DGR-MIL: Exploring Diverse Global Representation in Multiple Instance Learning for Whole Slide Image Classification

Wenhui Zhu[1]* Xiwen Chen[2]* Peijie Qiu[3]*
Aristeidis Sotiras[3] Abolfazl Razi[2] Yalin Wang[1]

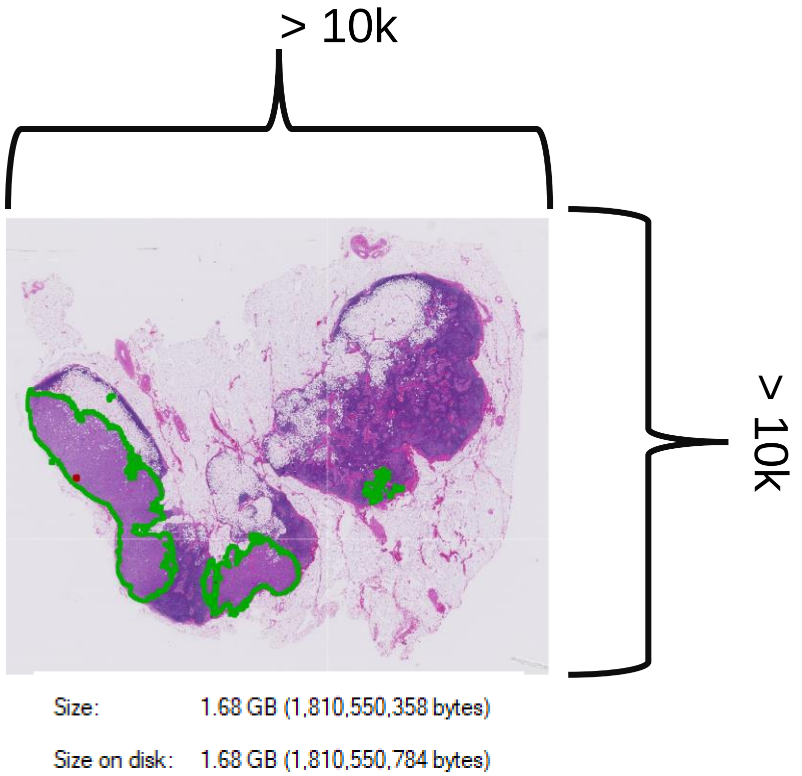
1 Arizona State University, 2 Clemson University,
3 Washington University in St. Louis

EUROPEAN CONFERENCE ON COMPUTER VISION
MILANO 2024

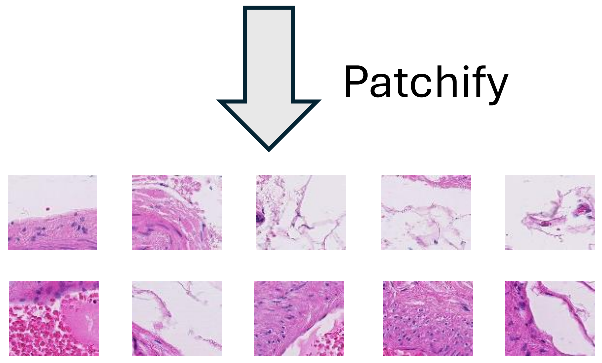
Paper & Code

Background

- Histological whole slide images (WSIs) are commonly used to diagnose a variety of cancers, e.g., breast cancer, lung cancer, etc.
- **Challenge: An WSI is often gigapixels.**
  - **Typical ML cannot process it.**
  - **Labor-intensive to annotate**
- Pipeline of Multiple Instance Learning (MIL):
  - Crop it into some small patches (~10k).
  - Each patch is an **instance**, and an WSI image is a **bag** that contains a collection of instances.
  - If at least one instance is positive (has tumor), the bag is positive.



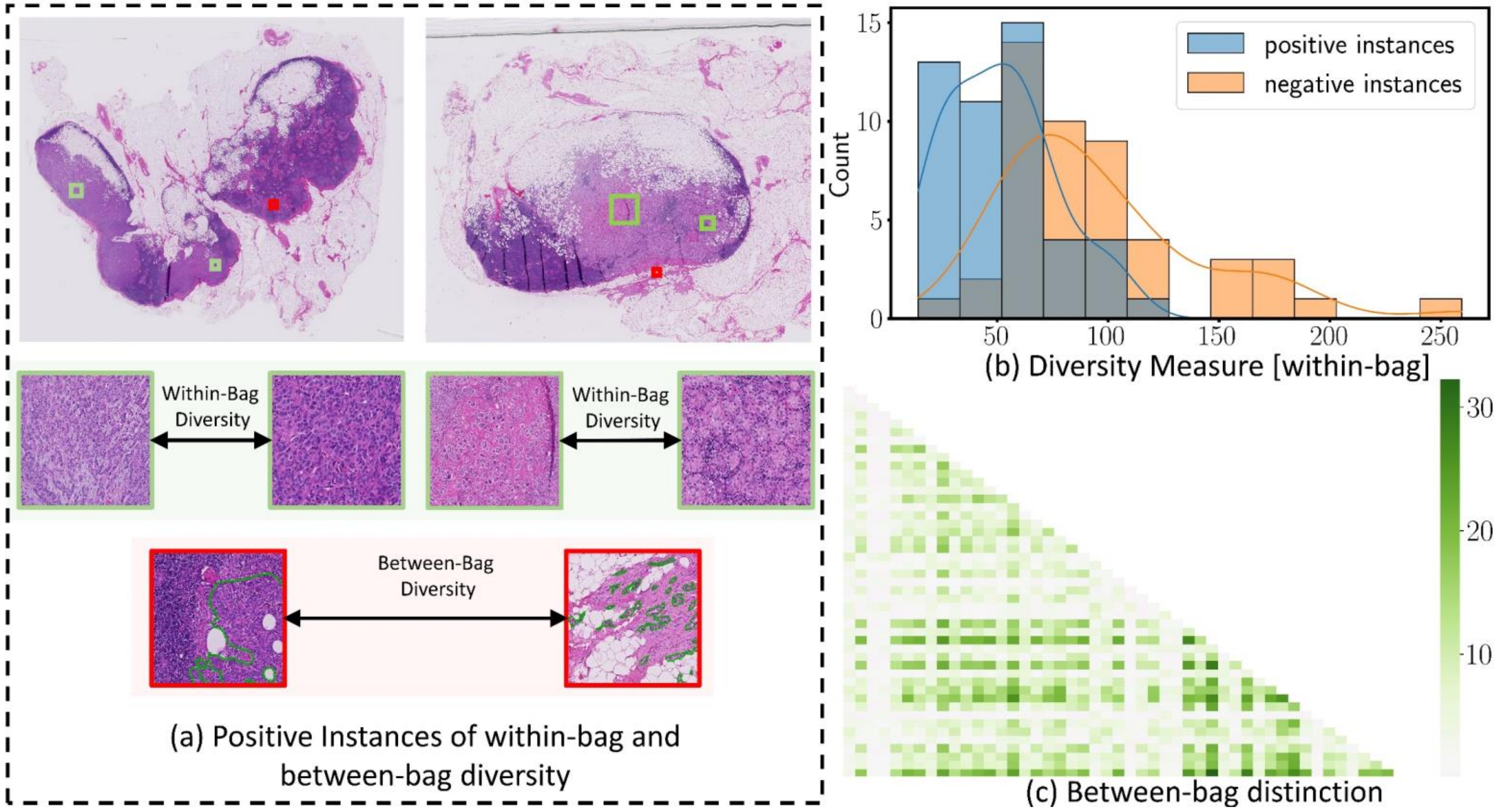
whole-slide images (WSIs) of sentinel lymph node



