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 BufferMIL 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 $\times$ 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 [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, ensuring the retention of essential visual features (Lu et al., 2021) [5]. To extract nuclei-specific numerical features such as cell counts and density, we utilize Cellpose, a state-of-the-art segmentation algorithm (Stringer et al., 2021) [6]. Simultaneously, we employ DINO, a self-supervised vision transformer, to generate embeddings representing the visual content of each patch (Caron et al., 2021) [7]. 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, our model improves sensitivity to diagnostically relevant regions. This dual-feature approach enhances both interpretability and predictive performance, addressing long-standing challenges in WSI classification.

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 (MIL) has revolutionized computational pathology by enabling efficient WSI classification without exhaustive pixel-level annotations. Despite its potential, traditional MIL approaches face limitations in capturing inter-instance relationships and integrating domain-specific knowledge. Recent advancements have sought to address these challenges.

BufferMIL [8] is a notable framework that enhances MIL by incorporating explicit domain knowledge. Unlike traditional attention-based models that treat instances independently or rely on implicit spatial relationships, BufferMIL enables the integration of prior information, significantly improving interpretability and robustness. This aligns with the efforts of DeeMILIP, which demonstrated how probabilistic constraints can map domain entities to model components, enhancing generalization under weak supervision [**?**].

The Context-Aware MIL (CAMIL) model introduced neighbor-constrained attention mechanisms, leveraging spatial dependencies among WSI tiles to achieve superior performance in cancer subtyping and metastasis detection [9]. Similarly, the Nuclei-Level Prior Knowledge Constrained MIL (NPKC-MIL) framework highlighted the benefits of combining handcrafted nuclei-level features with deep learning, demonstrating improvements in interpretability and classification accuracy for breast cancer WSIs [10].

Building on these advancements, our approach integrates nuclei-specific numerical features extracted via Cellpose into the BufferMIL framework. This integration enriches the bag representation with critical cellular attributes, such as density and morphological diversity, while leveraging high-dimensional visual embeddings from DINO. This combination bridges the gap between domain-specific insights and generalizable deep learning models, pushing the boundaries of WSI analysis.

In summary, the integration of prior knowledge into MIL frameworks, exemplified by BufferMIL, CAMIL, and NPKC-MIL, represents a paradigm shift. These models enhance both classification performance and interpretability, offering promising tools for computational pathology and personalized medicine.

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] 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.

[9] 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.

[10] 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.