CHONGQING UNIVERSITY

THE HONG KONG UNIVERSITY OF SCIENCE AND TECHNOLOGY

Multiple Instance Learning Framework with Masked Hard Instance Mining for Whole Slide Image Classification

Wenhao Tang, Sheng Huang, Xiaoxian Zhang, Fengtao Zhou, Yi Zhang, Bo Liu*

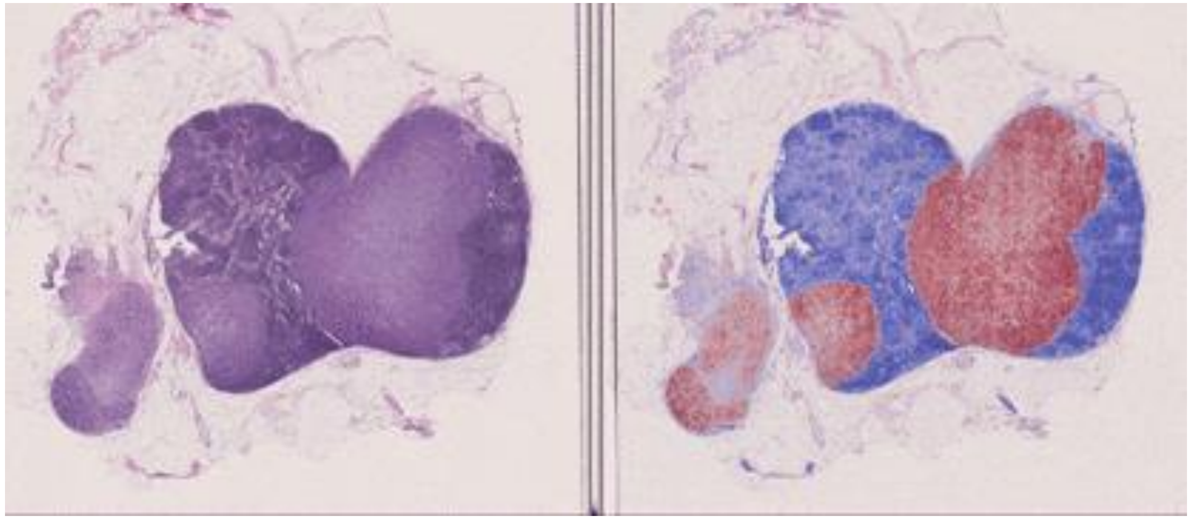
Reporter: Wenhao Tang   PID: 6302

GitHub

arXiv

Background: WSI Classification

ICCV23 PARIS



[*CLAM. Nature Biomedical Engineering 2021.*]

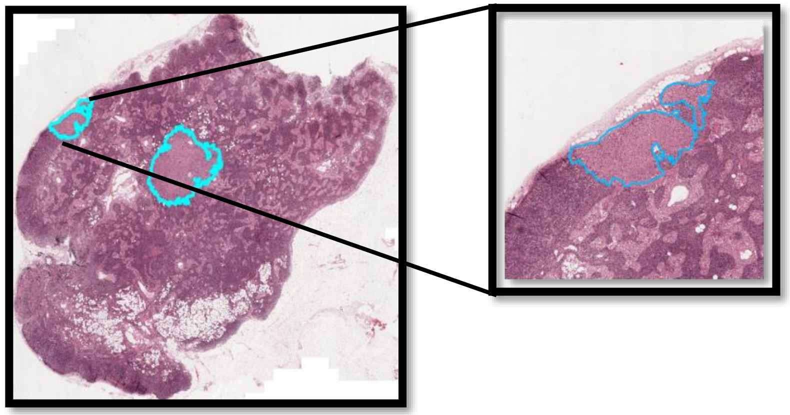
≈ 150000 × 150000 pixels per image

≈ 1000 images per Dataset

- **Whole-slide Images (WSI) Classification**
  - Many patches: **2k-20k**
  - Feature-level: use **features** as input
  - Weakly Supervised: **patch** label is **not** available
  - **Offline Feature Extractor**
  - **Redundancy**
    - **Noise fools the model**
    - **Low Efficiency (Transformer)**

Lu M Y, Williamson D F K, Chen T Y, et al. Data-efficient and weakly supervised computational pathology on whole-slide images. Nature biomedical engineering, 2021. Cite: 593