https://camelyon16.grand-challenge.org/Data/

Motivation

- Most MIL models for analyzing WSIs use the attention-based MIL (AB-MIL) framework, which treats instances **independently and ignores correlations.**
- While follow-up models address this by focusing **on instance correlations** within the same category, they overlook variations in phenotype, size, and spatial diversity, leading to incorrect correlations.
- Using **rate-distortion theory**, we quantify the diversity of instances, showing both between- and within-bag variations.

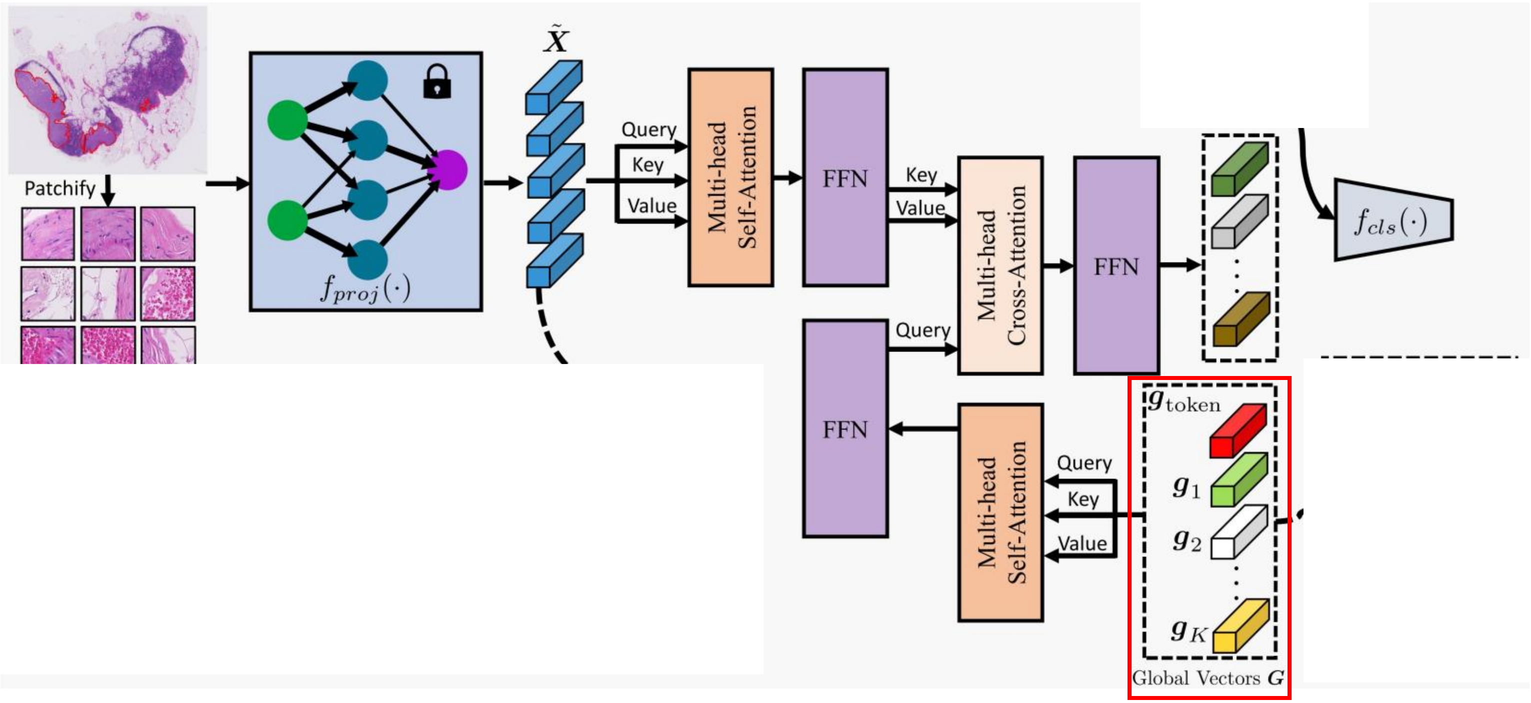


(a) Positive Instances of within-bag and between-bag diversity

(b) Diversity Measure [within-bag]

