Act Like a Pathologist: Tissue-Aware Whole Slide Image Reasoning

Wentao Huang, Weimin Lyu, Peiliang Lou, Qingqiao Hu, Xiaoling Hu,
Shahira Abousamra, Wenchao Han, Ruifeng Guo, Jiawei Zhou, Chao Chen, Chen Wang

Stony Brook University

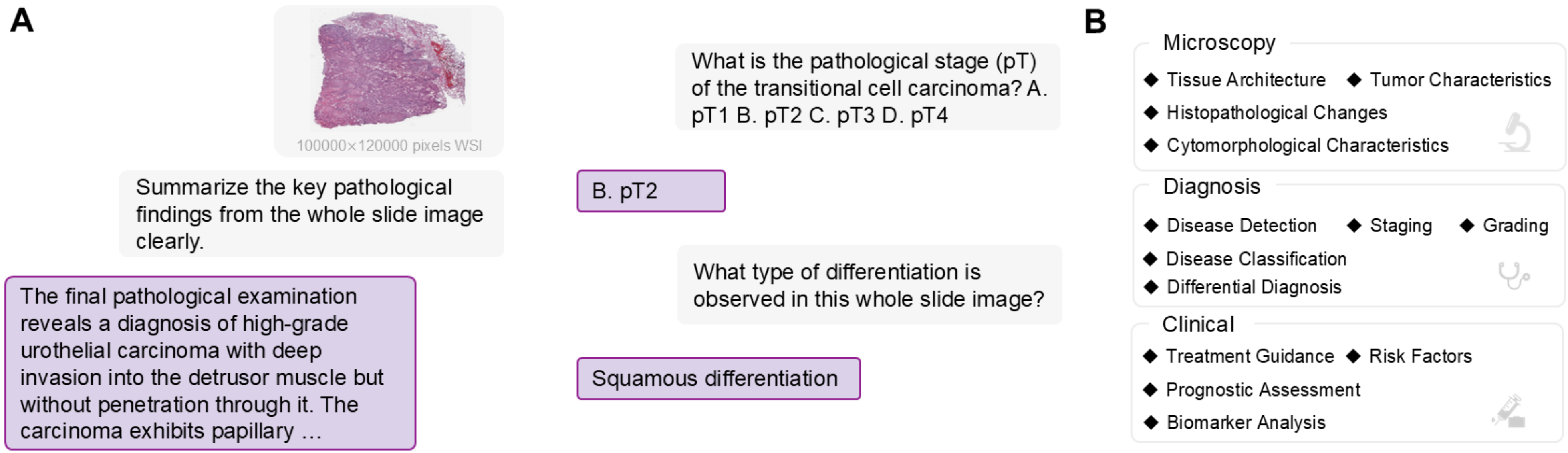
MAYO CLINIC

VE RI TAS

CVPR
JUNE 3-7, 2026
DENVER
COLORADO

Pathology Visual Question Answering

- **Goal: answer clinical questions about a whole-slide image (WSI).**
  - Diagnostic reasoning, report generation.
- **Why it matters:**
  - Scale pathologist expertise.
  - Clinically trustworthy, interpretable answers.

