哈爾濱工業大學 HARBIN INSTITUTE OF TECHNOLOGY

ICCV OCT 19-23, 2025 HONOLULU HAWAII

Vilab Visual Intelligence Lab

https://vilab.hit.edu.cn/

GMMamba: Group Masking Mamba for Whole Slide Image Classification

Tingting Zheng[1] Hongxun Yao[1,*] Kui Jiang[1] Yi Xiao[2] Sicheng Zhao[3]

1Harbin Institute of Technology

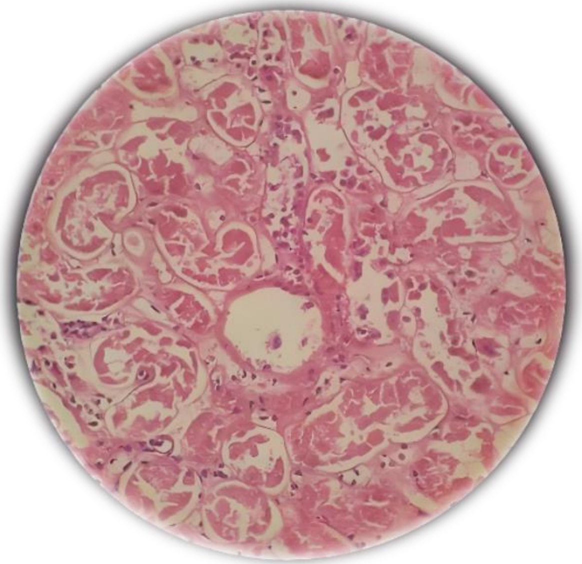
2Wuhan University 3Tsinghua University

ICCV 2025 (*Oral Presentation*)

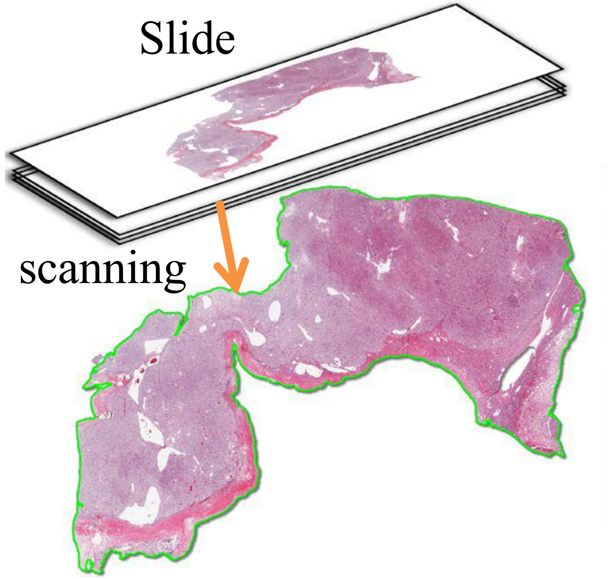
https://github.com/titizheng/GMMamba

Content

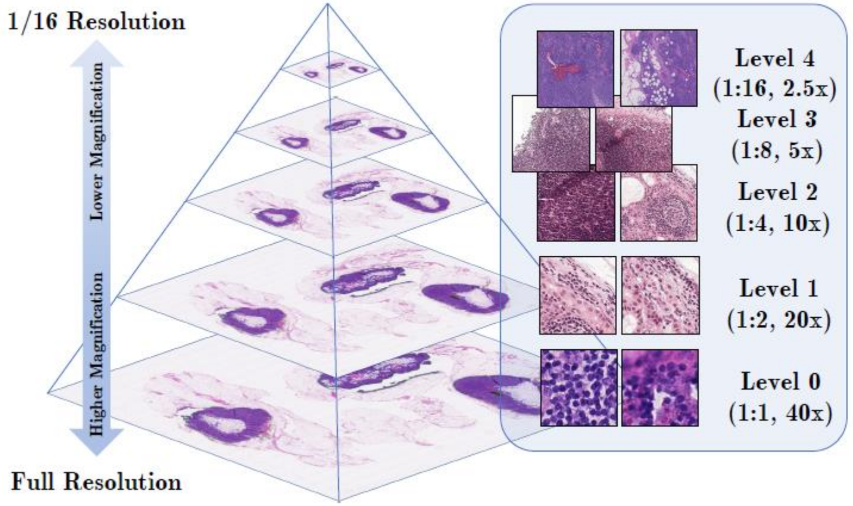
Introduction



Histopathology images

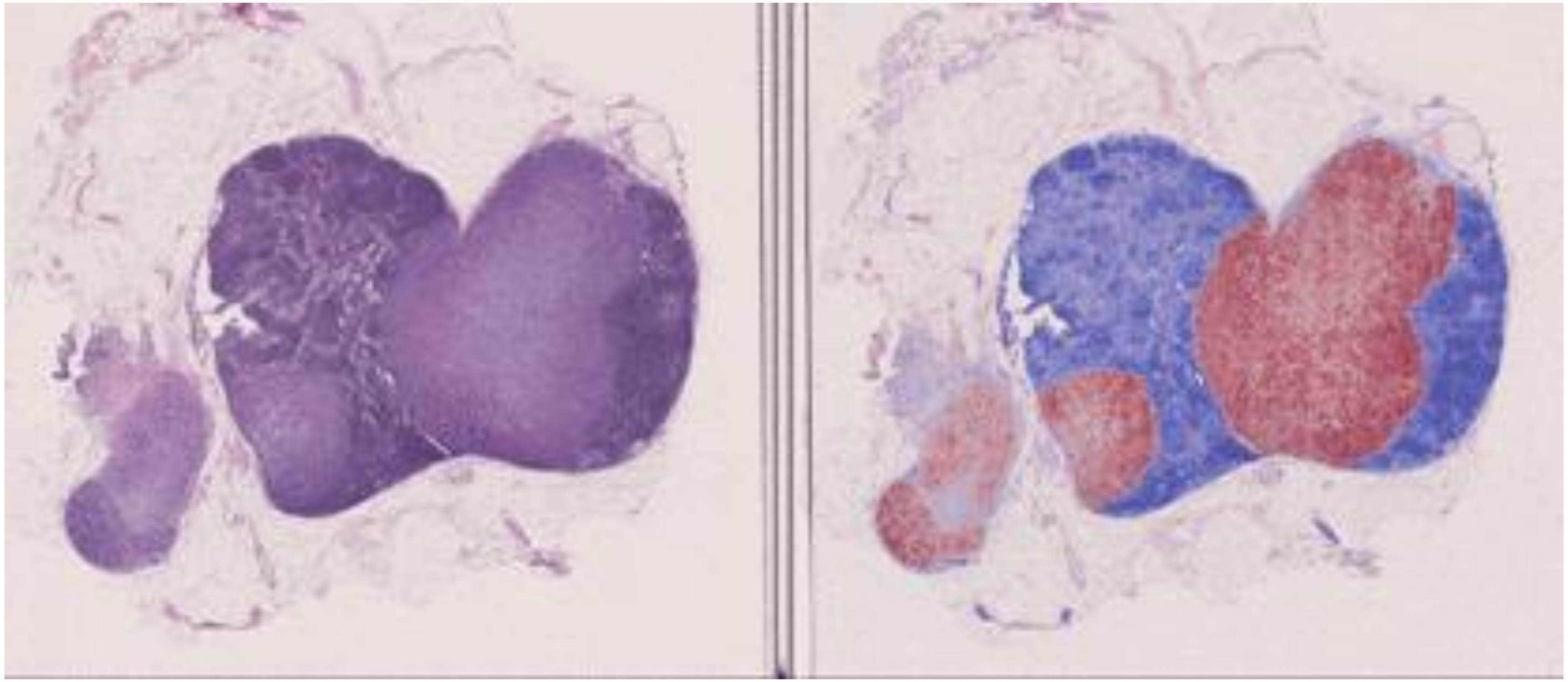


Whole Slide Image (WSI)