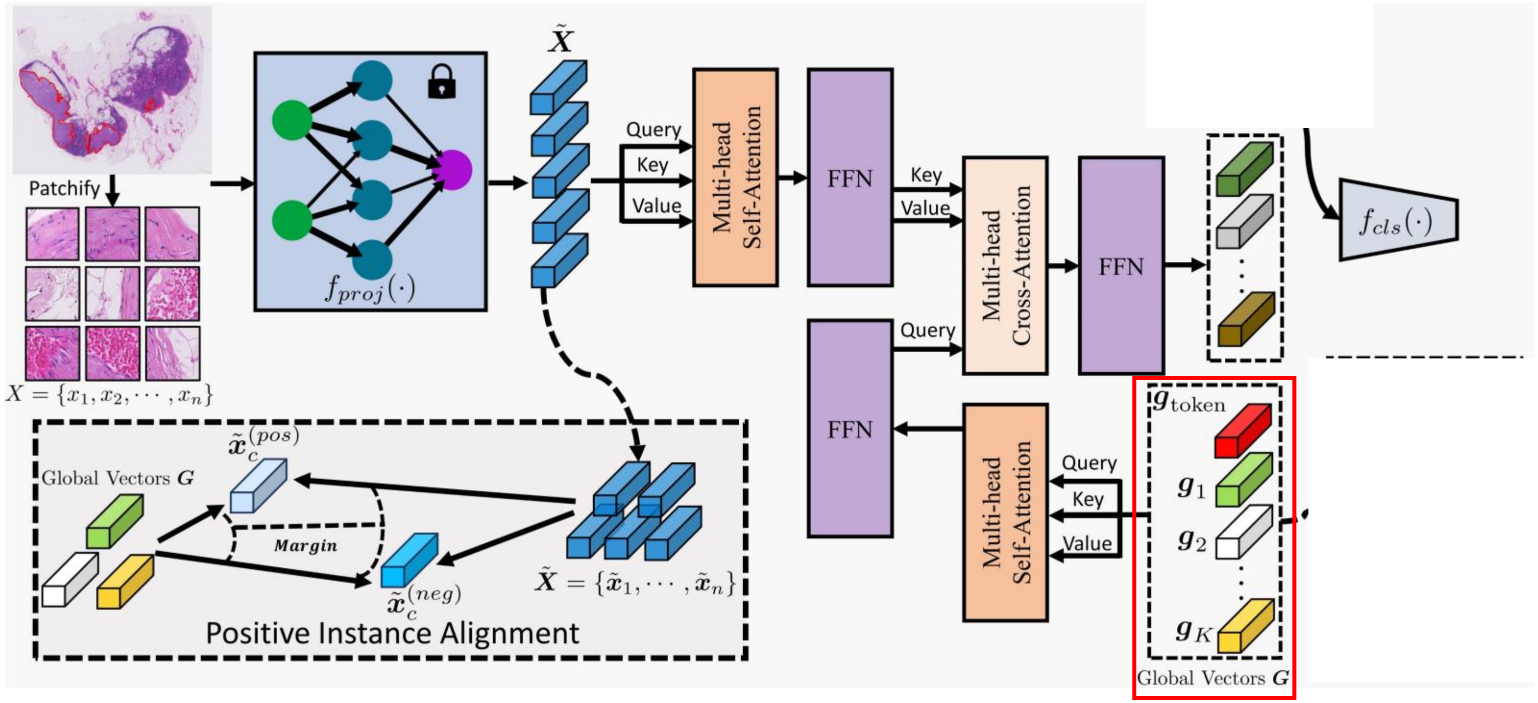
(c) Between-bag distinction

Proposed Method



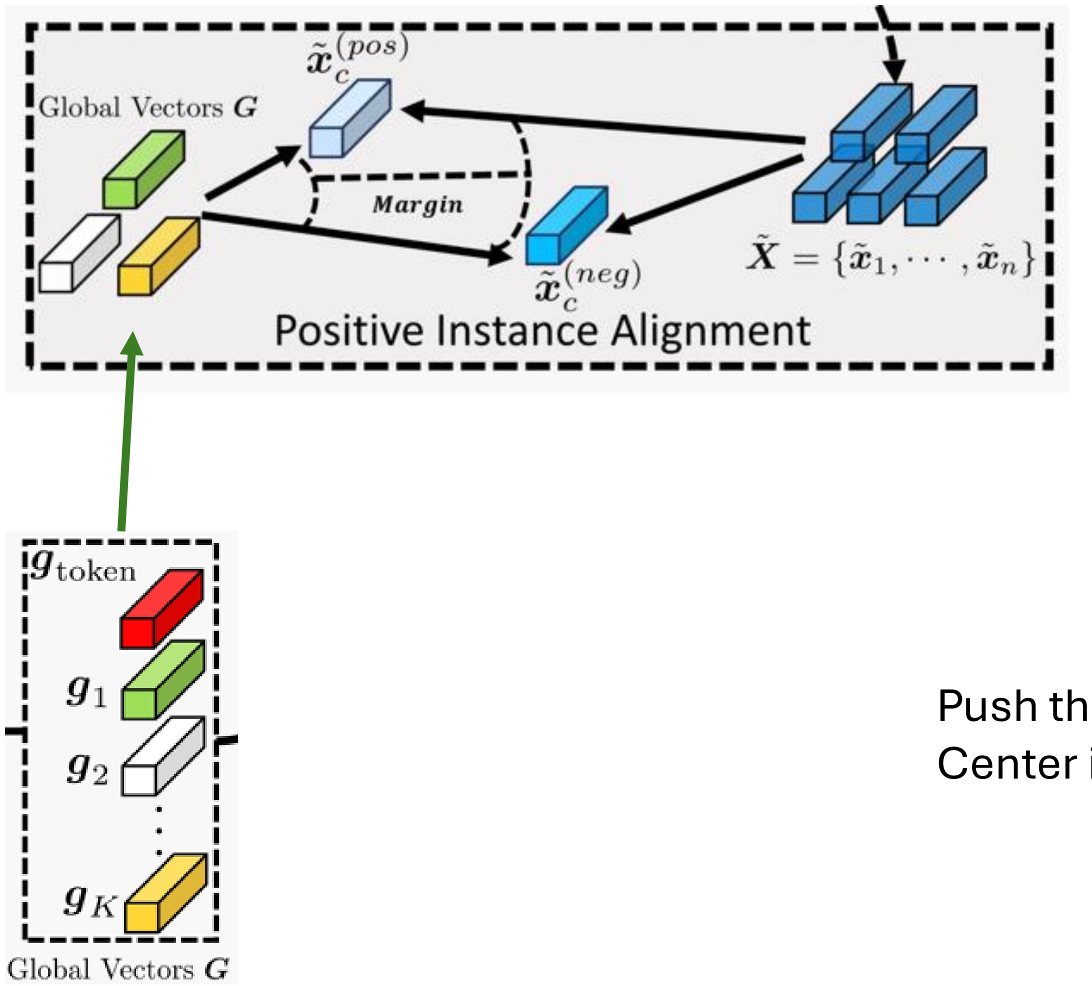
- We suggest using global learnable vectors to help network to learning diversity. The global vector will gather similar instance together by cross attention.
- K,V from instance embeddings, Q from the global vectors.
- Two mechanisms to learn a reliable and diverse global vector:
    - Positive instance alignment
    - DPP diversity loss (theoretical guaranteed).
- A class token to summarize all global vectors for final bag-level classification.