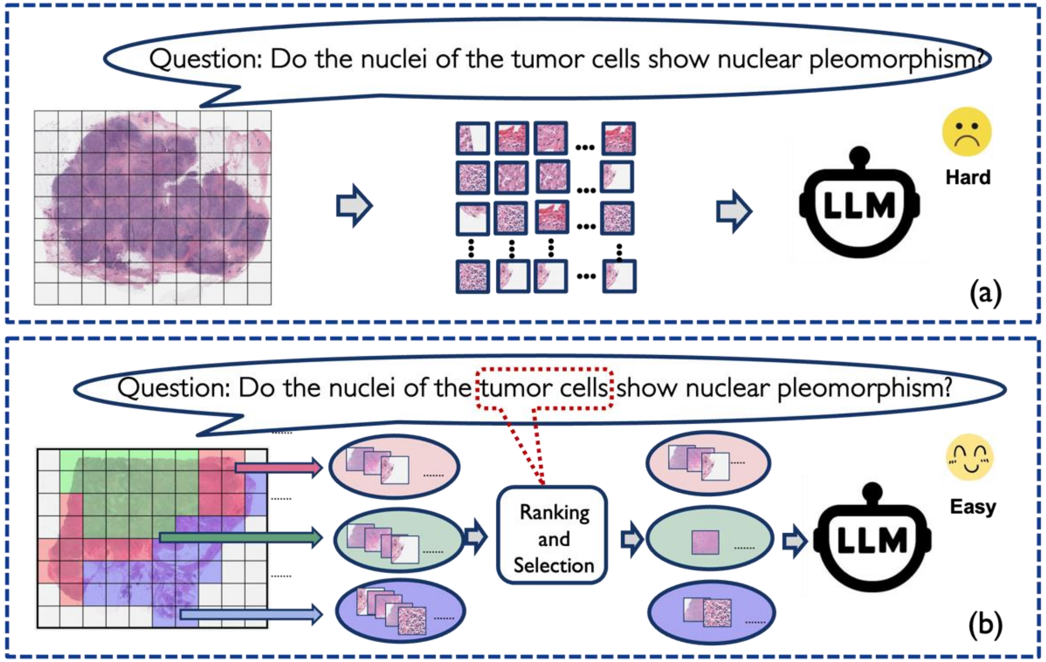


[1] Ying Chen, et al. "SlideChat: A Large Vision-Language Assistant for Whole-Slide Pathology Image Understanding." CVPR 2025

From Problem to Motivation

The Problem

- **A gigapixel slide contains far more information than necessary for a given question.**
  - A single WSI yields tens of thousands of patches.
  - Only a small subset carries the answer.
- **Current WSI-MLLMs treat every patch equally.**
  - Strict LLM token limits force uniform sampling or pooling.
  - Redundant + question-irrelevant tokens overwhelm the LLM.
- **Predictions are black-box and unverifiable.**
  - No patch-level attribution → pathologists can't verify the reasoning.

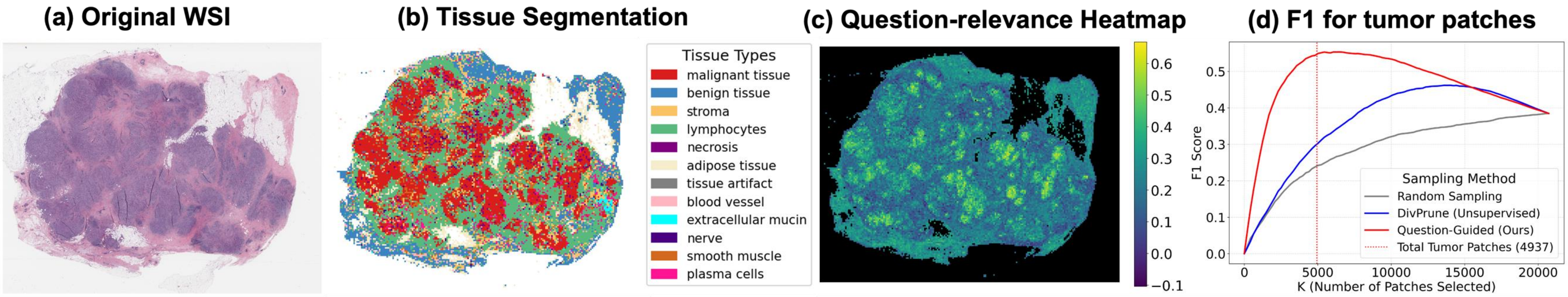


Our Motivation

- Pathologists scan broadly, then zoom in based on the clinical question.

