Localized Contrastive and Attention Based Multiple Instance Learning for Automatic Staging of Histopathology Images

Narmada Naik,[1] Angela Crabtree,[2] Carlo Bifulco,[2] Brian Piening,[2] Ganapati Srinivasa,[1] Kevin Matlock[1]

[1]dātma Health Science, Beaverton, Oregon [2]Earle A. Chiles Research Institute, Portland, Oregon

JUNE 18-22, 2023 CVPR VANCOUVER, CANADA

Summary

Background: A key tool in the diagnosis and treatment of cancer lies in the analysis of digital pathology images. To categorize the tumor, the pathologist evaluates a number of morphological characteristics, such as the size and shape of cancer cells, their level of differentiation, and whether or not they have invaded nearby tissues. Due to the difficulty and time involved in such evaluations, providing computer vision methods to alleviate the burden of a pathologists would benefit society by reducing the cost and increasing availability of health care.

Contribution: We propose a new framework that will automatically classify digital pathology images scanned from Hematoxylin and Eosin (H&E) stained tissues biopsies from breast cancer tumors into their pathological stage.

Dataset: We evaluated on both the Breast Invasive Carcinoma (TCGA-BRCA) and the Lung Adenocarcinoma (TCGA-LUAD) datasets. For each dataset we downloaded and parsed their Whole Slide Image (WSI). The full distribution of each dataset with regards to their pathological stage is shown in the table below.

Table 1. Dataset

Dataset	1	2	3	4	Total
TCGA-BRCA	42	40	42	18	142
TCGA-LUAD	47	47	46	46	186

Methodology

Preprocessing: Tissue region are first extracted from the background using a HSV thresholding method (Figure 2). An additional layer of filtering is applied that extracts only the tiles that contains tumorous tissues[1]. As our model relies on multi-resolution features, both 40x and 10x magnified tiles are extracted from the source image where four 40x tiles are extracted from their 10x equivalent regions using a square grid pattern.

Training: A representational model of the WSI using broad labels applied to the image is trained using contrastive learning[2]. Finally, the WSI image is classified into the appropriate stage by classifying each tile contained within it.

Multi-Instance Learning: Making a final prediction using a bagging classifier using Multi Instance Learning. In this scenario a singular WSI would represent a single "bag" wherein the extracted sets of tiles are the individual instances.

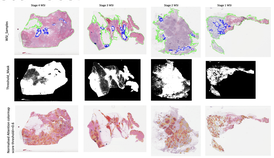

Figure 2. WSI Segmentation and Extraction of Instances

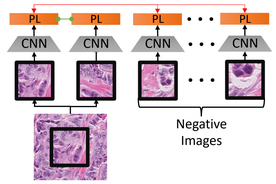

Figure 3. Contrastive Learning Pipeline[2]

Overview

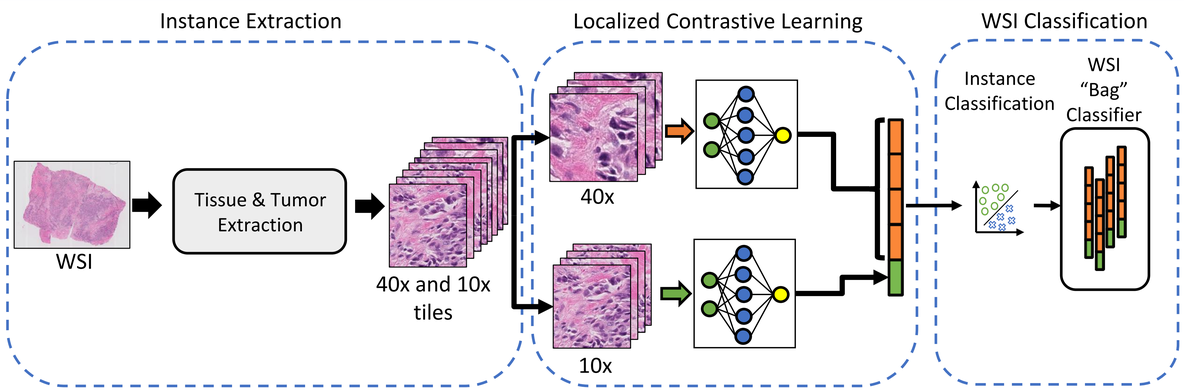

Figure 1.High Level Overview of the Proposed Framework. Multi-scale tiles are selectively extracted from WSI based on their tissue content (left). Separate Contrastive Learning layers are trained for each scale to create a one-dimensional representation which are concatenated to create a set of instance features (middle). These features are passed through an instance classifier, the outputs of which are used to create the final WSI or "bag" classification (right).