Proposed Method



- **Two mechanisms to learn a reliable and diverse global vector:**
  - Positive instance alignment
  - DPP diversity loss (theoretical guaranteed).
- **A class token to summarize all global vectors for final bag-level classification.**

Positive Instance Alignment (reliable G)



Center of positive bag:

$$\tilde{\boldsymbol{x}}_c^{(pos)} = m\tilde{\boldsymbol{x}}_c^{(pos)} + (1-m)\frac{1}{|\mathcal{I}_{pos}|}\sum_{i\in\mathcal{I}_{pos}} \tilde{x}_i$$

Center of negative bag:

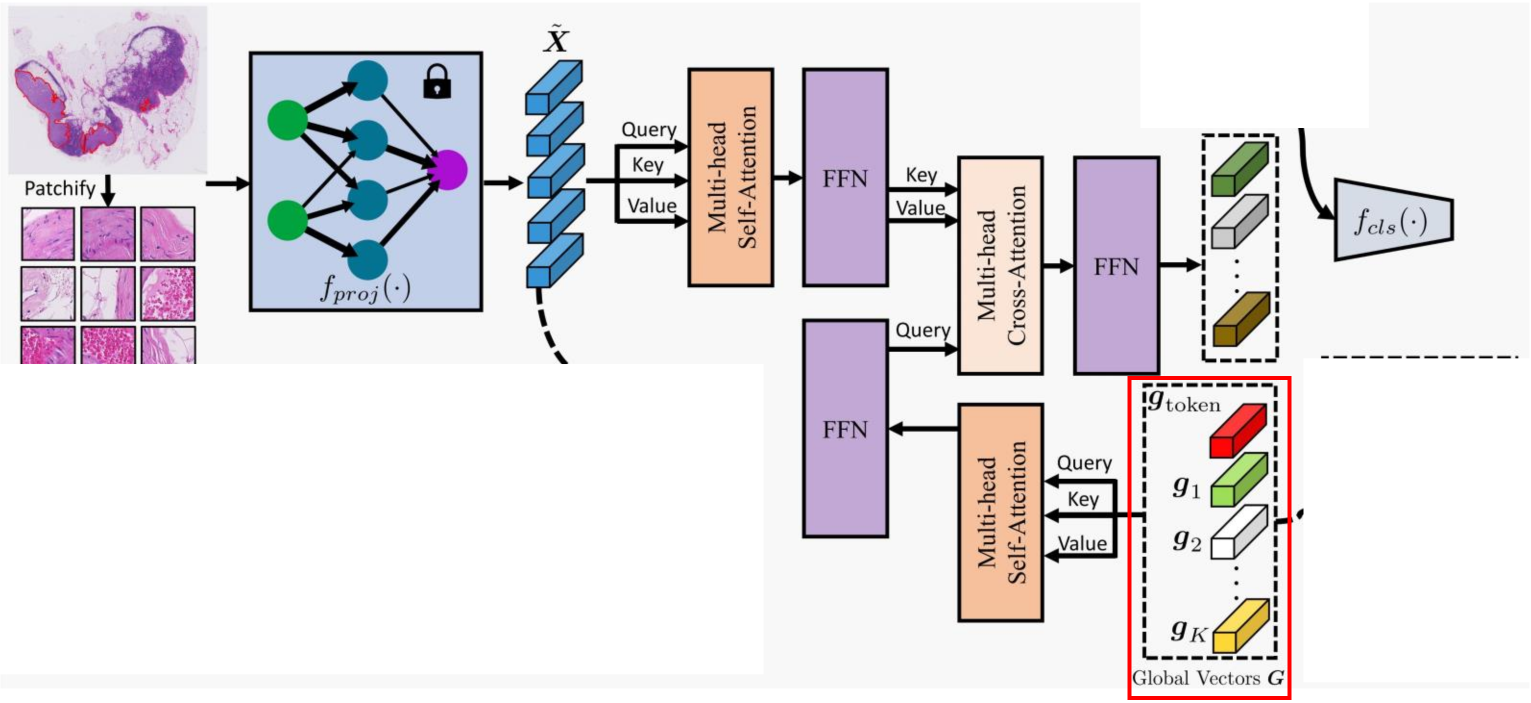
$$\tilde{\boldsymbol{x}}_c^{(neg)} = m\tilde{\boldsymbol{x}}_c^{(neg)} + (1-m)\frac{1}{|\mathcal{I}_{neg}|}\sum_{i\in\mathcal{I}_{neg}} \tilde{x}_i,$$