Multi-level high-resolution images

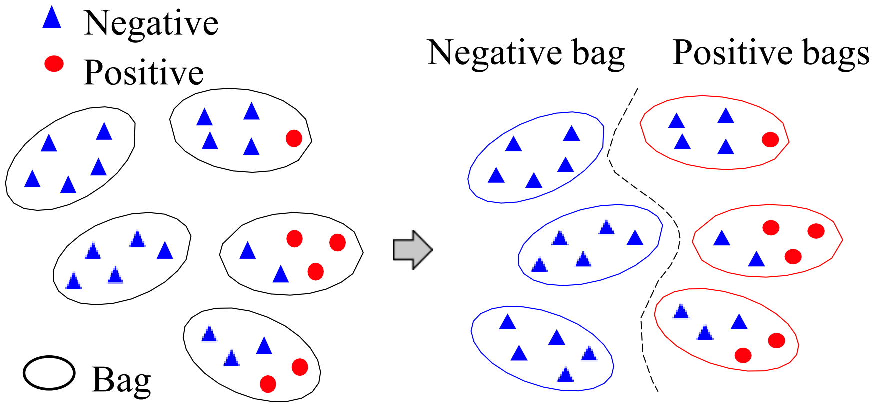
Introduction



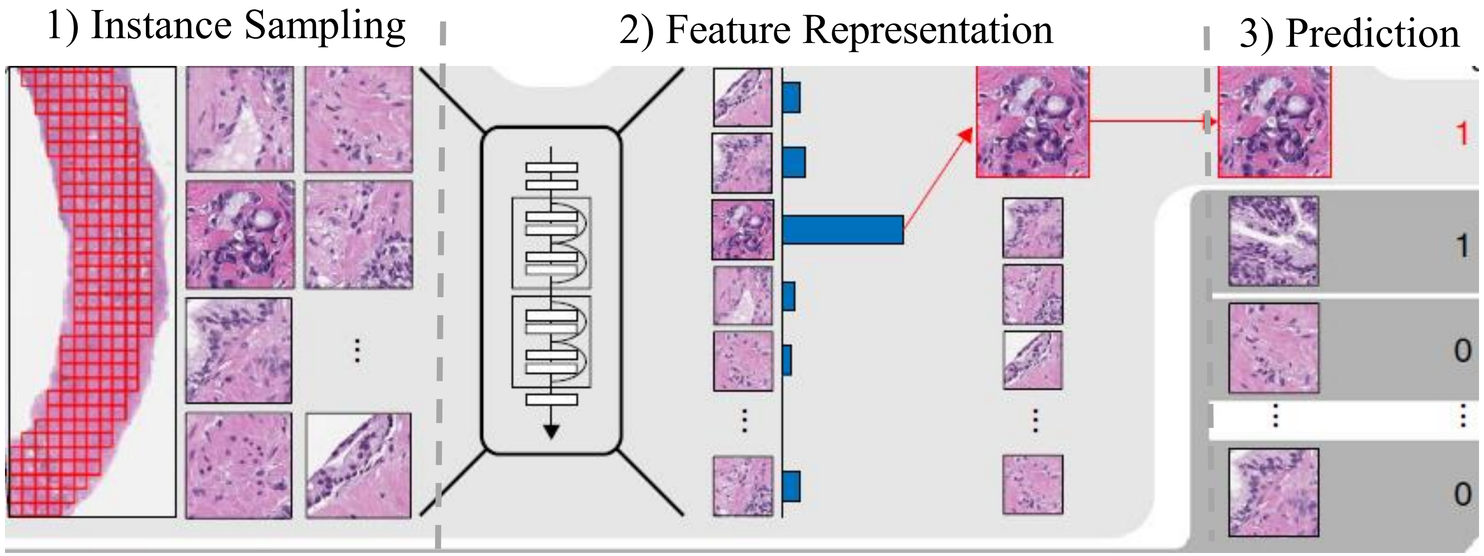
CLAM (Lu et al Nat. Biomed. Eng’2021)

How to extend the supervised methods to the weakly supervised methods for efficient and accurate classification of WSIs ?

Introduction



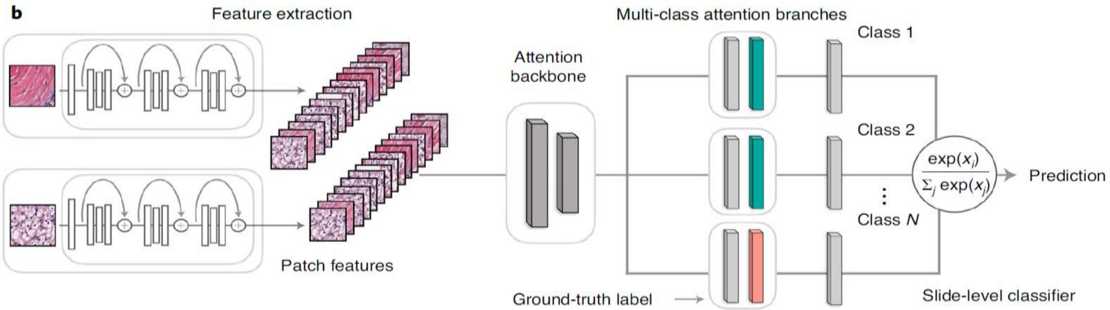
DAMIL (Hashimoto et al CVPR’2020)



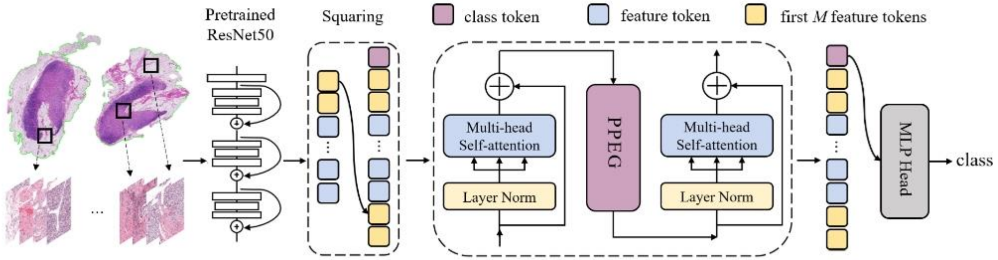
RNNMIL (Campanella et al Nature Medicite’2019)

Multiple instance learning (MIL) techniques aggregate disease-relevant information from label-free instances for bag label Prediction.

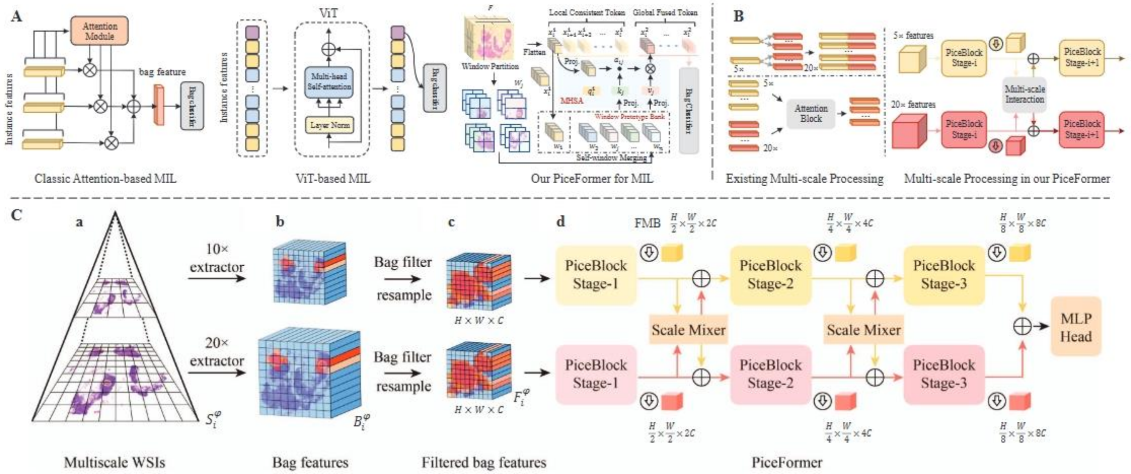
Introduction



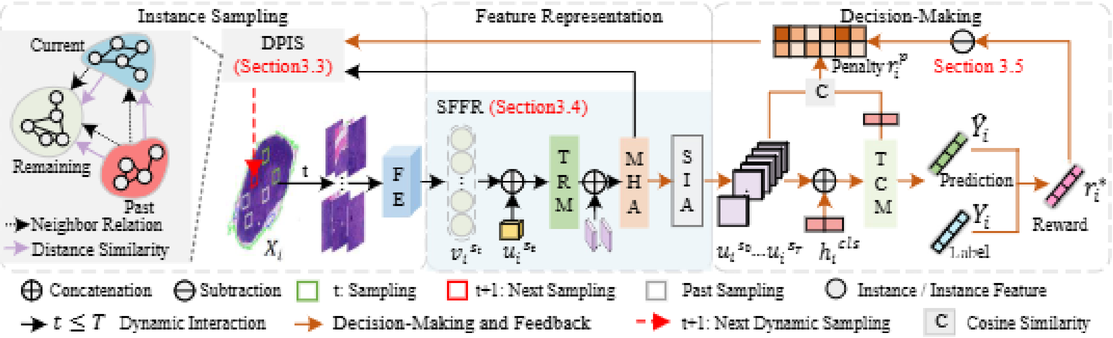
Attention-based CLAM (Nat. Biomed. Eng'2021)



