Bridging Visual-Textual Modalities: Weakly Supervised Histopathology Segmentation

Sicong Gao[1], Matthew AB Baker[2], Maurice Pagnucco[1], Zhiwei Gao[3], Yang Song[1]

[1]School of Computer Science and Engineering, and [2]School of Biotechnology and Biomolecular Sciences, University of New South Wales, Australia

[3]Hebei Key Laboratory of Electromagnetic Environmental Effects and Information Processing, China

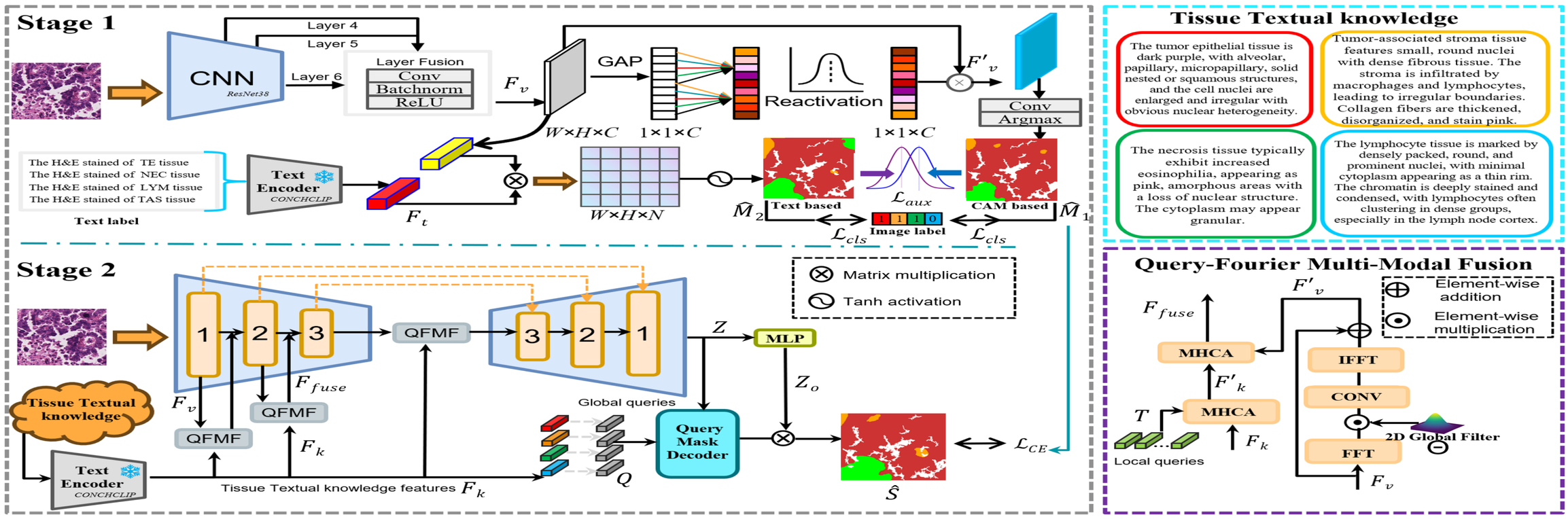
1.INTRODUCTION

Weakly supervised histopathology segmentation often relies on Class Activation Maps (CAMs) to produce pseudo-masks, which usually highlight only the most discriminative regions and result in poor boundary accuracy. To address this, we propose a two-stage framework that builds a bidirectional shared latent space between visual and textual modalities, improving pseudo-mask quality. In the segmentation stage, complex tissue descriptions are incorporated as external knowledge through a multi-stage fusion module based on learnable query tokens and Fourier transforms. Experiments on LUAD-HistoSeg and BCSS-WSSS show that our method achieves state-of-the-art performance.

2.OBJECTIVES

- Modality-Complementary Strategy: Activates overlooked visual regions and leverages textual prompts to refine incomplete or noisy pseudo-masks.
- Query-Fourier Knowledge Fusion (QFKF): Combines learnable query tokens for discriminative text selection with frequency-domain visual filters to capture structural cues.
- State-of-the-Art Performance: Achieves superior results on LUAD-HistoSeg and BCSS-WSSS benchmarks under weak supervision.