Triplet loss:

$$\mathcal{L}_{tri} = \sum_{k=1}^{K}[d_+(G_k, \tilde{x}_c^{(pos)}) - d_-(G_k, \tilde{x}_c^{(neg)}) + \mu]_+,$$

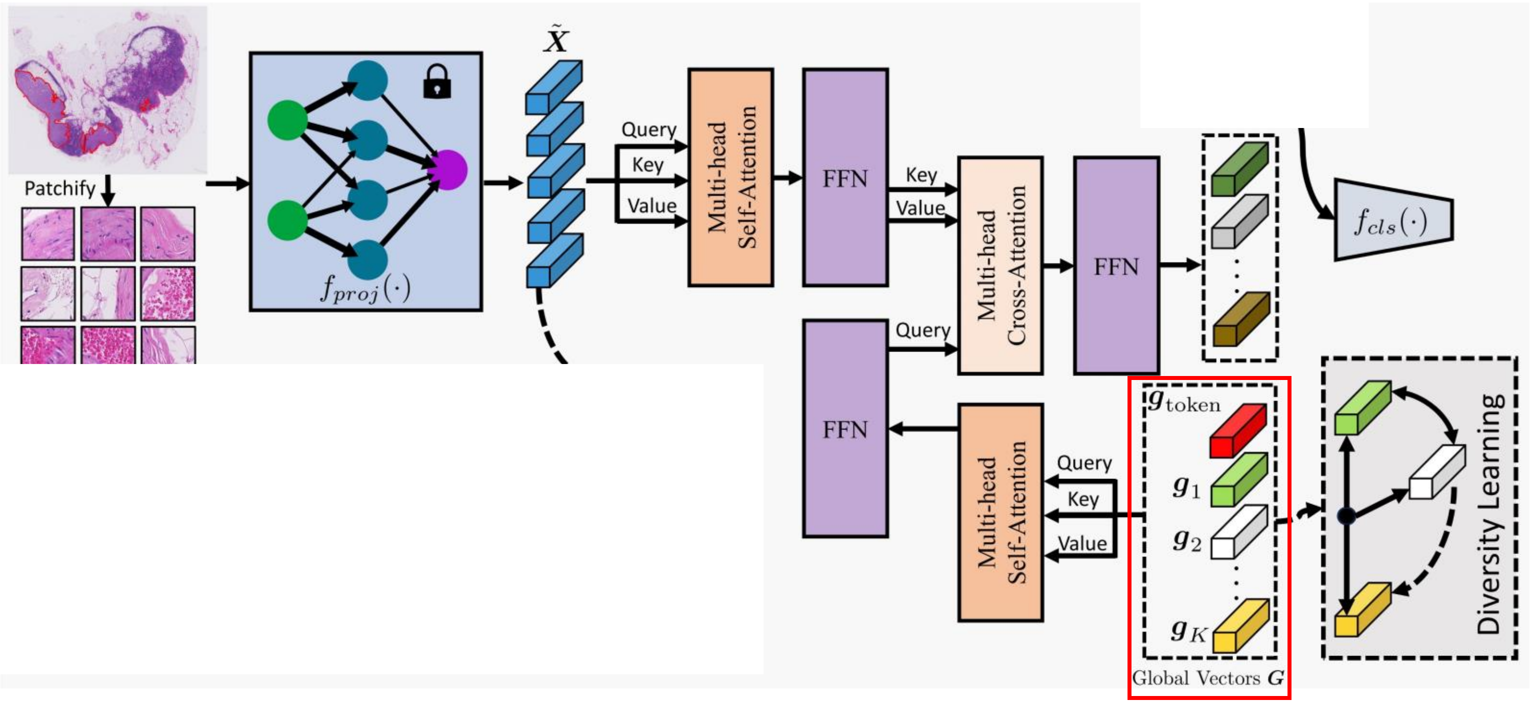
Push the global vectors close to the center of positive bags.
Center is updated in a momentum fashion for stable training.

Proposed Method



- **We suggest using global learnable vectors to help network to learning diversity. The global vector will gather similar instance together by cross attention.**
- **Two mechanisms to learn a reliable and diverse global vector:**
  - Positive instance alignment
  - DPP diversity loss (theoretical guaranteed).
- **A class token to summarize all global vectors for final bag-level classification.**

Proposed Method



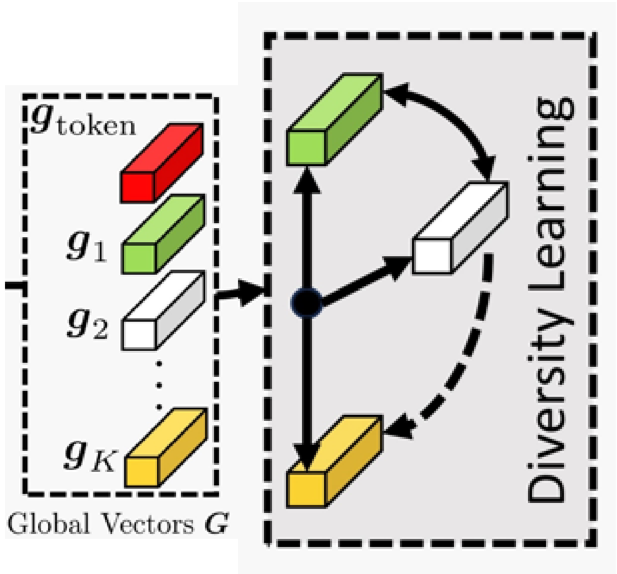
- **We suggest using global learnable vectors to help network to learning diversity. The global vector will gather similar instance together by cross attention.**
- **Two mechanisms to learn a reliable and diverse global vector:**
  - Positive instance alignment
  - DPP diversity loss (theoretical guaranteed).
- **A class token to summarize all global vectors for final bag-level classification.**

DPP-based Diversity Loss

- Determinantal Point Process (DPP): a probabilistic model of repulsion to select diverse subsets.
- Instead of using DPP to select subsets, we use it as a differentiable diversity measurement.
- It is theoretic guaranteed.