Transformer-based TransMIL (NIPS'2021)



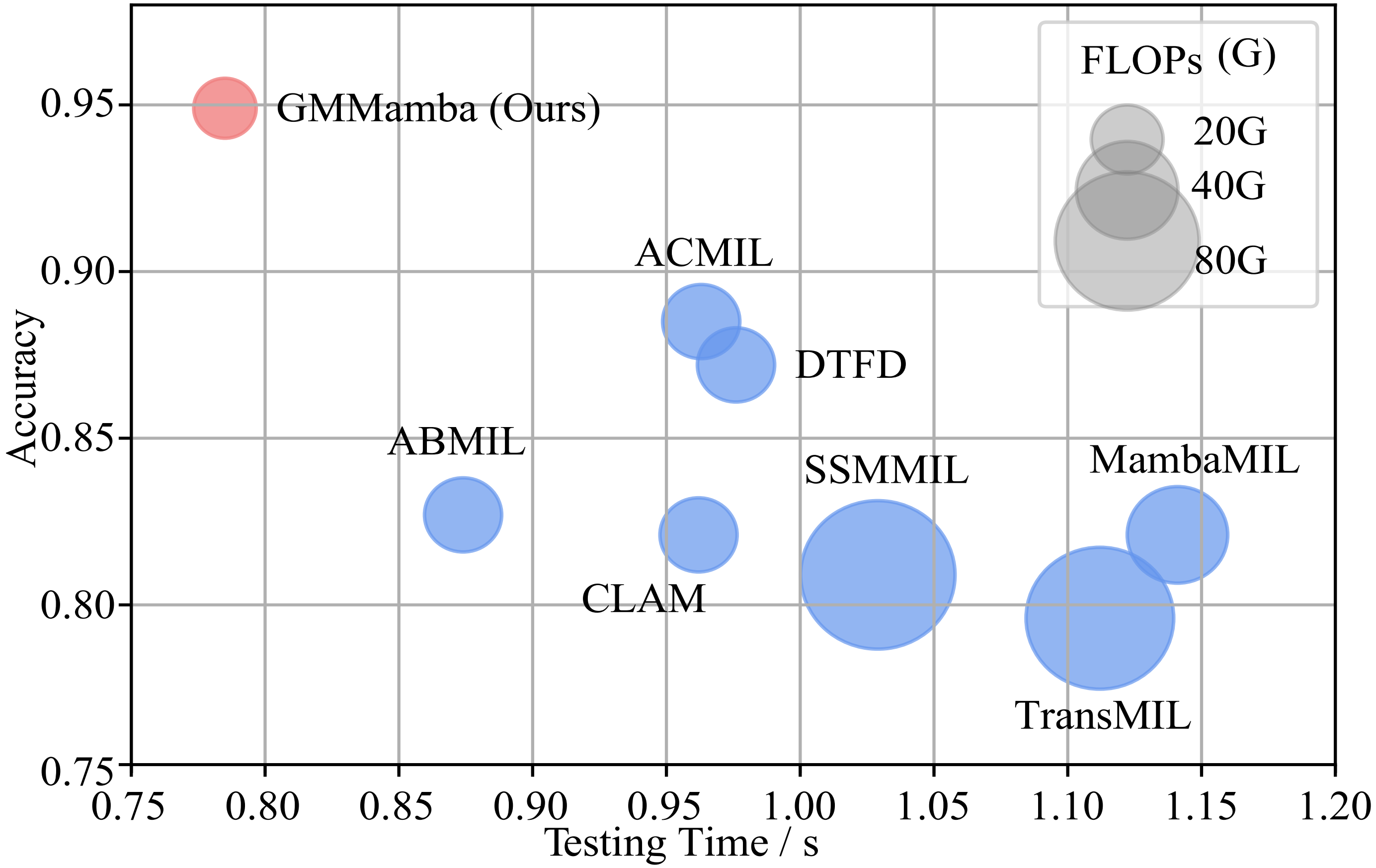
Transformer-based MILBooster (ICCV'2023)



Transformer-based PAMIL (CVPR'2024)

Prior work leverages instance correlations to build global representations for bag prediction.