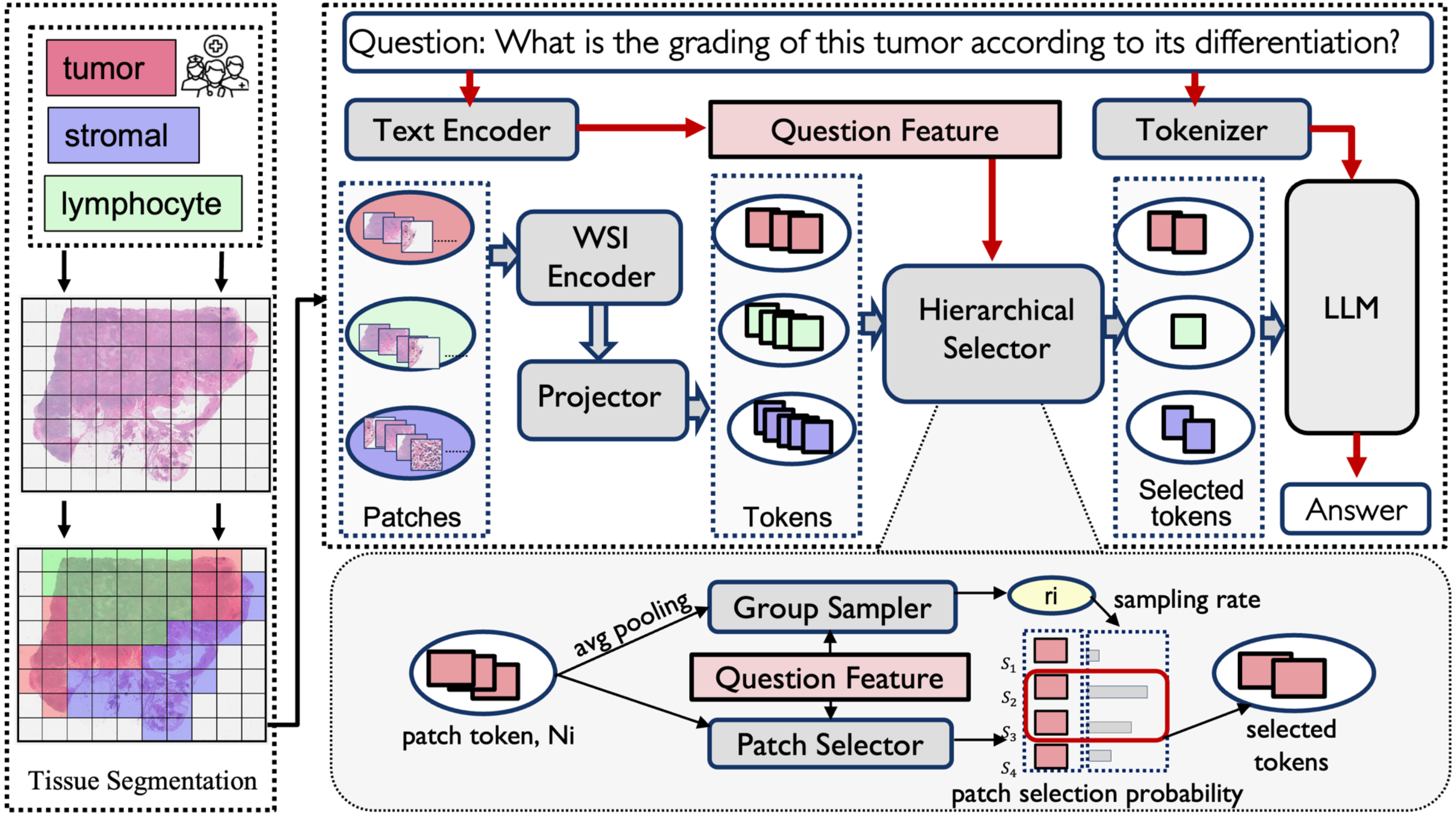
Pre-analysis

- **Pathologists scan broadly, then zoom in selectively, guided by the question.**
- **Pre-analysis on TCGA-BRCA for patch retrieval:**
  - Patch–question similarity heatmap overlaps the tumor mask.
  - Question-guided ≫ diversity ≫ random in F1 score for retrieving tumor patches.
  - Diversity is not enough — patch selection must be driven by the question.



Our Approach: HistoSelect

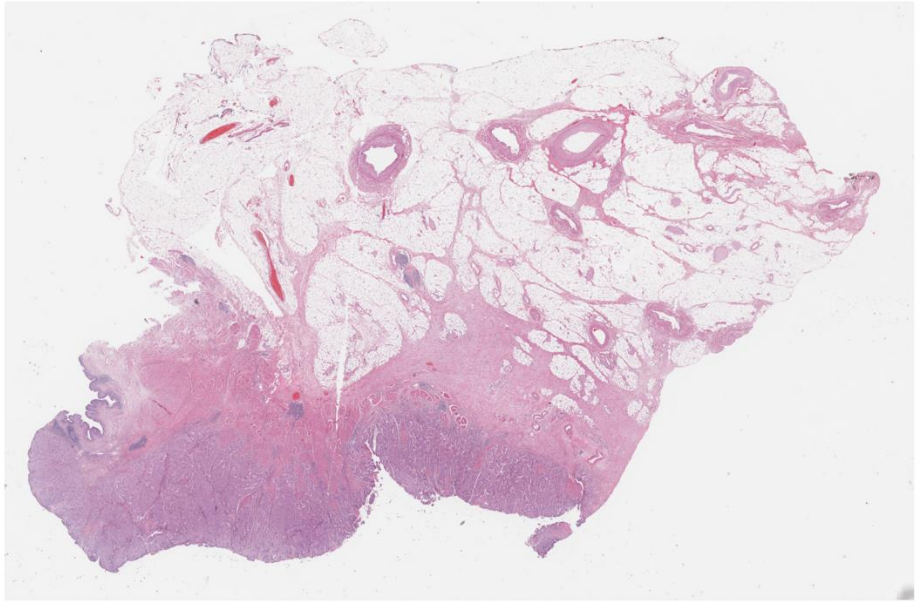
A question-guided, tissue-aware, coarse-to-fine patch selector grounded in the Information Bottleneck (IB).