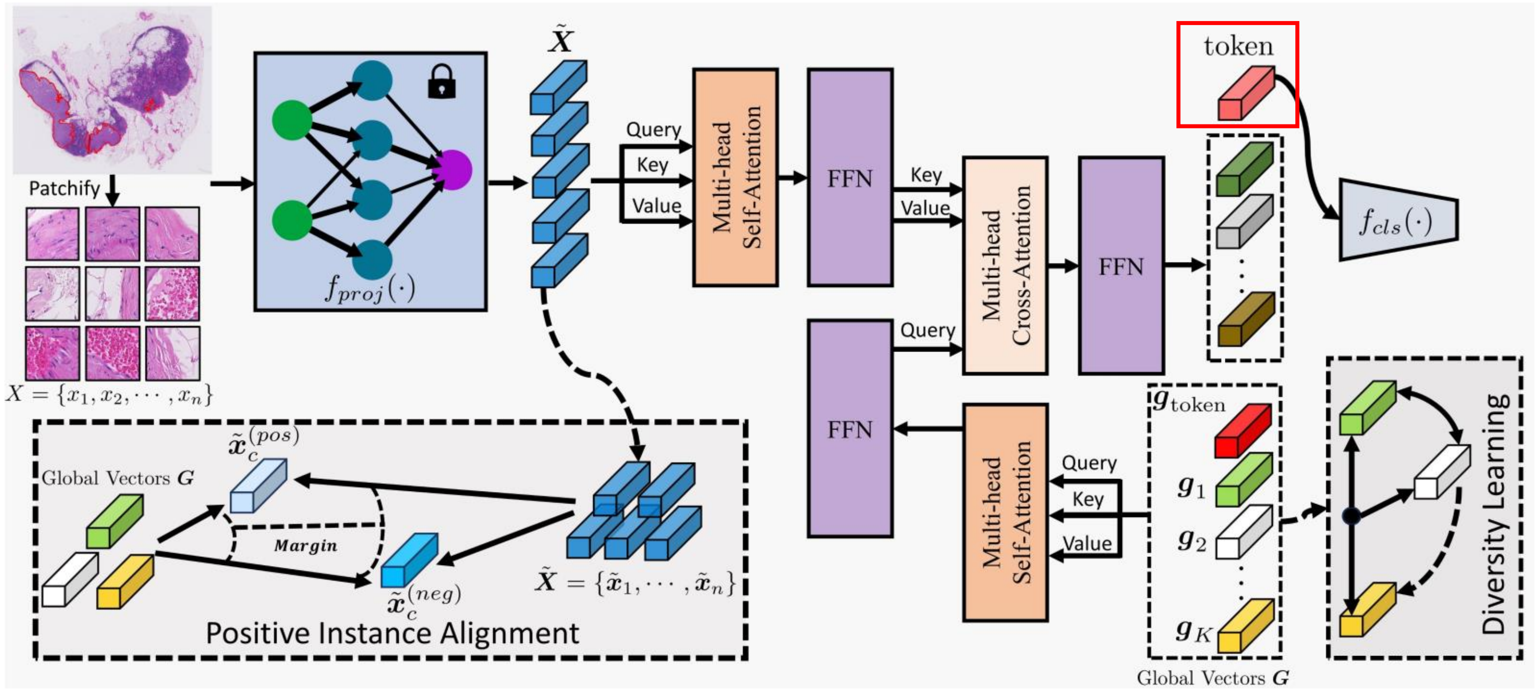
$$\boldsymbol{L} = \boldsymbol{G}\boldsymbol{G}^\top \in \mathbb{R}^{K\times K}$$

$$\mathcal{L}_{div} = -\log\det(\boldsymbol{G}\boldsymbol{G}^\top), \quad \text{s.t. } \|\boldsymbol{g}_i\| = 1 = C.$$



Theorem 1. *Given a set of global vectors* $\boldsymbol{G} = [\boldsymbol{g}_1^\top, \cdots, \boldsymbol{g}_K^\top]$ *with* $\|\boldsymbol{g}_i\| = C, \forall i \in [K]$, *maximizing the DPP-based diversity (i.e.* $\max\det(\boldsymbol{G}\boldsymbol{G}^\top)$*) results in orthogonal global vectors with* $\boldsymbol{g}_i \perp \boldsymbol{g}_j$, $\forall i \neq j, i, j \in [K]$.

Proposed Method



- **We suggest using global learnable vectors to help network to learning diversity. The global vector will gather similar instance together by cross attention.**
- **Two mechanisms to learn a reliable and diverse global vector:**
  - Positive instance alignment
  - DPP diversity loss (theoretical guaranteed).
- **A class token to summarize all global vectors for final bag-level classification.**

Objective Function

$$\mathcal{L}_{final} = \boxed{\mathcal{L}_{ce}} + \boxed{\lambda_{tri}\mathcal{L}_{tri}} + \boxed{\lambda_{div}\mathcal{L}_{div}},$$

Cross-entropy for bag-level classification

Positive instance alignment

Diversity loss

Experimental Results

Outperforms all recent SOTAs!