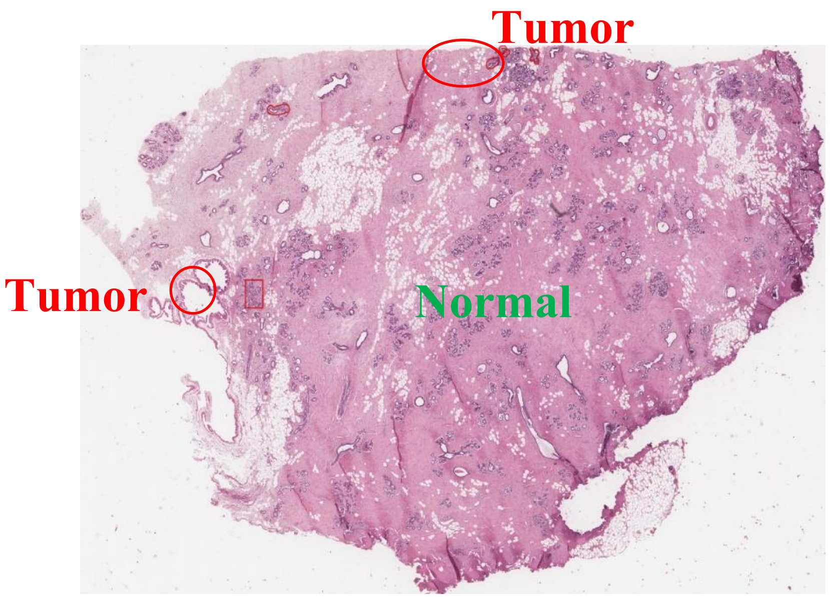
Introduction

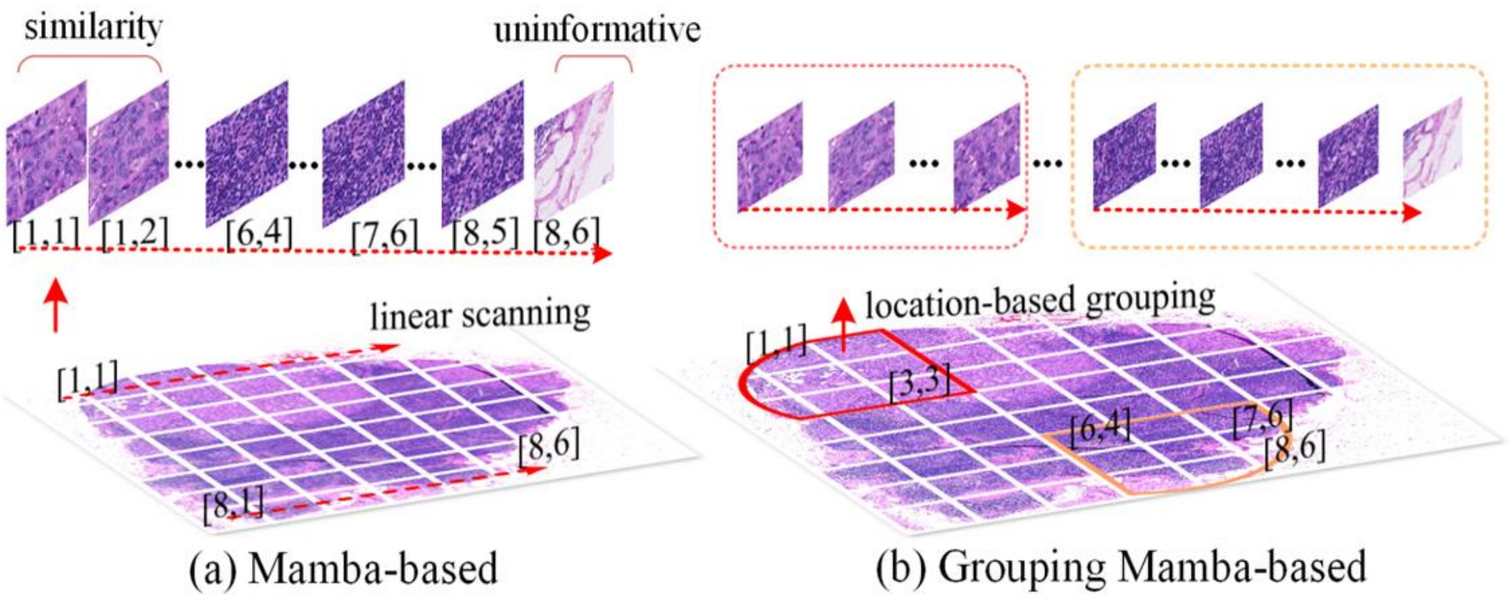


Comparative results of FLOPs, testing time, and accuracy with representative MIL methods. The size of each represents the FLOPs.

Attention- and Transformer-based MIL are limited by uninformative instances and heavy computation, leading to suboptimal representations and prolonged inference.

Introduction





(a) Mamba-based

(b) Grouping Mamba-based

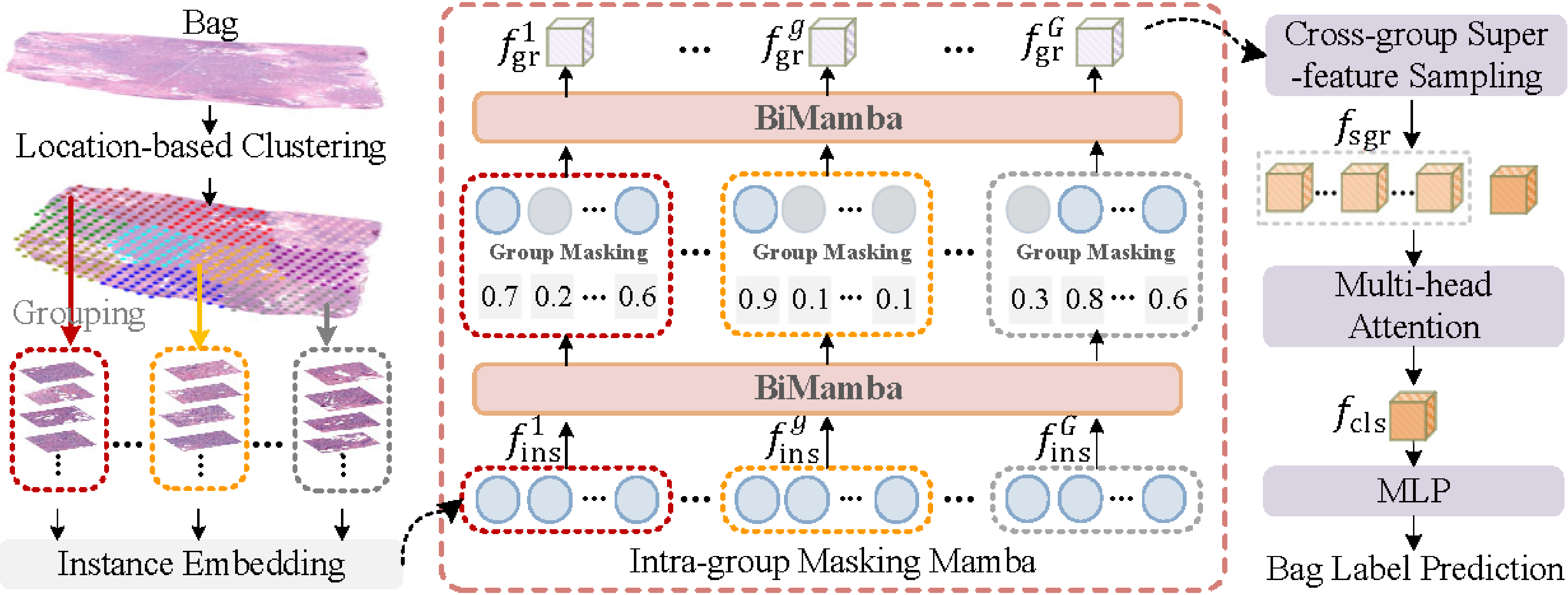
Limitations:

- **Redundant Local Modeling: Mamba processes all instances uniformly, causing overhead from redundant features and loss of critical diagnostic cues.**
- **Sparse Global Representation: Tumor regions are spatially dispersed and sparse, and existing methods fail to model inter-group correlations, leading to weak global representations.**

Content

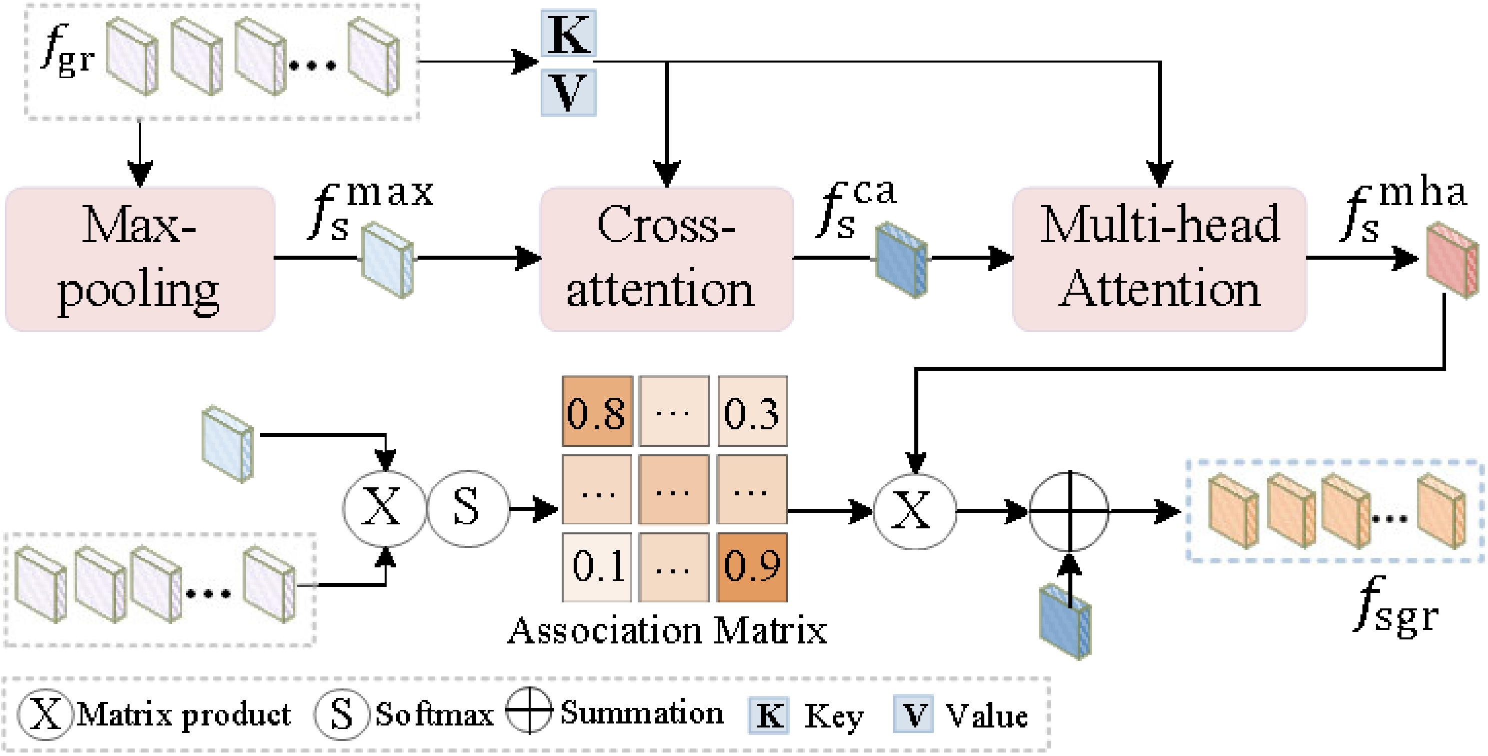
Method

Group Masking Mamba (GMMamba): Reduce redundant interference and enhance inter-bag interactions to improve representation quality and prediction performance.



