MUSTANG: Multi-Stain Self-Attention Graph Multiple Instance Learning Pipeline for Histopathology Whole Slide Images

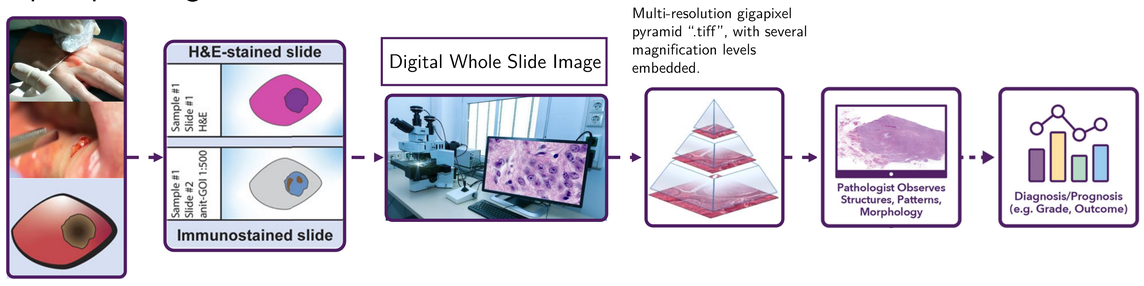
Amaya Gallagher-Syed[1,2], Luca Rossi[4], Felice Rivellese[3], Costantino Pitzalis[3], Myles J. Lewis[3], Michael R. Barnes[1,2], Gregory Slabaugh[2]

{[1]Centre for Translational Bioinformatics [2]Digital Environment Research Institute [3]Centre for Experimental Medicine & Rheumatology}, Queen Mary University of London, UK

[4]Department of Electronic and Information Engineering, The Hong Kong Polytechnic University, Hong Kong

Background

- Whole Slide Images (WSIs) are gigapixel (100k x 100k pixels) multi-resolution digital stained pathology images. Semi-quantitative examination of multi-stain sets of WSIs is used as the gold standard for diagnosis and subtyping of cancer/autoimmune diseases by expert pathologists.



- Due to their size WSIs present a challenging computer vision task, with Transformer solutions rendered intractable due to their quadratic complexity.
- Multiple Instance Learning (MIL) techniques divide the images into thousands of smaller patches, each inheriting the noisy patient label, with trainable pooling layers then used to aggregate the information.
- These approaches often employ attention, which can fail to capture long-range dependencies. Applying self attention between patches is desirable to accurately catch complex diagnostic patterns.