	CAMELYON16			TCGA-NSCLC		
	Accuracy	F1	AUC	Accuracy	F1	AUC
ResNet-50 ImageNet Pretrained						
Classic AB-MIL (*ICML'18*)	$0.845_{(0.839,0.851)}$	$0.780_{(0.769,0.791)}$	$0.854_{(0.848,0.860)}$	$0.869_{0.032}$	$0.866_{0.021}$	$0.941_{0.028}$
DS-MIL (*CVPR'21*)	$0.856_{(0.843,0.869)}$	$0.815_{(0.797,0.832)}$	$0.899_{(0.890,0.908)}$	$0.888_{0.013}$	$0.876_{0.011}$	$0.939_{0.019}$
CLAM-SB (*Nature Bio. Eng.'21*)	$0.837_{(0.809,0.865)}$	$0.775_{(0.755,0.795)}$	$0.871_{(0.856,0.885)}$	$0.875_{0.041}$	$0.864_{0.043}$	$0.944_{0.023}$
CLAM-MB (*Nature Bio. Eng.'21*)	$0.823_{(0.795,0.850)}$	$0.774_{(0.752,0.795)}$	$0.878_{(0.861,0.894)}$	$0.878_{0.043}$	$0.874_{0.028}$	$0.949_{0.019}$
PMIL (*MedIA'23*)	$0.831_{(0.799,0.863)}$	$0.816_{(0.779,0.853)}$	$0.845_{(0.813,0.876)}$	$0.873_{0.010}$	$0.875_{0.011}$	$0.933_{0.007}$
Trans-MIL (*NeurIPS'21*)	$0.858_{(0.848,0.868)}$	$0.797_{(0.776,0.818)}$	$0.906_{(0.875,0.937)}$	$0.883_{0.022}$	$0.876_{0.021}$	$0.949_{0.013}$
DTFD-MIL (MaxS) (*CVPR'22*)	$0.864_{(0.848,0.880)}$	$0.814_{(0.802,0.826)}$	$0.907_{(0.894,0.919)}$	$0.868_{0.040}$	$0.863_{0.029}$	$0.919_{0.037}$
DTFD-MIL (MaxMinS) (*CVPR'22*)	$0.899_{(0.887,0.912)}$	$0.865_{(0.848,0.882)}$	$0.941_{(0.936,0.944)}$	$0.894_{0.033}$	$0.891_{0.027}$	$0.961_{0.021}$
DTFD-MIL (AFS) (*CVPR'22*)	$0.908_{(0.892,0.925)}$	$0.882_{(0.861,0.903)}$	$0.946_{(0.941,0.951)}$	$0.891_{0.033}$	$0.883_{0.025}$	$0.951_{0.022}$
ILRA-MIL (*ICLR'23*)	$0.848_{(0.844,0.853)}$	$0.826_{(0.823,0.829)}$	$0.868_{(0.852,0.883)}$	$0.895_{0.017}$	$0.896_{0.017}$	$0.946_{0.014}$
Our	$\mathbf{0.917}_{(0.902,0.931)}$	$\mathbf{0.913}_{(0.898,0.928)}$	$\mathbf{0.957}_{(0.951,0.963)}$	$\mathbf{0.908}_{0.015}$	$\mathbf{0.911}_{0.018}$	$\mathbf{0.963}_{0.008}$
ResNet-18 ImageNet Pretrained						
Classic AB-MIL (*ICML'18*)	$0.805_{(0.772,0.837)}$	$0.786_{(0.757,0.815)}$	$0.843_{(0.827,0.858)}$	$0.874_{0.005}$	$0.873_{0.006}$	$0.937_{0.001}$
DS-MIL (*CVPR'21*)	$0.791_{(0.739,0.843)}$	$0.776_{(0.712,0.840)}$	$0.814_{(0.754,0.875)}$	$0.831_{0.012}$	$0.838_{0.008}$	$0.896_{0.009}$
CLAM-SB (*Nature Bio. Eng.'21*)	$0.792_{(0.769,0.815)}$	$0.766_{(0.746,0.786)}$	$0.811_{(0.777,0.845)}$	$0.869_{0.010}$	$0.869_{0.010}$	$0.931_{0.006}$
CLAM-MB (*Nature Bio. Eng.'21*)	$0.786_{(0.754,0.818)}$	$0.770_{(0.746,0.795)}$	$0.825_{(0.808,0.843)}$	$0.880_{0.016}$	$0.880_{0.016}$	$0.944_{0.012}$
PMIL (*MedIA'23*)	$0.800_{(0.775,0.825)}$	$0.784_{(0.765,0.804)}$	$0.829_{(0.807,0.851)}$	$0.856_{0.006}$	$0.862_{0.003}$	$0.933_{0.010}$
Trans-MIL (*NeurIPS'21*)	$0.839_{(0.822,0.856)}$	$0.827_{(0.805,0.848)}$	$0.854_{(0.823,0.886)}$	$0.877_{0.009}$	$0.879_{0.008}$	$0.938_{0.014}$
DTFD-MIL (MaxS) (*CVPR'22*)	$0.856_{(0.824,0.887)}$	$0.792_{(0.742,0.842)}$	$0.878_{(0.862,0.893)}$	$0.830_{0.014}$	$0.821_{0.020}$	$0.893_{0.015}$
DTFD-MIL (MaxMinS) (*CVPR'22*)	$0.833_{(0.807,0.858)}$	$0.768_{(0.747,0.788)}$	$0.878_{(0.872,0.883)}$	$0.853_{0.012}$	$0.850_{0.021}$	$0.925_{0.013}$
DTFD-MIL (AFS) (*CVPR'22*)	$0.817_{(0.791,0.843)}$	$0.734_{(0.687,0.781)}$	$0.868_{(0.841,0.896)}$	$0.870_{0.007}$	$0.864_{0.012}$	$0.935_{0.010}$
ILRA-MIL (*ICLR'23*)	$0.831_{(0.768,0.895)}$	$0.819_{(0.768,0.871)}$	$0.852_{(0.811,0.893)}$	$0.878_{0.002}$	$0.879_{0.001}$	$0.937_{0.004}$
Our	$\mathbf{0.873}_{(0.862,0.884)}$	$\mathbf{0.862}_{(0.852,0.871)}$	$\mathbf{0.898}_{(0.886,0.909)}$	$\mathbf{0.891}_{0.029}$	$\mathbf{0.890}_{0.021}$	$\mathbf{0.955}_{0.023}$
Vision Transformer ImageNet Pretrained						
Classic AB-MIL (*ICML'18*)	$0.851_{(0.837,0.865)}$	$0.835_{(0.810,0.860)}$	$0.873_{(0.840,0.906)}$	$0.904_{0.011}$	$0.904_{0.010}$	$0.953_{0.013}$
DS-MIL (*CVPR'21*)	$0.810_{(0.741,0.879)}$	$0.806_{(0.742,0.869)}$	$0.871_{(0.836,0.906)}$	$0.875_{0.020}$	$0.879_{0.016}$	$0.933_{0.016}$
CLAM-SB (*Nature Bio. Eng.'21*)	$0.839_{(0.831,0.847)}$	$0.816_{(0.799,0.834)}$	$0.864_{(0.841,0.887)}$	$0.907_{0.008}$	$0.907_{0.001}$	$0.954_{0.014}$
CLAM-MB (*Nature Bio. Eng.'21*)	$0.826_{(0.806,0.846)}$	$0.804_{(0.795,0.813)}$	$0.851_{(0.825,0.878)}$	$0.911_{0.007}$	$0.911_{0.007}$	$0.959_{0.008}$
PMIL (*MedIA'23*)	$0.843_{(0.831,0.856)}$	$0.826_{(0.814,0.838)}$	$0.843_{(0.820,0.867)}$	$0.882_{0.009}$	$0.884_{0.006}$	$0.940_{0.006}$
Trans-MIL (*NeurIPS'21*)	$0.862_{(0.841,0.883)}$	$0.846_{(0.823,0.869)}$	$0.860_{(0.848,0.873)}$	$0.909_{0.009}$	$0.909_{0.009}$	$0.953_{0.006}$
DTFD-MIL (MaxS) (*CVPR'22*)	$0.846_{(0.832,0.860)}$	$0.767_{(0.746,0.787)}$	$0.859_{(0.842,0.876)}$	$0.904_{0.011}$	$0.904_{0.010}$	$0.953_{0.013}$
DTFD-MIL (MaxMinS) (*CVPR'22*)	$0.839_{(0.826,0.851)}$	$0.752_{(0.742,0.763)}$	$0.862_{(0.836,0.888)}$	$0.895_{0.013}$	$0.892_{0.016}$	$0.952_{0.011}$
DTFD-MIL (AFS) (*CVPR'22*)	$0.831_{(0.818,0.844)}$	$0.759_{(0.737,0.781)}$	$0.880_{(0.864,0.897)}$	$0.901_{0.005}$	$0.900_{0.008}$	$0.959_{0.012}$
ILRA-MIL (*ICLR'23*)	$0.850_{(0.825,0.875)}$	$0.838_{(0.812,0.865)}$	$0.864_{(0.843,0.885)}$	$0.902_{0.007}$	$0.904_{0.007}$	$0.954_{0.006}$
Our	$\mathbf{0.893}_{(0.889,0.897)}$	$\mathbf{0.882}_{(0.877,0.886)}$	$\mathbf{0.891}_{(0.884,0.899)}$	$\mathbf{0.926}_{0.008}$	$\mathbf{0.925}_{0.008}$	$\mathbf{0.969}_{0.004}$