Method

Cross-group Super-feature Sampling (CSS) Module：capture dispersed tumor information across groups for comprehensive and discriminative representations



Content

Experiments

Evaluation methods:

We compare our proposed GMMamba against representative MIL methods, including：

- **Attention-based:** ABMIL (Ilse, ICML’18), CLAM (Lu, Nat. Biomed. Eng’21) , DSMIL (Li, CVPR’21), DTFD (Zhang, CVPR’22), MHIM-ABMIL (Tang, ICCV’23), IBMIL-ABMIL (Zhang, Lin’23), ILRA-MIL (Xiang, ICIL’23), ACMIL (Zhang, ECCV’25);
- **Transformer-based:** TransMIL (Shao, CVPR’21), MHIM-TransMIL (Tang, ICCV’23) ;
- **Mamba-based:** SSMMIL (Leo, MICCAI’23), MambaMIL (Yang, MICCAI’24) .

Evaluation metrics:

- Aunder the receiver operating characteristic curve (AUC), accuracy, and F1 score (F1)
- 5-fold Training-validation-Testing

Experiments-Datasets

TCGA-BRCA(Breast Cancer)

- 952 WSIs (749 IDC, 203 ILC), split 65/10/25.
- ~3.1M patches at 10× magnification.

TCGA-ESCA(Esophageal Cancer)

- 156 WSIs (90 SCC, 66 ADC), split 3/1/1.
- ~0.5M patches at 20× magnification.

BRACS(Breast Carcinoma Subtyping)

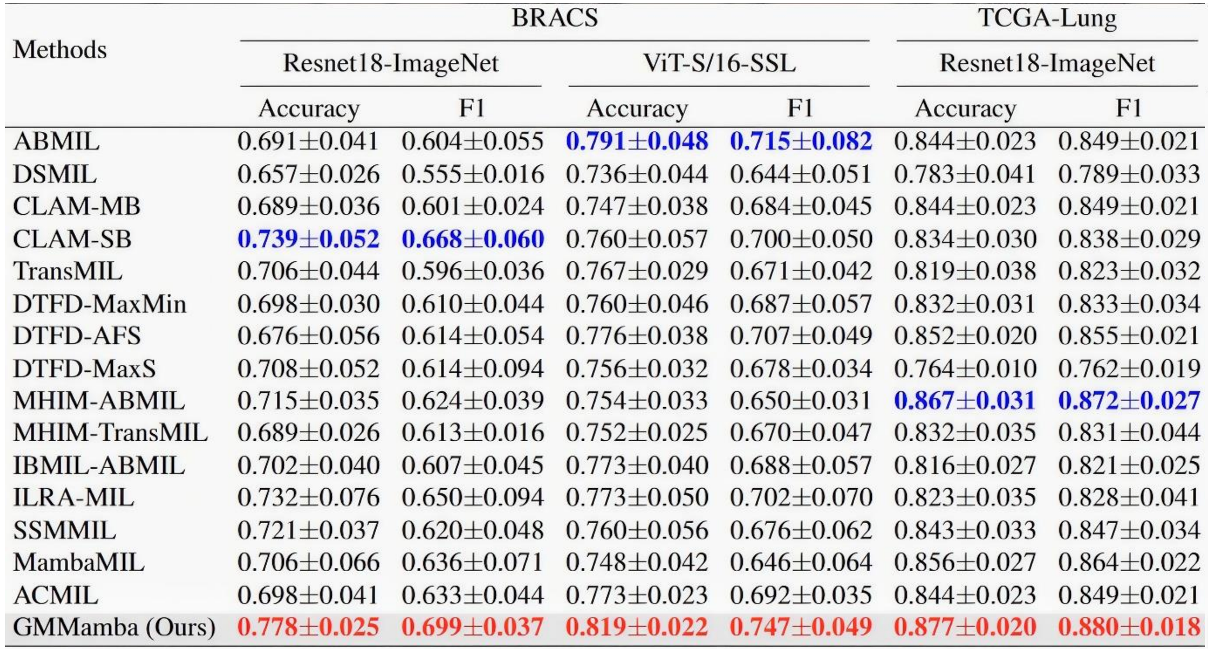
- 547 WSIs (265 benign, 89 atypical, 193 malignant).
- ~1.4M patches at 10× magnification.

TCGA-Lung(Lung Cancer)

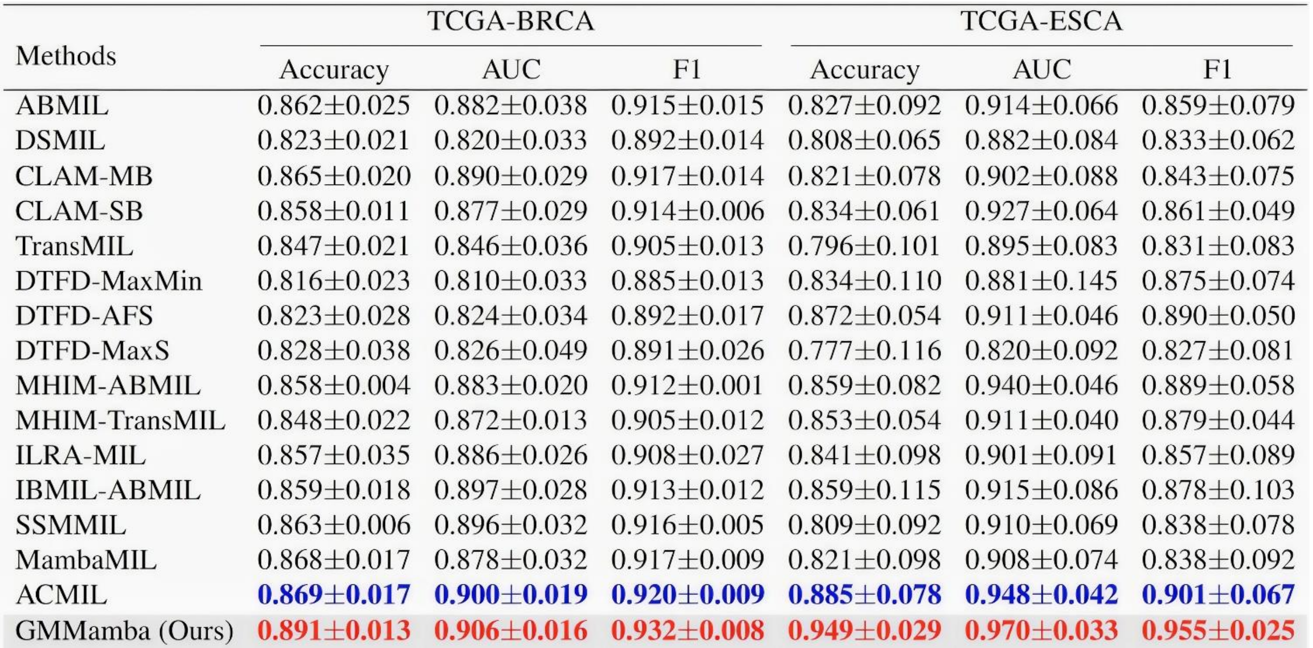
- 1053 WSIs (541 LUAD, 512 LUSC), split 65/10/25.
- ~4.1M patches at 20× magnification.

