



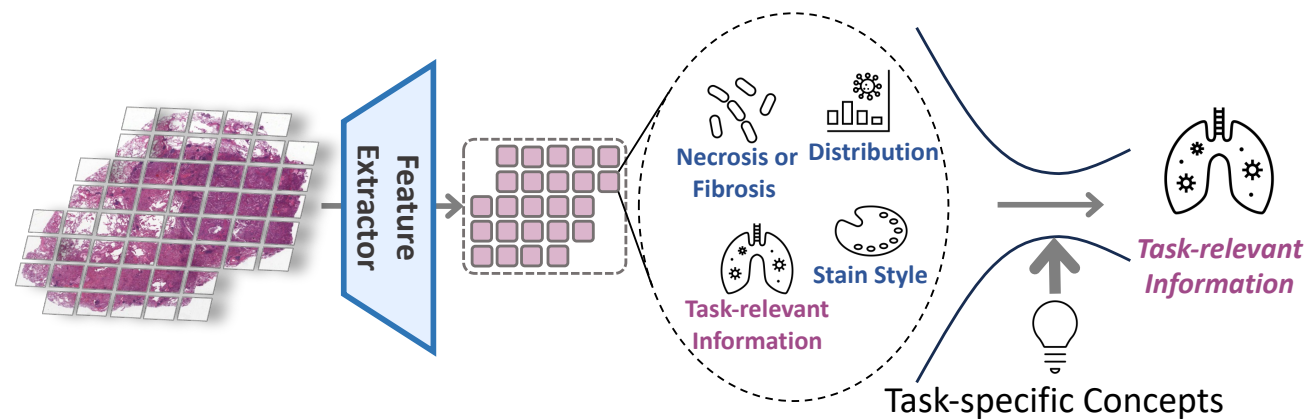
Free Lunch in Pathology Foundation Model: Task-specific Model Adaptation with Concept-Guided Feature Enhancement

Yanyan Huang¹, Weiqin Zhao¹, Yihang Chen¹, Yu Fu², Lequan Yu¹

¹The University of Hong Kong, ²Lanzhou University

Motivation

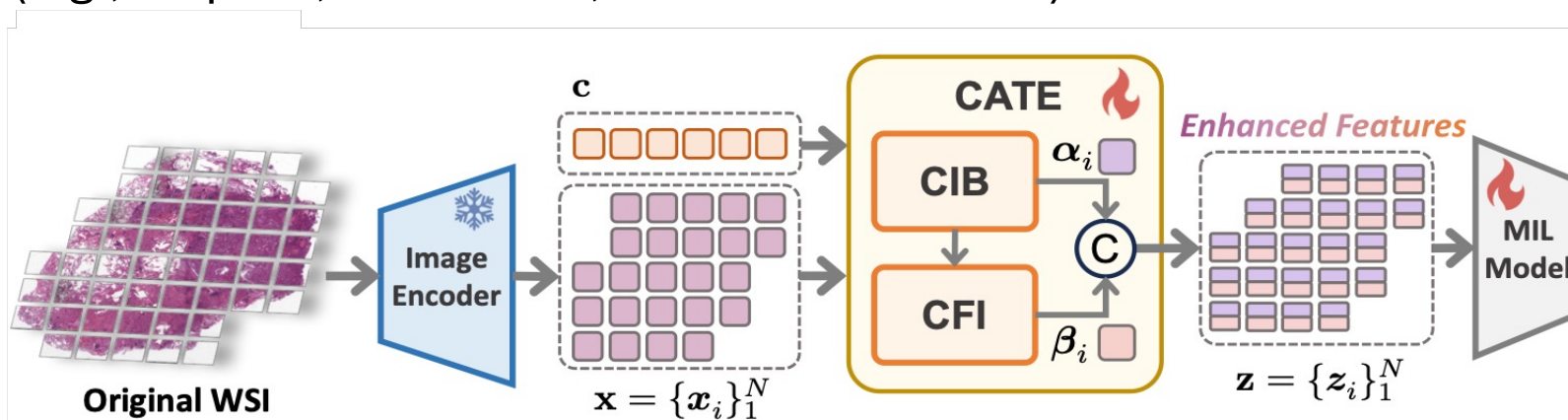
- *Can we undertake task-specific adaptation to further enhance pathology visual-language foundation model?*