Previous Works: Only for Salient Area



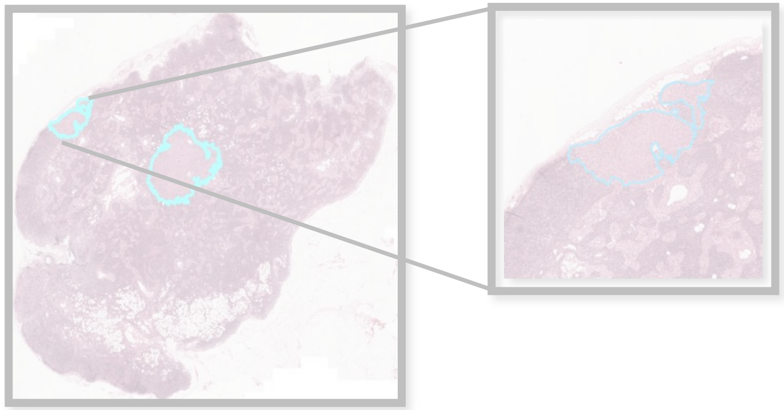
Whole-slide Image

Gigapixel resolution requires us to **zoom in**, **zoom in**, and **zoom in** again ...

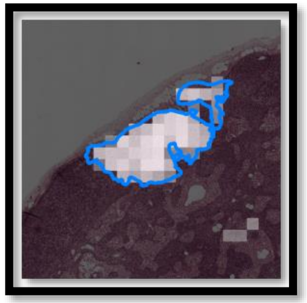
TINY details determine the classification of the **HUGE** image!

Ilse M, Tomczak J, Welling M. Attention-based deep multiple instance learning. ICLR 2018. Cite: 1245

Previous Works: Only for Salient Area



Whole-slide Image



Attention

Used Model: AB-MIL *(ResNet in MIL)*

Only for Salient Area

It seems Perfect ?

$$\hat{Y} = C(F), \qquad F = \sum_{i=1}^{N} a_i z_i$$

a_i Attention Score, z_i Patch Feature

Ilse M, Tomczak J, Welling M. Attention-based deep multiple instance learning. ICLR 2018. Cite: 1245

Previous Works: Only for Salient Area

ICCV23 PARIS

Used Model: AB-MIL *(ResNet in MIL)* Tumor Probabilities Algorithm: DTFD *(SoTA in MIL)*

Whole-slide Image Attention Tumor Probabilities

Not Perfect in Tumor Probabilities

Attention Score ≠ Tumor Probabilities

It was believed infeasible to explicitly infer instance probabilities under AB-MIL frameworks [*DSMIL. CVPR 2021 Oral*]

Zhang H, Meng Y, Zhao Y, et al. DTFD-MIL: Double-tier feature distillation multiple instance learning for histopathology whole slide image classification. CVPR 2022. Cite: 84

Previous Works: Only for Salient Area

ICCV23 PARIS

Used Model: AB-MIL *(ResNet in MIL)* / Tumor Probabilities Algorithm: DTFD *(SoTA in MIL)*