Experiments-Compare with SOTA Methods

GMMamba consistently outperforms baseline methods across multiple WSI datasets by effectively reducing redundancy and enhancing inter-group interactions.



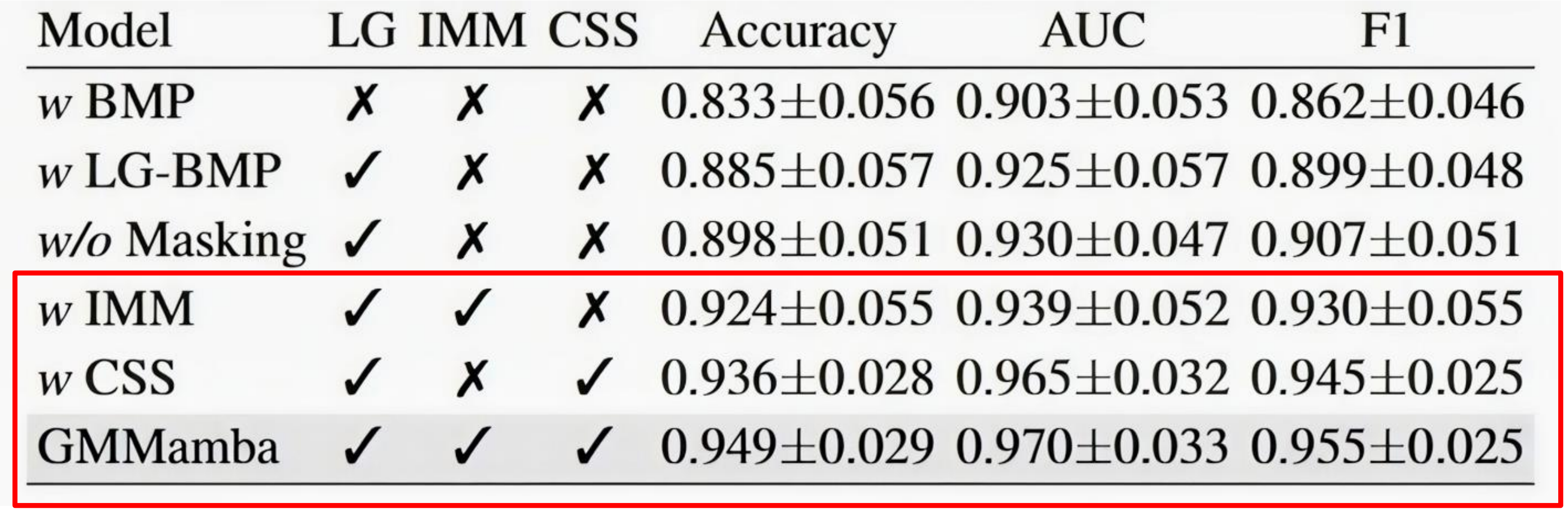
Methods	BRACS				TCGA-Lung	
	Resnet18-ImageNet		ViT-S/16-SSL		Resnet18-ImageNet	
	Accuracy	F1	Accuracy	F1	Accuracy	F1
ABMIL	0.691±0.041	0.604±0.055	**0.791±0.048**	**0.715±0.082**	0.844±0.023	0.849±0.021
DSMIL	0.657±0.026	0.555±0.016	0.736±0.044	0.644±0.051	0.783±0.041	0.789±0.033
CLAM-MB	0.689±0.036	0.601±0.024	0.747±0.038	0.684±0.045	0.844±0.023	0.849±0.021
CLAM-SB	**0.739±0.052**	**0.668±0.060**	0.760±0.057	0.700±0.050	0.834±0.030	0.838±0.029
TransMIL	0.706±0.044	0.596±0.036	0.767±0.029	0.671±0.042	0.819±0.038	0.823±0.032
DTFD-MaxMin	0.698±0.030	0.610±0.044	0.760±0.046	0.687±0.057	0.832±0.031	0.833±0.034
DTFD-AFS	0.676±0.056	0.614±0.054	0.776±0.038	0.707±0.049	0.852±0.020	0.855±0.021
DTFD-MaxS	0.708±0.052	0.614±0.094	0.756±0.032	0.678±0.034	0.764±0.010	0.762±0.019
MHIM-ABMIL	0.715±0.035	0.624±0.039	0.754±0.033	0.650±0.031	**0.867±0.031**	**0.872±0.027**
MHIM-TransMIL	0.689±0.026	0.613±0.016	0.752±0.025	0.670±0.047	0.832±0.035	0.831±0.044
IBMIL-ABMIL	0.702±0.040	0.607±0.045	0.773±0.040	0.688±0.057	0.816±0.027	0.821±0.025
ILRA-MIL	0.732±0.076	0.650±0.094	0.773±0.050	0.702±0.070	0.823±0.035	0.828±0.041
SSMMIL	0.721±0.037	0.620±0.048	0.760±0.056	0.676±0.062	0.843±0.033	0.847±0.034
MambaMIL	0.706±0.066	0.636±0.071	0.748±0.042	0.646±0.064	0.856±0.027	0.864±0.022
ACMIL	0.698±0.041	0.633±0.044	0.773±0.023	0.692±0.035	0.844±0.023	0.849±0.021
GMMamba (Ours)	**0.778±0.025**	**0.699±0.037**	**0.819±0.022**	**0.747±0.049**	**0.877±0.020**	**0.880±0.018**



Methods	TCGA-BRCA			TCGA-ESCA		
	Accuracy	AUC	F1	Accuracy	AUC	F1
ABMIL	0.862±0.025	0.882±0.038	0.915±0.015	0.827±0.092	0.914±0.066	0.859±0.079
DSMIL	0.823±0.021	0.820±0.033	0.892±0.014	0.808±0.065	0.882±0.084	0.833±0.062
CLAM-MB	0.865±0.020	0.890±0.029	0.917±0.014	0.821±0.078	0.902±0.088	0.843±0.075
CLAM-SB	0.858±0.011	0.877±0.029	0.914±0.006	0.834±0.061	0.927±0.064	0.861±0.049
TransMIL	0.847±0.021	0.846±0.036	0.905±0.013	0.796±0.101	0.895±0.083	0.831±0.083
DTFD-MaxMin	0.816±0.023	0.810±0.033	0.885±0.013	0.834±0.110	0.881±0.145	0.875±0.074
DTFD-AFS	0.823±0.028	0.824±0.034	0.892±0.017	0.872±0.054	0.911±0.046	0.890±0.050
DTFD-MaxS	0.828±0.038	0.826±0.049	0.891±0.026	0.777±0.116	0.820±0.092	0.827±0.081
MHIM-ABMIL	0.858±0.004	0.883±0.020	0.912±0.001	0.859±0.082	0.940±0.046	0.889±0.058
MHIM-TransMIL	0.848±0.022	0.872±0.013	0.905±0.012	0.853±0.054	0.911±0.040	0.879±0.044
ILRA-MIL	0.857±0.035	0.886±0.026	0.908±0.027	0.841±0.098	0.901±0.091	0.857±0.089
IBMIL-ABMIL	0.859±0.018	0.897±0.028	0.913±0.012	0.859±0.115	0.915±0.086	0.878±0.103
SSMMIL	0.863±0.006	0.896±0.032	0.916±0.005	0.809±0.092	0.910±0.069	0.838±0.078
MambaMIL	0.868±0.017	0.878±0.032	0.917±0.009	0.821±0.098	0.908±0.074	0.838±0.092
ACMIL	**0.869±0.017**	**0.900±0.019**	**0.920±0.009**	**0.885±0.078**	**0.948±0.042**	**0.901±0.067**
GMMamba (Ours)	**0.891±0.013**	**0.906±0.016**	**0.932±0.008**	**0.949±0.029**	**0.970±0.033**	**0.955±0.025**

Ablation Studies-Validation on Basic Components

GMMamba improve bag-level representations by aggregating dispersed tumor information and mitigating redundancy, boosting accuracy, F1, and AUC.



