicators for tumors. To address these limitations, we propose an enhanced MIL framework that integrates pre-contextual numerical information derived from semantic segmentation. Specifically, our approach combines visual features with nuclei-level numerical attributes, such as cell density and morphological diversity, extracted using advanced segmentation tools like Cellpose. These enriched features are then fed into a modified Buffer-MIL model for WSI classification. We evaluate our method on subtyping non-small cell lung cancer (TCGA-NSCLC) and detecting lymph node metastases (CAMELYON16 and CAMELYON17), achieving test AUCs of (INSERIRE I VALORI DOPO EVALUATION) respectively.

1. Introduction

In recent years, computational pathology has emerged as a transformative tool for cancer research, leveraging Whole Slide Images (WSIs) to extract meaningful insights into tissue architecture and cellular composition. These large, high-resolution images are invaluable for diagnosing and prognosticating cancer, yet their sheer size, heterogeneity, and reliance on detailed annotations pose substantial challenges. One computational challenge is the large size of WSIs, of the order of 100,000 × 100,000 pixels. Processing images of such size with deep neural network directly is not possible with the GPUs commonly available. Overcoming this problem, previous work proposes to tessellate each WSI into thousands of smaller images called tiles and global survival prediction per slide is obtained in two steps. The tiles are first embedded into a space of lower dimension using a pre-trained feature extractor model, and a MIL model is trained to predict survival from the set of tiles embeddings of a WSI (Herrera et al., 2016) [1].

Multiple Instance Learning (MIL) has become a pivotal paradigm for WSI analysis. By treating a slide as a "bag" of smaller patches (instances), MIL allows slide-level predictions without the need for pixel-level annotations, streamlining the analysis pipeline (Ilse et al., 2018; Campanella et al., 2019) [2, 3]. Despite its utility, traditional MIL approaches often overlook critical contextual and numerical information that can enhance interpretability and predictive accuracy.

One limitation of MIL is the assumption that tiles from the same WSI are independent (Ilse et al., 2018) [2]. In particular, MIL models take into account only the visual knowledge comes from WSIs. In contrast, pathologists take into account also other aspects of WSIs in their analysis. Addressing these limitations requires innovative approaches capable of combining visual and numerical features from WSIs effectively (Litjens et al., 2017; Campanella et al., 2019) [3, 4].

In this work, we introduce a novel pipeline that integrates cutting-edge tools and methodologies to overcome these limitations. We preprocess WSIs using the CLAM framework (Lu et al., 2021) [5], ensuring the retention of essential visual features. To extract nuclei-specific numerical features such as cell counts and density, we utilize Cellpose (Stringer et al., 2021) [6], a state-of-the-art segmentation algorithm. Simultaneously, we employ DINO (Caron et al., 2021) [7], a self-supervised vision transformer, to generate embeddings representing the visual content of each patch. By concatenating these numerical and visual features, we construct a richer, more informative representation for each patch.

Our key innovation lies in adapting the BufferMIL framework to incorporate these enriched embeddings, by assigning greater importance to patches with high cell density or other critical numerical features

This paper is structured as follows: Section 2 reviews key advancements in MIL and its applications in computational pathology. Section 3 describes our methodology, detailing preprocessing, feature extraction, and the enhancements made to BufferMIL. Section 4 presents experimental results, discusses their implications, and outlines potential future directions. By combining numerical and visual features, our work seeks to advance computational pathology and provide deeper insights into the analysis of WSIs.

The source code is publicly available at `https://github.com/andrea-grandi/bio_project`.

2. Related Work

Multiple Instance Learning has revolutionized computational pathology by enabling efficient WSI classification without exhaustive pixel-level annotations. Under MIL formulation, the prediction of a WSI label can come either directly from the tile predictions (instance-based) (Campanella et al.,2019) [3], or from a higher-level bag representation resulting from aggregation of the tile features (bag embedding-based) (Ilse et al., 2018) [2]. The bag embedding-based approach has empirically demonstrated superior performance (Sharma et al., 2021) [8]. Most recent bag embedding-based approaches employ attention mechanisms, which assign an attention score to every tile reflecting its relative contribution to the collective WSI-level representation. Attention scores enable the automatic localization of sub-regions of high diagnostic value in addition to informing the WSI level label.

One of the first important work in this field was DS-MIL (Li et al., 2021) [9]. This model utilizes a dual-stream framework, where patches are extracted from different magnifications (e.g., $5\times$ and $20\times$ in their study) of Whole Slide Images. These patches are processed separately for self-supervised contrastive learning. The embeddings obtained from patches at various resolutions are then concatenated to train the MIL aggregator, which assigns an importance or criticality score to each patch. The most critical patch is selected and compared to all others in a one-vs-all manner. This comparison uses a distance metric inspired by attention mechanisms, though it differs significantly by comparing two queries instead of the traditional key-query setup. Finally, the distances are aggregated to generate the final bag-level prediction.