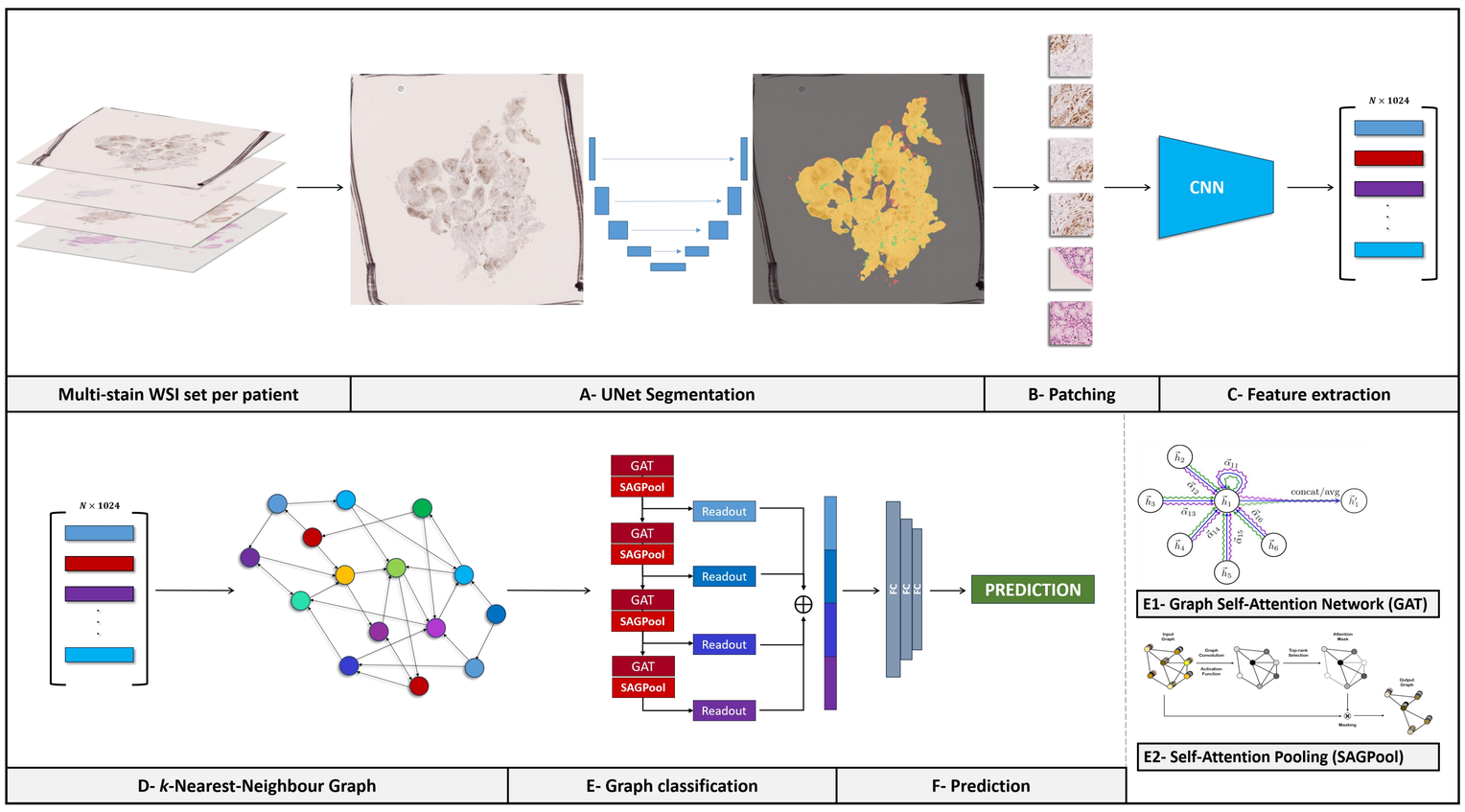
Contribution

- We propose a novel end-to-end multi-stain self-attention graph (MUSTANG) multiple instance learning pipeline. MUSTANG solves a weakly-supervised gigapixel multi-image classification task, where the label is assigned at the patient-level across a set of images, but no image/patch-level labels are available.
- The pipeline introduces a self-attention-based approach by restricting attention operations to a highly sparse k-Nearest Neighbour Graph (k-NNG) of embedded WSI patches based on Euclidean distance.
- Our approach does not require registration of WSIs, preprocessing or labelling of regions of interest, nor any feature engineering for the embedded feature vectors, making it straightforward and flexible to apply to real-world clinical datasets.

Proposed Method

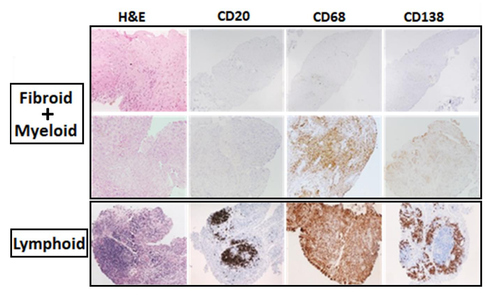
Paper Code

- **A - Segmentation**: UNet segmentation of tissue areas on the WSIs.
- **B - Patching**: The tissue area is divided into N patches.
- **C - Feature extraction**: The image patches are each passed through a CNN feature extractor and embedded into a *[1×1024]* feature vector. All feature vectors for a patient - across the multi-stain WSI set - are aggregated into a *[N×1024]* matrix.