Model	LG	IMM	CSS	Accuracy	AUC	F1
w BMP	✗	✗	✗	0.833±0.056	0.903±0.053	0.862±0.046
w LG-BMP	✓	✗	✗	0.885±0.057	0.925±0.057	0.899±0.048
w/o Masking	✓	✗	✗	0.898±0.051	0.930±0.047	0.907±0.051
w IMM	✓	✓	✗	0.924±0.055	0.939±0.052	0.930±0.055
w CSS	✓	✗	✓	0.936±0.028	0.965±0.032	0.945±0.025
GMMamba	✓	✓	✓	0.949±0.029	0.970±0.033	0.955±0.025

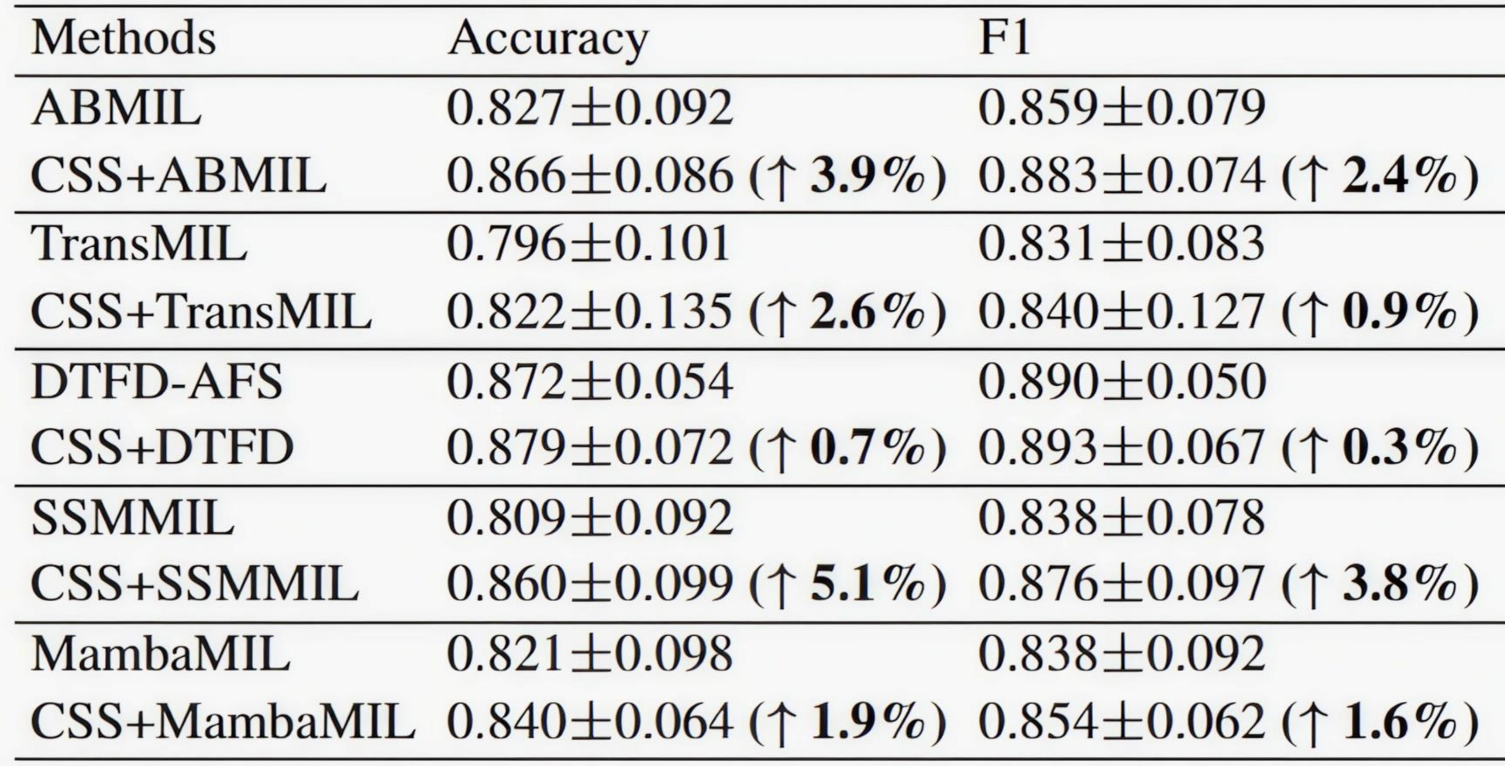
Location-based Grouping (**LG**)

BiMamba with Max-Pooling (**BMP**)

Intra-group Masking Mamba (**IMM**)

Cross-group Super-feature Sampling (**CSS**)

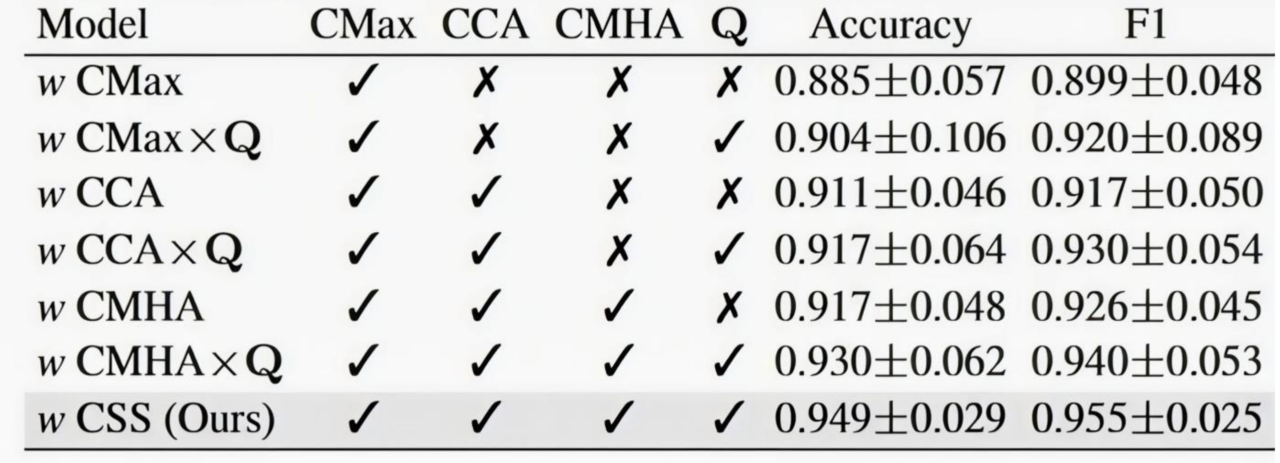
Experiments-Compare with SOTA Methods



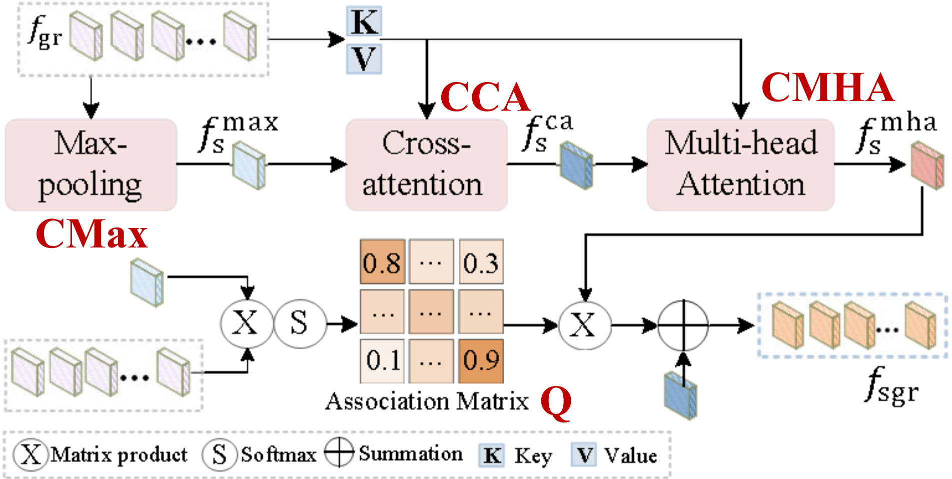
Methods	Accuracy	F1
ABMIL	0.827±0.092	0.859±0.079
CSS+ABMIL	0.866±0.086 (↑ **3.9%**)	0.883±0.074 (↑ **2.4%**)
TransMIL	0.796±0.101	0.831±0.083
CSS+TransMIL	0.822±0.135 (↑ **2.6%**)	0.840±0.127 (↑ **0.9%**)
DTFD-AFS	0.872±0.054	0.890±0.050
CSS+DTFD	0.879±0.072 (↑ **0.7%**)	0.893±0.067 (↑ **0.3%**)
SSMMIL	0.809±0.092	0.838±0.078
CSS+SSMMIL	0.860±0.099 (↑ **5.1%**)	0.876±0.097 (↑ **3.8%**)
MambaMIL	0.821±0.098	0.838±0.092
CSS+MambaMIL	0.840±0.064 (↑ **1.9%**)	0.854±0.062 (↑ **1.6%**)