Another work is BufferMIL (Bontempo et al, 2023) [10], which is a notable framework that enhances MIL by incorporating explicit domain knowledge for histopathological image analysis, particularly addressing challenges like class imbalance and covariate shift. In this approach, a buffer is maintained to store the most representative instances from each disease-positive slide in the training set. An attention mechanism then compares all instances against this buffer to identify the most critical ones within a given slide. This strategy ensures that the model focuses on the most informative instances, thereby improving its generalization performance. By leveraging a buffer to track critical instances and employing an attention mechanism for comparison, Buffer-MIL effectively mitigates issues related to class imbalance and covariate shift. This approach enhances the model's ability to focus on the most informative patches within WSIs, leading to more accurate and reliable predictions in histopathological image analysis.

Building upon the attention-based methodologies of frameworks like BufferMIL, Context-Aware MIL (CAMIL) (Fourkioti et al., 2024) [11] extends the concept of informed instance selection by introducing neighbor-constrained attention mechanisms. CAMIL leverages spatial dependencies among WSI tiles to achieve superior performance in cancer subtyping and metastasis detection, showcasing the importance of spatial context in WSI analysis. Similarly, the Nuclei-Level Prior Knowledge Constrained MIL (NPKC-MIL) (Wang et al., 2024) [12] highlights the value of combining handcrafted nuclei-level features with deep learning, demonstrating improvements in interpretability and classification accuracy for breast cancer WSIs.

3. Methods

4. Experiments and Results

5. Conclusions

References

[1] Herrera, Francisco and Ventura, Sebastián and Bello, Rafael and Cornelis, Chris and Zafra, Amelia and Sánchez-Tarragó, Dánel and Vluymans, Sarah. *Multiple instance learning : foundations and algorithms.* Springer, 2016.

[2] Maximilian Ilse, Jakub M Tomczak, and Max Welling. Attention-based deep multiple instance learning. *arXiv preprint arXiv:1802.04712*, 2018.

[3] Gabriele Campanella, Matthew G Hanna, Liron Geneslaw, et al. Clinical-grade computational pathology using weakly supervised deep learning on whole slide images. *Nature medicine*, 25(8):1301–1309, 2019.

[4] Geert Litjens, Thijs Kooi, Babak Ehteshami Bejnordi, et al. A survey on deep learning in medical image analysis. *Medical image analysis*, 42:60–88, 2017.

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

[6] Carsen Stringer, Tim Wang, Michael Michaelos, and Marius Pachitariu. Cellpose: a generalist algorithm for cellular segmentation. *Nature methods*, 18(1):100–106, 2021.

[7] Mathilde Caron, Hugo Touvron, Ishan Misra, et al. Emerging properties in self-supervised vision transformers. *arXiv preprint arXiv:2104.14294*, 2021.

[8] Yash Sharma, Aman Shrivastava, Lubaina Ehsan, Christopher A. Moskaluk, Sana Syed, and Donald E. Brown. Cluster-to-conquer: A framework for end-to-end multi-instance learning for whole slide image classification. In Mattias P. Heinrich, Qi Dou, Marleen de Bruijne, Jan Lellmann, Alexander Schlaefer, and Floris Ernst, editors, *MIDL*, volume 143 of *Proceedings of Machine Learning Research*, pages 682–698. PMLR, 2021.

[9] Bin Li, Yin Li, and Kevin W. Eliceiri. Dual-stream multiple instance learning network for whole slide

image classification with self-supervised contrastive learning, 2021.

[10] Gianpaolo Bontempo, Luca Lumetti, Angelo Porrello, Federico Bolelli, Simone Calderara, and Elisa Ficarra. Buffer-mil: Robust multi-instance learning with a buffer-based approach. In *Image Analysis and Processing – ICIAP 2023: 22nd International Conference, ICIAP 2023, Udine, Italy, September 11–15, 2023, Proceedings, Part II*, page 1–12, Berlin, Heidelberg, 2023. Springer-Verlag.

[11] Olga Fourkioti, Matt De Vries, Chen Jin, Daniel C. Alexander, and Chris Bakal. Camil: Context-aware multiple instance learning for cancer detection and subtyping in whole slide images, 2023.

[12] Xunping Wang and Wei Yuan. Nuclei-level prior knowledge constrained multiple instance learning for breast histopathology whole slide image classification. *iScience*, 27(6):109826, 2024.