Results

Results on LUAD and BRCA datasets:

Stages	10x	40x	Multiscale
Stage 1	0.68±0.016	0.58±0.002	**0.69±0.0012**
Stage 2	0.61±0.018	0.60±0.07	**0.63±0.031**
Stage 3	0.61±0.021	0.58±0.006	**0.62±0.0030**
Stage 4	0.79±0.0011	0.77±0.0019	**0.81±0.013**

Table 2. Per Stage(class) AUC for Localized contrastive Learning (CL) based feature encoder and MIL based classification for TCGA BRCA.

Features	SVM Kernel	Accuracy
Multiscale	Linear	0.84±0.0097
	RBF	0.84±0.0082
	Polynomial	0.81±0.091
	Sigmoid	0.68±0.0015
40x	Linear	0.54±0.0032
	RBF	0.55±0.0043
	Polynomial	0.53±0.0017
	Sigmoid	0.39±0.0015
10x	Linear	0.79±0.0172
	RBF	0.78±0.0097
	Polynomial	0.76±0.0912
	Sigmoid	0.59±0.0181 height

Table 3. Tile based AUC for Localized Contrastive based feature and LDA Projection with SVM instance classification.

Models	10x	40x	Multiscale
Resnet18	0.61±0.0021	0.58±0.0017	NA
Resnet50	0.64±0.0010	0.62±0.0029	NA
EfficientNet V2	0.66±0.0017	0.64±0.0091	NA
Resnet18+MIL	0.64±0.0012	0.57±0.0072	0.66
Resnet50+MIL	0.67±0.0002	0.60±0.0018	0.69±0.0097
EfficientNet V2+ MIL	0.68±0.0037	0.62±0.0072	0.70±0.00147
DSMIL	0.62±0.031	0.59±0.0142	0.71±0.0088
CLAM	NA	0.67±0.0068	NA
Resnet50 + Localized CL + SVM	0.70±0.0018	0.65±0.0190	0.72±0.0010
EfficientNetV2 + Localized CL +MIL	0.72±0.0043	0.69±0.0039	0.74±0.0021

Table 4. Comparison of different models for WSI classification on TCGA-BRCA dataset. Accuracy is measured using AUC. Best performing model is in bold.

Models	5x	20x	Multiscale
Resnet50 + Local.CL +SVM	0.74 ±0.0032	0.72 ±0.00137	0.76 ±0.0133
Resnet50 + Local.CL+ MIL	0.73 ±0.0148	0.69 ±0.0044	0.75 ±0.0167

Table 5. AUC with different pretrained models TCGA-LUAD, the bold represent the best performance of each model with respect to scale

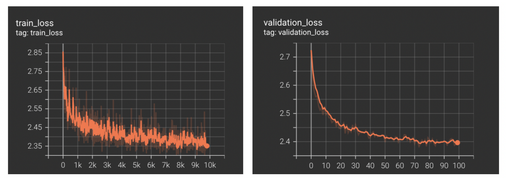

Figure 4. Contrastive loss for Training the TCGA-BRCA dataset with EfficientNet-v2, Y Axis shows the Supervised NT-Xent loss across the training iterations.

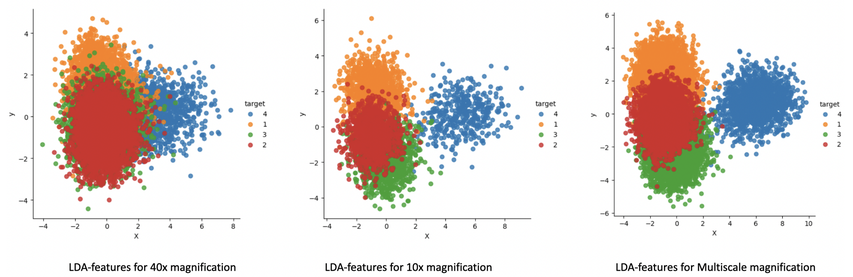

Figure 5. LDA projection of the Contrastive Features generated from the TCGA-BRCA dataset.

Future Work: While digital pathology is capable in characterizing the tumor environment, it is insufficient to fully characterize a tumor[3]. To achieve this, we must start incorporating advance molecular testing in the form of Next-Generation Sequencing (NGS). By combining morphological and genomic characteristics we can enhance our models and better predict patient prognosis.

References:

[1] Lu *et al.* Data-efficient and weakly supervised computational pathology on whole-slide images, Nature Biomedical Engineering, 2021.

[2] Tavolara *et al.* Contrastive Multiple Instance Learning: An Unsupervised Framework for Learning Slide Level Representations of Whole Slide Histopathology Images without Labels. MDPI, 2022.

[3] Crabtree *et al.* Automated staging of breast cancer histopathology images using deep learning. In NeurIPS, 2022.