Visualization

Different global vectors focus on
different regions of interest.

Detected
tumor area

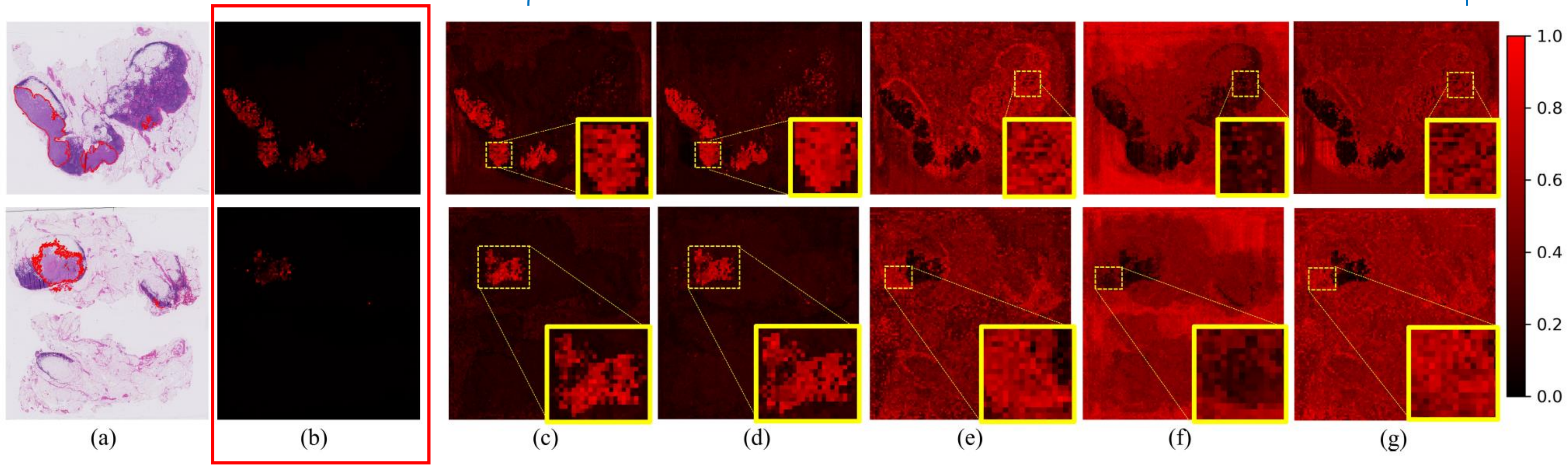


Fig. 5: Visualization of the attention map: (a) raw WSI with the ground-truth annotation, (b) the attention map computes using the tokenized global vectors, and (c-g) the attention map computes using the other $(K-1)$ global vectors with $K=6$ in our experiment.

Thanks for Watching!