- The features extracted by the image encoder of a pathology VLM can include both **task-relevant information** and **task-irrelevant elements**.
- We present **Concept Anchor-guided Task-specific Feature Enhancement (CATE)**, an adaptable paradigm that can boost the expressivity and discriminativeness of pathology foundation models for specific downstream tasks.

Overview

- **Objective:** Obtain the enhanced task-specific feature set \mathbf{z} from the original feature \mathbf{x} with the guidance of pre-extracted concepts anchors \mathbf{c} .
- Task-specific concept anchors are generated by the text encoder of pathology VLM with prompt p , including **class-specific concepts** (e.g., subtyping classes) and **class-agnostic concepts** (e.g., adipose, connective, and normal tissues).



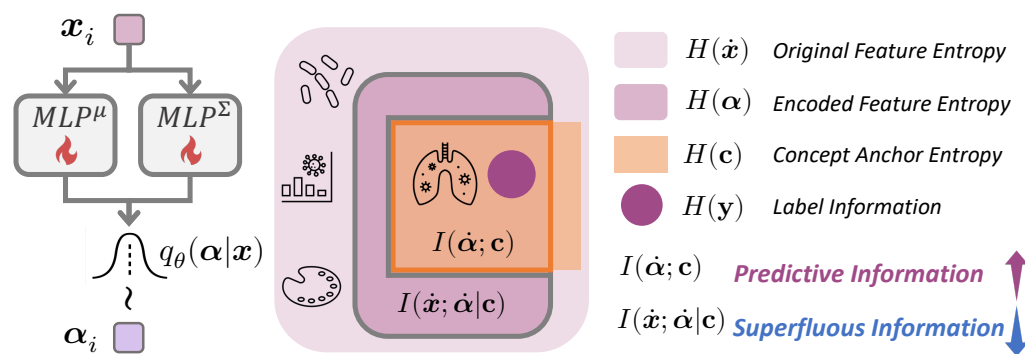
$$\mathbf{z} = \mathbf{CATE}(\mathbf{x}, \mathbf{c}), \hat{y} = \text{MIL}(\mathbf{z}).$$

$$z_i = \text{Concat}[\alpha_i, \beta_i] = \text{Concat}[\mathbf{CIB}(\mathbf{x}_i, \mathbf{c}), \mathbf{CFI}(\mathbf{x}_i, \mathbf{c})].$$

Concept-guided Information Bottleneck (CIB)

The mutual information between the representative subset of the original feature set and the corresponding enhanced feature set can be decomposed into **Predictive Information** and **Superfluous Information**.

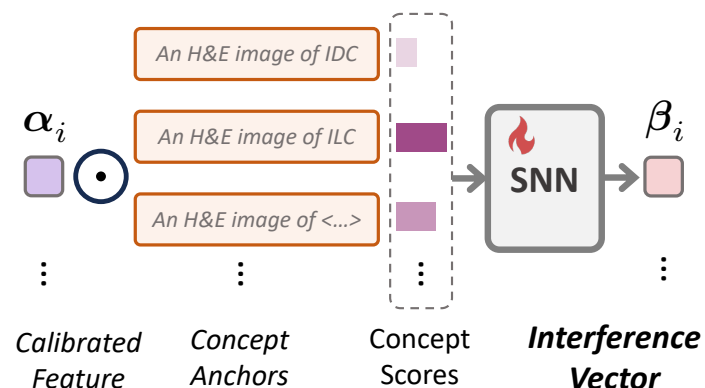
- The predictive information can be maximized by maximizing the **InfoNCE** mutual information lower bound.
- The superfluous information can be minimized by utilizing **Variational Information Bottleneck** to minimize the mutual information between original feature set and corresponding enhanced feature set.



$$I(\hat{x}; \hat{\alpha}) = \underbrace{I(\hat{\alpha}; c)}_{\text{Predictive Information}} + \underbrace{I(\hat{x}; \hat{\alpha}|c)}_{\text{Superfluous Information}}$$

Concept-Feature Interference (CFI)

- Calculate the cosine similarity between each CIB encoded feature α_i and each class specific concept c_i^{CS} .
- A SNN layer is utilized to encode the similarity vector to interference vector β_i to be aligned with the calibrated feature α_i .