3.OVERALL FRAMEWORK



4.EXPERIMENTAL RESULTS

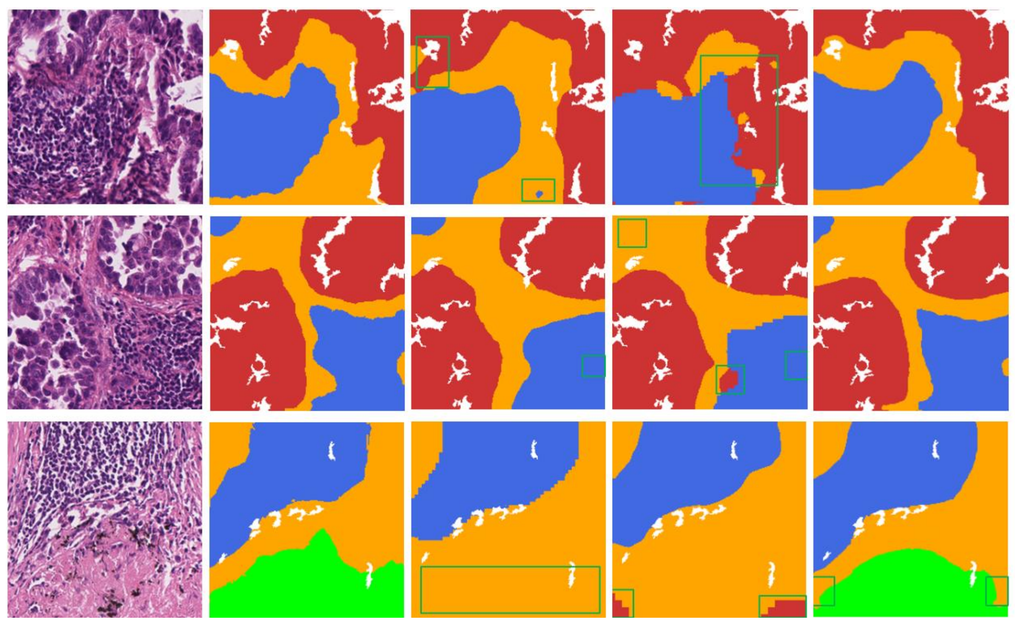
Dataset		LUAD-HistoSeg [17]					BCSS-WSSS [2]				
Method	Modal	TE	NEC	LYM	TAS	mIoU	TUM	STR	LYM	NEC	mIoU
Grad-CAM [38]	Vision	70.07	66.01	70.18	64.76	67.76	65.96	56.71	43.36	30.04	49.02
SC-CAM [5]	Vision	68.29	64.28	62.07	61.79	64.10	64.28	56.16	42.87	30.14	48.36
MLPS [17]	Vision	75.68	75.59	74.01	69.97	73.81	72.98	68.13	56.19	55.99	63.32
OEEM [26]	Vision	69.66	72.62	72.58	66.88	70.44	66.83	58.71	49.41	51.12	56.52
HAMIL [47]	Vision	72.82	69.79	69.82	70.96	70.85	69.84	59.45	49.98	51.29	57.64
PCSformer [30]	Vision	71.82	73.55	75.31	68.98	72.42	74.54	67.36	55.29	61.47	64.67
HisynSeg [12]	Vision	75.67	75.83	75.10	71.94	74.64	**78.73**	67.41	59.47	65.31	67.73
CLIP-ES [29]	Vision-Language	70.12	69.34	68.25	66.58	68.57	65.83	61.29	51.32	55.27	58.43
TPRO [46]	Vision-Language	74.82	**77.55**	76.40	70.98	74.94	77.18	63.77	54.95	61.43	64.33
Ours	Vision-Language	**76.32**	76.50	**77.20**	**76.78**	**76.70**	78.14	**68.32**	**63.24**	**67.77**	**69.37**

Table 1: Comparison of the pseudo mask results.

Dataset		LUAD-HistoSeg [17]					BCSS-WSSS [2]				
Method	Modal	TE	NEC	LYM	TAS	mIoU	TUM	STR	LYM	NEC	mIoU
HistoSegNet [4]	Vision	45.59	36.30	58.28	50.82	47.75	33.14	46.46	29.05	1.91	27.64
MLPS [17]	Vision	77.70	79.32	73.41	71.98	75.60	78.84	73.16	57.30	66.39	68.92
OEEM [26]	Vision	73.81	70.49	71.89	69.48	71.42	74.86	64.68	48.91	61.03	62.37
HAMIL [47]	Vision	73.46	75.83	72.94	70.86	73.27	71.65	62.37	51.52	54.29	59.96
PCSformer [30]	Vision	72.41	74.66	75.39	70.60	73.27	74.92	69.58	59.71	62.21	66.61
HisynSeg [12]	Vision	78.41	77.67	76.39	73.05	76.38	**81.67**	73.85	60.74	66.82	70.77
LAVT [44]	Vision-Language	71.36	68.74	68.19	67.28	68.89	67.72	65.21	50.99	59.83	60.94
Bridger [43]	Vision-Language	69.79	72.33	70.76	68.49	70.34	71.25	62.14	51.73	60.82	61.49
TPRO [46]	Vision-Language	75.80	**80.56**	78.14	72.69	76.80	77.95	65.10	54.55	64.96	65.64
Ours	Vision-Language	**79.21**	79.34	**80.22**	**77.30**	**79.02**	81.41	**74.22**	**65.21**	**69.24**	**72.52**

Table 2: Comparison of the final segmentation results.

5.VISUALIZATIONS



6.REFERENCES

[1] C. Han et al., "Multi-layer pseudo-supervision for histopathology tissue semantic segmentation using patch-level classification labels," *Medical Image Analysis*, vol. 80, 2022.

[2] S. Zhang et al., "TPRO: Text-prompting-based weakly supervised histopathology tissue segmentation," in *MICCAI*. Springer, 2023, pp. 109–118.

[3] Y. Lu et al., "A visual-language foundation model for computational pathology," *Nature Medicine*, 2024.

BMVC 2025

The 36th British Machine Vision Conference

24th - 27th November 2025, Sheffield, UK