- **D - k-Nearest-Neighbour Graph**: The *[N×1024]* matrix is used to create a sparse directed k-NNG using Euclidean distance, with $k = 5$. The attribute of each node corresponds to a [1×1024] feature vector and the graph represent the WSIs set. This graph is used as input to the GNN.
- **E - Graph classification**: The k-NNG is successively passed through four Graph Attention Network layers (E1 - GAT) and SAGPooling (E2 – SAGPool) layers. The readouts from each layer are concatenated and passed through three MLP layers and finally classified.
- **F - Prediction**: A diagnosis prediction is obtained at the patient-level.

Dataset



Binary classification of multi-stain pathology subtypes. Lymphoid vs Fibroid + Myeloid. These groups differ in disease trajectory, and drug response.

The **Rheumatoid Arthritis R4RA clinical trial dataset** was gathered in 20 European centres and recruited a total of 164 patients who underwent ultrasound-guided synovial biopsy of a clinically active joint. The synovial tissue samples were then stained with Hematoxylin & Eosin (H&E) and Immunochemistry (IHC) stains. Each dye highlights different cellular components: H&E nuclei, CD20+ B cells, CD68+ macrophages and IHC+ CD138 plasma cells and each contain complimentary information about the underlying disease process. The dataset has a total of 651 WSIs, with a variable number of WSIs per patient and a total of 309,248 non-overlapping 224×224 pixels patches extracted at 10x magnification.

Results

MUSTANG outperforms the benchmark by 5 percentage points, with a substantially similar test runtime. It accurately identifies both correct and true positives (Sens.=0.93), an important consideration in healthcare. MUSTANG does not show accuracy loss on single-stain WSIs, whereas CLAM performs better on single-stain WSIs. This shows MUSTANG can better integrate long range dependencies across stains, identifying complex spatial arrangements pertaining to disease subtyping.

Multi-stain results

	F1-score	AUC	Sens	Spec	Params [M]	Test runtime [min]
CLAM	0.84	0.88	0.86	0.82	0.47	10
MUSTANG (ours)	**0.89**	**0.92**	**0.93**	0.82	3.29	11

Single-stain results – F1-score

	CD138	CD68	CD20	HE
CLAM	0.85	0.87	**0.88**	0.76
MUSTANG (ours)	**0.89**	**0.89**	0.87	**0.78**

The graph becomes weakly connected in k=5, coinciding with increased accuracy, indicating the sparsest weakly connected graph efficiently propagates information across stains. Conversely, the graph loses structure with each GNN layer: we posit upper layers capture long range dependencies and lower micro tissue architecture.

Figure 6: On left a set of WSI corresponding to a patient and on right its corresponding k-NNG ($k = 5$) plotted using Networkx's spring-layout.

The spring layout shows closely connected nodes clustering together.

There is a good degree of mixing between WSIs, indicating information can flow between them.

Figure 4: k-NNG layout for different values of k. From left to right, $k = 1$, $k = 2$, $k = 3$, $k = 4$ and $k = 5$.

Figure 5: k-NNG graph connectivity layout after each GAT + SAGPool layer, with pooling ratio of 0.5

Future work

- Visualisation of attention heatmaps for added interpretability of clinical results.
- Benchmarking on public datasets (TCGA/Camelyon16).
- Structure aware graph transformer.

References

- P. Velickovic, G. Cucurull, A. Casanova, A. Romero, P. Liò, Y. Bengio. Graph Attention Networks. In Proc. 6th ICLR, 2018.
- J. Lee, I. Lee, and J. Kang. Self-attention graph pooling. In Proc. 36th ICML, 2019.
- M. Y. Lu, D. F. K. Williamson, T. Y. Chen, et al., "Data-efficient and weakly supervised computational pathology on whole-slide images," Nat. Biomed. Eng, 2021.

Queen Mary University of London · Digital Environment Research Institute · R4RA clinical trial · wellcome · BMVC 2023