- The overall training objective can be represented as the combination of the cross entropy loss \mathcal{L}_{CE} , the predictive information maximization loss \mathcal{L}_{PIM} , and the superfluous information minimization loss \mathcal{L}_{SIM} .

$$\mathcal{L} = \mathcal{L}_{CE} + \lambda_P \mathcal{L}_{PIM} + \lambda_S \mathcal{L}_{SIM}$$

Experiments

- We selected several sites as **IND** data (in-domain, the testing and training data are from the same sites) and used data from other sites as **OOD** data (out-of-domain, the testing and training data are from different sites), and reported the testing performance on both IND and OOD data.

BRCA ($N_{IND}=1$)									
Method	CATE	OOD-AUC		Gain		OOD-ACC		Gain	
		OOD-AUC	Gain	OOD-ACC	Gain	IND-AUC	Gain	IND-ACC	Gain
ABMIL	✗	0.914±0.015	N/A	0.852±0.014	N/A	0.963±0.044	N/A	0.888±0.053	N/A
CLAM	✗	0.907±0.017	N/A	0.802±0.053	N/A	0.965±0.049	N/A	0.888±0.068	N/A
DSMIL	✗	0.925±0.020	N/A	0.836±0.048	N/A	0.969±0.040	N/A	0.900±0.080	N/A
DTFD-MIL	✗	0.912±0.012	N/A	0.858±0.020	N/A	0.944±0.058	N/A	0.894±0.070	N/A
TransMIL	✗	0.918±0.015	N/A	0.832±0.046	N/A	0.969±0.036	N/A	0.918±0.067	N/A
R ² T-MIL [†]	✗	0.901±0.027	N/A	0.816±0.051	N/A	0.965±0.033	N/A	0.894±0.022	N/A
ABMIL	✓	0.951 ±0.003	↑4.05%	0.897 ±0.026	↑5.28%	0.998 ±0.006	↑3.63%	0.965 ±0.045	↑8.67%
CLAM	✓	0.951 ±0.005	↑4.85%	0.906 ±0.020	↑12.97%	0.998 ±0.006	↑3.42%	0.965 ±0.037	↑8.67%
DSMIL	✓	0.936±0.007	↑1.19%	0.866±0.036	↑3.59%	0.990±0.022	↑2.17%	0.959±0.044	↑6.56%
DTFD-MIL	✓	0.947±0.004	↑3.84%	0.906 ±0.009	↑5.59%	0.985±0.028	↑4.34%	0.953±0.042	↑6.60%
TransMIL	✓	0.938±0.005	↑2.18%	0.880±0.023	↑5.77%	0.998 ±0.006	↑2.99%	0.965 ±0.027	↑5.12%