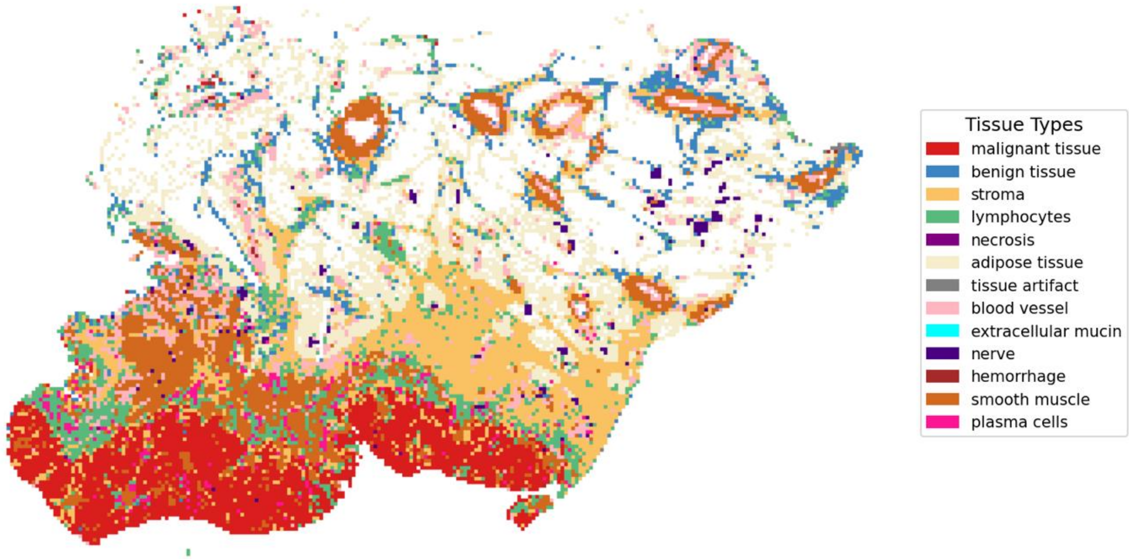
Tissue Segmentation

- **M pathologist-designed tissue prompts: tumor / stroma / lymphocyte / necrosis / …**
- **Match each patch to a prompt by visual–text cosine similarity (CONCH encoders).**
  - Patch i is assigned to the tissue group with the highest similarity.
- **Yields a partition of the WSI into M semantically coherent regions — the tissue segmentation map.**

Original WSI (Thumbnail)



Predicted Tissue Class Visualization



Hierarchical Selector + IB

- **Group Sampler**
  - predicts a question-aware sampling rate per tissue group.
- **Patch Selector**
  - scores each patch by relevance to the question; top-k_j kept per group.
- **Hierarchical Information Bottleneck**
  - regularizes selection toward question-aligned tokens.
- **Straight-Through Estimator** makes top-k differentiable for end-to-end VQA training.

$$r_j = \sigma(\mathcal{F}_{\text{group}}([\mathbf{g}_j \,;\, \mathbf{q}])) \qquad s_i = \sigma(\mathcal{F}_{\text{patch}}([\mathbf{x}_i \,;\, \mathbf{q}]))$$

$$\mathcal{L}_{\text{final}} = \mathcal{L}_{\text{VQA}} + \beta_g \, \mathcal{L}_{\text{group}} + \beta_p \, \mathcal{L}_{\text{patch}}$$