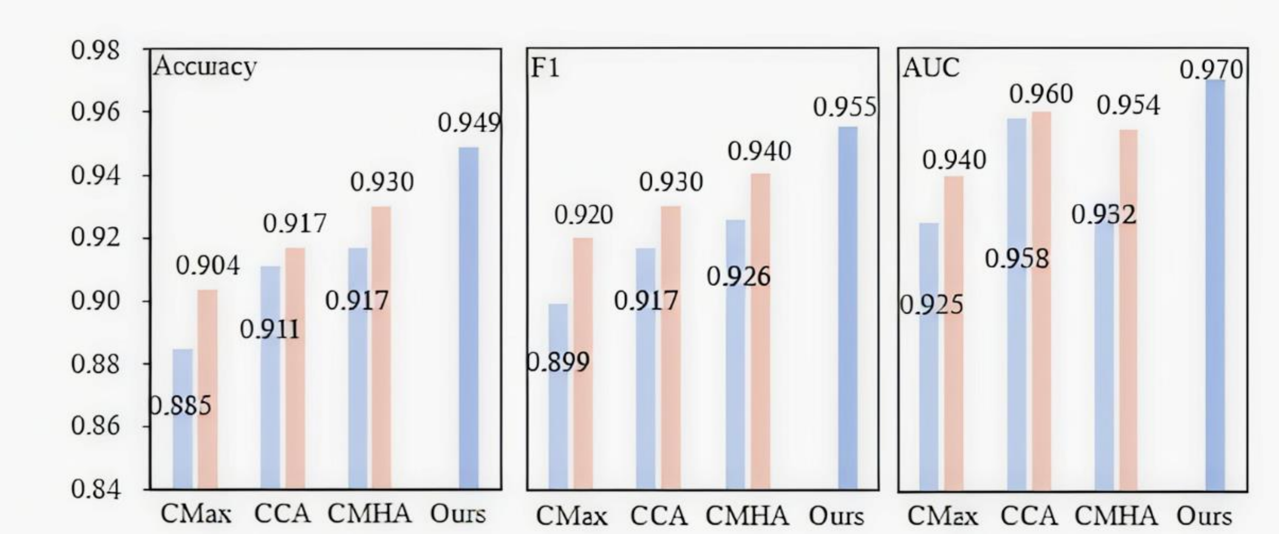
CSS enhances bag-level methods by effectively exploring instance relationships and aggregating dispersed tumor features.

Ablation Studies-CSS Variants



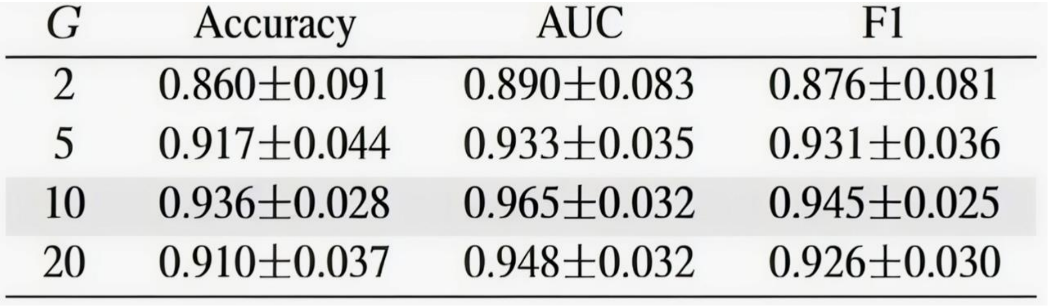
Model	CMax	CCA	CMHA	Q	Accuracy	F1
w CMax	✓	✗	✗	✗	0.885±0.057	0.899±0.048
w CMax×Q	✓	✗	✗	✓	0.904±0.106	0.920±0.089
w CCA	✓	✓	✗	✗	0.911±0.046	0.917±0.050
w CCA×Q	✓	✓	✗	✓	0.917±0.064	0.930±0.054
w CMHA	✓	✓	✓	✗	0.917±0.048	0.926±0.045
w CMHA×Q	✓	✓	✓	✓	0.930±0.062	0.940±0.053
w CSS (Ours)	✓	✓	✓	✓	0.949±0.029	0.955±0.025



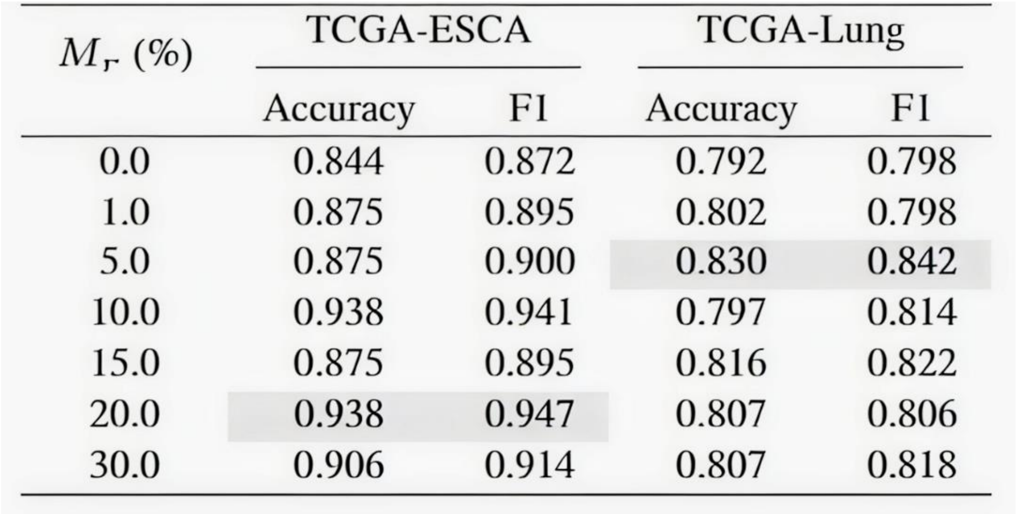


CSS consistently outperforms its variants, with the association matrix Q effectively bridging local and global interactions for improved group representations.

Ablation Studies-Hyperparameter Analysis



G	Accuracy	AUC	F1
2	0.860±0.091	0.890±0.083	0.876±0.081
5	0.917±0.044	0.933±0.035	0.931±0.036
10	0.936±0.028	0.965±0.032	0.945±0.025
20	0.910±0.037	0.948±0.032	0.926±0.030



M_r (%)	TCGA-ESCA		TCGA-Lung	
	Accuracy	F1	Accuracy	F1
0.0	0.844	0.872	0.792	0.798
1.0	0.875	0.895	0.802	0.798
5.0	0.875	0.900	0.830	0.842
10.0	0.938	0.941	0.797	0.814
15.0	0.875	0.895	0.816	0.822
20.0	0.938	0.947	0.807	0.806
30.0	0.906	0.914	0.807	0.818

G: grouping number $\boldsymbol{M_r}$: masking ratio

GMMamba clusters instances with similar tissues and structures into groups, reducing redundancy and producing more precise bag representations.

Content

1. Introduction
2. Method
3. Experiments
4. Conclusion

Conclusion

Key Innovations:

- Intra-group masking Mamba (**IMM**): removes local redundancy, yields compact group representations.
- Cross-group super-feature sampling (**CSS**): aggregates dispersed tumor features, enhances global representation.

Future Work:

Adaptive masking strategies and advanced instance selection.

If you have any questions or concerns, please do not hesitate to contact me.

Email: zhengtingting008@gmail.com

Phone: (+086)186388218568 (WeChat)