BRCA ($N_{IND}=2$)									
Method	CATE	OOD-AUC		Gain		OOD-ACC		Gain	
		OOD-AUC	Gain	OOD-ACC	Gain	IND-AUC	Gain	IND-ACC	Gain
ABMIL	✗	0.899±0.035	N/A	0.892±0.019	N/A	0.967±0.019	N/A	0.941±0.024	N/A
CLAM	✗	0.893±0.030	N/A	0.862±0.019	N/A	0.960±0.042	N/A	0.935±0.027	N/A
DSMIL	✗	0.881±0.032	N/A	0.852±0.028	N/A	0.946±0.057	N/A	0.940±0.020	N/A
DTFD-MIL	✗	0.909±0.019	N/A	0.878±0.014	N/A	0.973±0.023	N/A	0.945±0.041	N/A
TransMIL	✗	0.904±0.023	N/A	0.852±0.090	N/A	0.966±0.031	N/A	0.936±0.052	N/A
R ² T-MIL [†]	✗	0.902±0.028	N/A	0.873±0.027	N/A	0.946±0.060	N/A	0.929±0.048	N/A
ABMIL	✓	0.943±0.006	↑4.89%	0.907 ±0.018	↑1.68%	0.981 ±0.018	↑1.45%	0.948±0.030	↑0.74%
CLAM	✓	0.945±0.008	↑5.82%	0.896±0.030	↑3.94%	0.976±0.023	↑1.67%	0.938±0.043	↑0.32%
DSMIL	✓	0.919±0.015	↑4.31%	0.869±0.036	↑2.00%	0.958±0.051	↑1.27%	0.949 ±0.024	↑0.96%
DTFD-MIL	✓	0.946 ±0.005	↑4.07%	0.887±0.027	↑1.03%	0.977±0.023	↑0.41%	0.946±0.036	↑0.11%
TransMIL	✓	0.920±0.011	↑1.77%	0.867±0.046	↑1.76%	0.968±0.045	↑0.21%	0.940±0.026	↑0.43%

* The best results are highlighted in **bold**, and the second-best results are underlined.

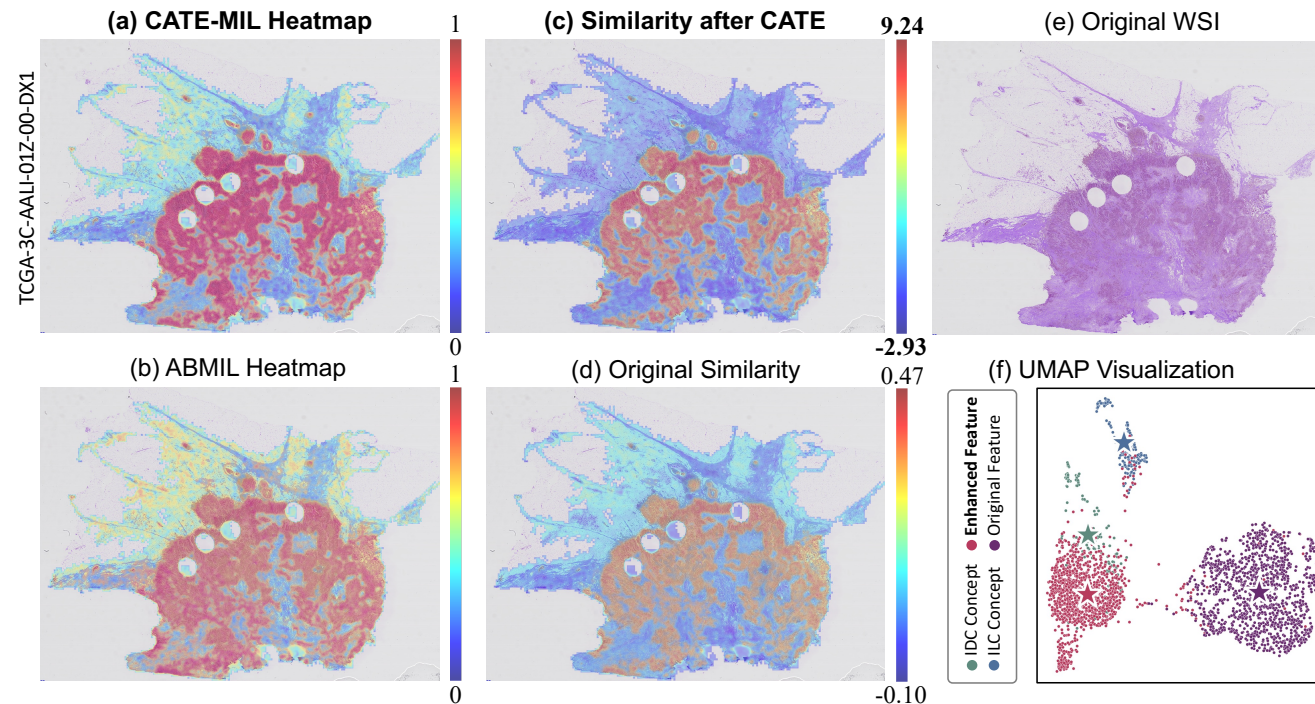
† R²T-MIL is designed for feature re-embedding that utilize ABMIL as base MIL model.