Quantitative Results

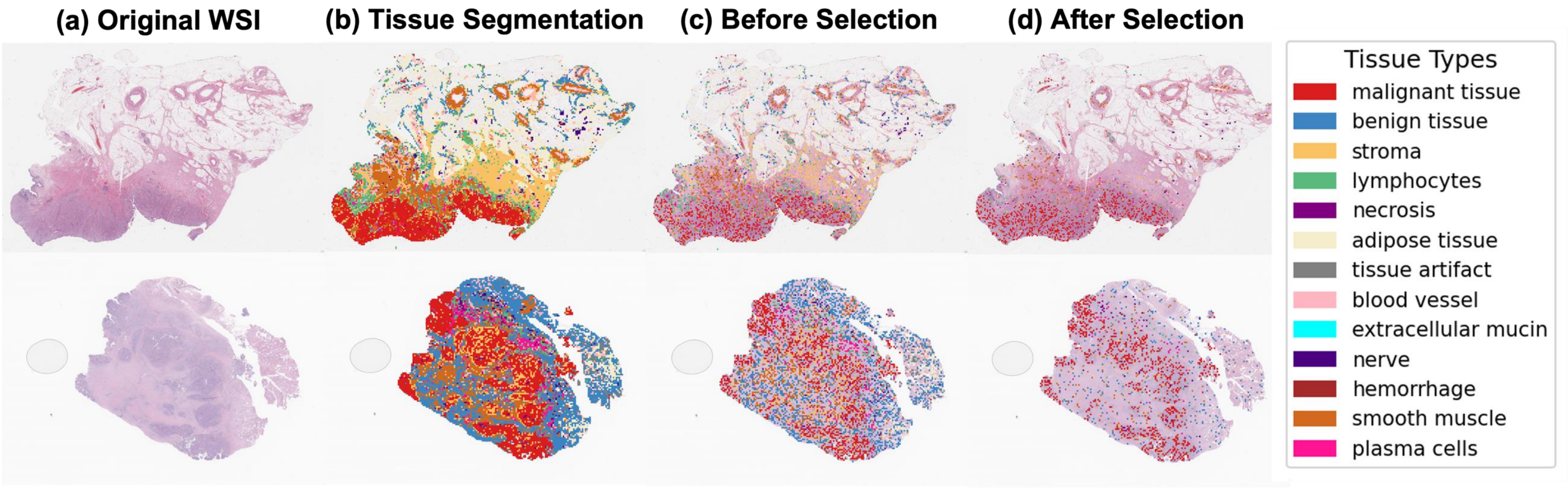
➢ Close Ended

Method	Input	SlideBench-VQA (TCGA) Microscopy	SlideBench-VQA (TCGA) Diagnosis	SlideBench-VQA (TCGA) Clinical	WSI-Bench (Close) Morphology	WSI-Bench (Close) Diagnosis	WSI-Bench (Close) Treatment	In-house Ovarian Diagnosis	Average
GPT-4o	Thumbnail	39.24	24.12	44.67	47.07	53.06	87.50	—	49.28
Quilt-LLaVA[32]	Thumbnail	52.39	30.19	49.33	94.13	84.13	97.92	70.67	68.39
LLaVA-Med [20]	Thumbnail	52.15	29.97	47.33	91.04	81.32	95.83	70.67	66.90
SlideChat [9]	WSI	83.15	71.36	75.33	91.34	82.15	93.75	69.33	80.88
HistoSelect	WSI	**84.62**	**73.09**	**77.30**	**94.57**	**85.79**	**97.92**	**73.33**	**83.80**

➢ Open Ended

Method	WSI-Bench Report Generation BLEU-1	BLEU-2	BLEU-3	BLEU-4	ROUGE-L	Morphology WSI-P	Morphology WSI-R	Diagnosis WSI-P	Diagnosis WSI-R	Treatment WSI-P	Treatment WSI-R
GPT-4o	0.202	0.069	0.030	0.016	0.132	0.220	0.204	0.472	0.457	0.513	0.704
WSI-VQA [7]	0.301	0.225	0.181	0.155	0.343	0.395	0.462	0.436	0.525	0.591	0.595
MI-Gen [6]	0.403	0.306	0.248	0.209	0.446	—	—	—	—	—	—
Histo-Gen [13]	0.406	0.307	0.248	0.208	0.448	—	—	—	—	—	—
Quilt-LLaVA [32]	0.421	0.316	0.257	0.216	0.455	0.453	0.484	0.521	0.552	0.751	**0.807**
SlideChat [9]	0.413	0.312	0.254	0.215	0.450	0.512	0.541	0.501	0.522	0.745	0.712
HistoSelect	**0.431**	**0.324**	**0.262**	**0.221**	**0.463**	**0.538**	**0.589**	**0.542**	**0.587**	**0.766**	0.801

Qualitative Result



Conclusion

- HistoSelect brings a pathologist-style coarse-to-fine, question-guided selection mechanism into WSI-VQA.
- Cuts visual tokens by ~70% while reaching SOTA performance on 3 different datasets.
- Delivers clinically attributable evidence — every prediction is backed by selected patches.

Thank You!

Code: https://github.com/winston52/HistoSelect

Contact: wentaohuang@cs.stonybrook.edu