Method	NSCLC ($N_{IND}=2$)				NSCLC ($N_{IND}=4$)			
	OOD-AUC	OOD-ACC	IND-AUC [#]	IND-ACC [#]	OOD-AUC	OOD-ACC	IND-AUC [#]	IND-ACC [#]
ABMIL	0.874±0.021	0.803±0.021	0.997±0.004	0.954±0.028	<u>0.951</u> ±0.023	0.883±0.029	0.974±0.018	0.910±0.036
CLAM	0.875±0.020	0.801±0.021	0.997±0.007	0.963±0.042	0.931±0.037	0.870±0.036	0.977±0.023	0.926±0.048
DSMIL	0.839±0.046	0.764±0.043	0.993±0.004	0.963±0.028	0.934±0.019	0.864±0.026	0.974±0.013	0.913±0.042
DTFD-MIL	0.903±0.023	0.836±0.026	0.990±0.009	0.958±0.049	0.949±0.010	0.893±0.012	0.981±0.012	0.918±0.040
TransMIL	0.790±0.028	0.712±0.024	0.997±0.004	0.954±0.033	0.917±0.022	0.832±0.031	0.977±0.014	0.923±0.029
R ² T-MIL [†]	0.739±0.088	0.690±0.075	0.999±0.002	0.971±0.036	0.892±0.041	0.800±0.059	0.977±0.018	0.916±0.045
CATE-MIL	0.945 ±0.016	0.840 ±0.043	0.985±0.011	0.938±0.037	0.969 ±0.003	0.906 ±0.011	0.967±0.019	0.905±0.054

Method	RCC ($N_{IND}=3$)				RCC ($N_{IND}=6$)			
	OOD-AUC	OOD-ACC	IND-AUC [#]	IND-ACC [#]	OOD-AUC	OOD-ACC	IND-AUC [#]	IND-ACC [#]
ABMIL	0.973±0.005	0.891±0.017	0.997±0.004	0.961±0.032	<u>0.971</u> ±0.007	0.885±0.010	0.973±0.010	0.897±0.023
CLAM	0.972±0.004	0.893±0.012	0.991±0.005	0.961±0.032	0.969±0.009	0.888±0.015	0.975±0.011	0.896±0.031
DSMIL	0.977±0.002	0.893±0.010	0.996±0.006	0.965±0.026	0.969±0.008	0.883±0.016	0.980±0.012	0.901±0.022
DTFD-MIL	0.975±0.003	0.897±0.012	0.996±0.004	0.943±0.046	<u>0.971</u> ±0.007	0.893±0.017	0.974±0.012	0.878±0.022
TransMIL	0.961±0.010	0.864±0.022	0.994±0.004	0.930±0.030	0.947±0.017	0.828±0.037	0.975±0.013	0.894±0.027
R ² T-MIL [†]	0.956±0.018	0.847±0.022	0.991±0.008	0.936±0.030	0.932±0.020	0.803±0.048	0.974±0.012	0.897±0.029
CATE-MIL	0.983 ±0.002	0.911 ±0.018	0.989±0.009	0.944±0.031	0.979 ±0.007	0.905 ±0.017	0.963±0.011	0.882±0.032

Visualization

- Attention heatmap comparisons reveal that CATE-MIL focuses more intensely on cancerous regions, with a clearer delineation between high and low attention areas.
- The enhanced similarity in cancerous regions is significantly higher than in original features.
- CATE could effectively align features with task-relevant concepts and enhance task-relevant information.



Thank You!