Whole-slide Image

Attention

Tumor Probabilities

Attention After Softmax

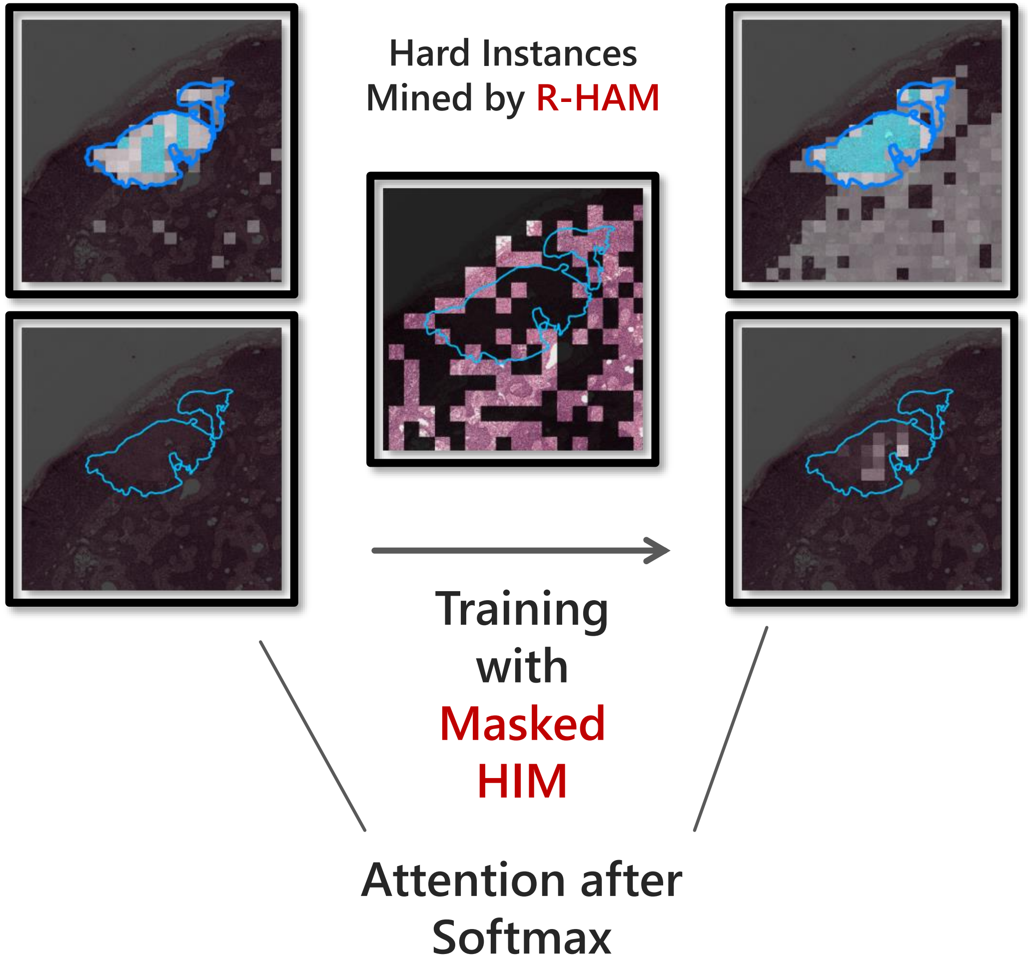
Attention Score ≠ Tumor Probabilities ≠ Final Prediction

$$\hat{Y} = C(F), \qquad F = \sum_{i=1}^{N} a_i z_i \,, \qquad A = Softmax(A)$$

Not Perfect in Final Prediction

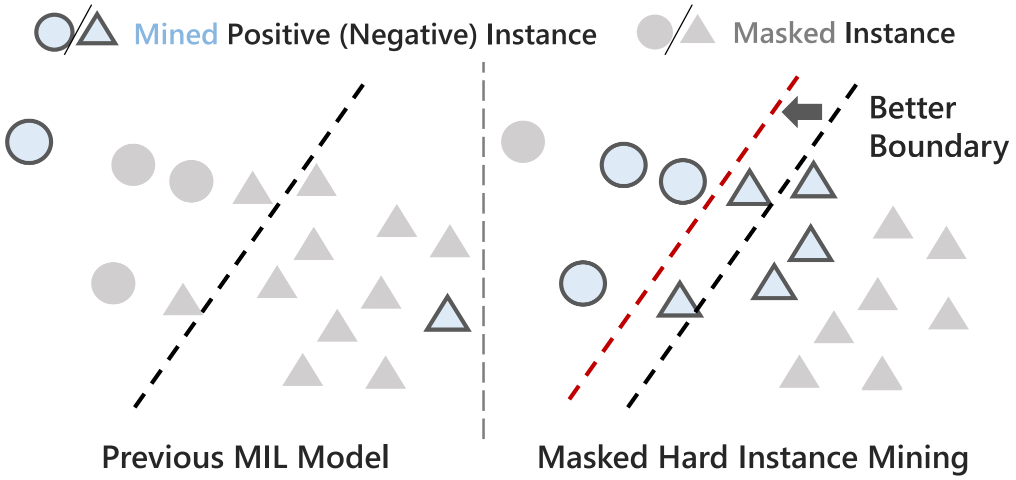
Ilse M, Tomczak J, Welling M. Attention-based deep multiple instance learning. ICLR 2018. Cite: 1245

MHIM: More ‘Useless’, More Powerful

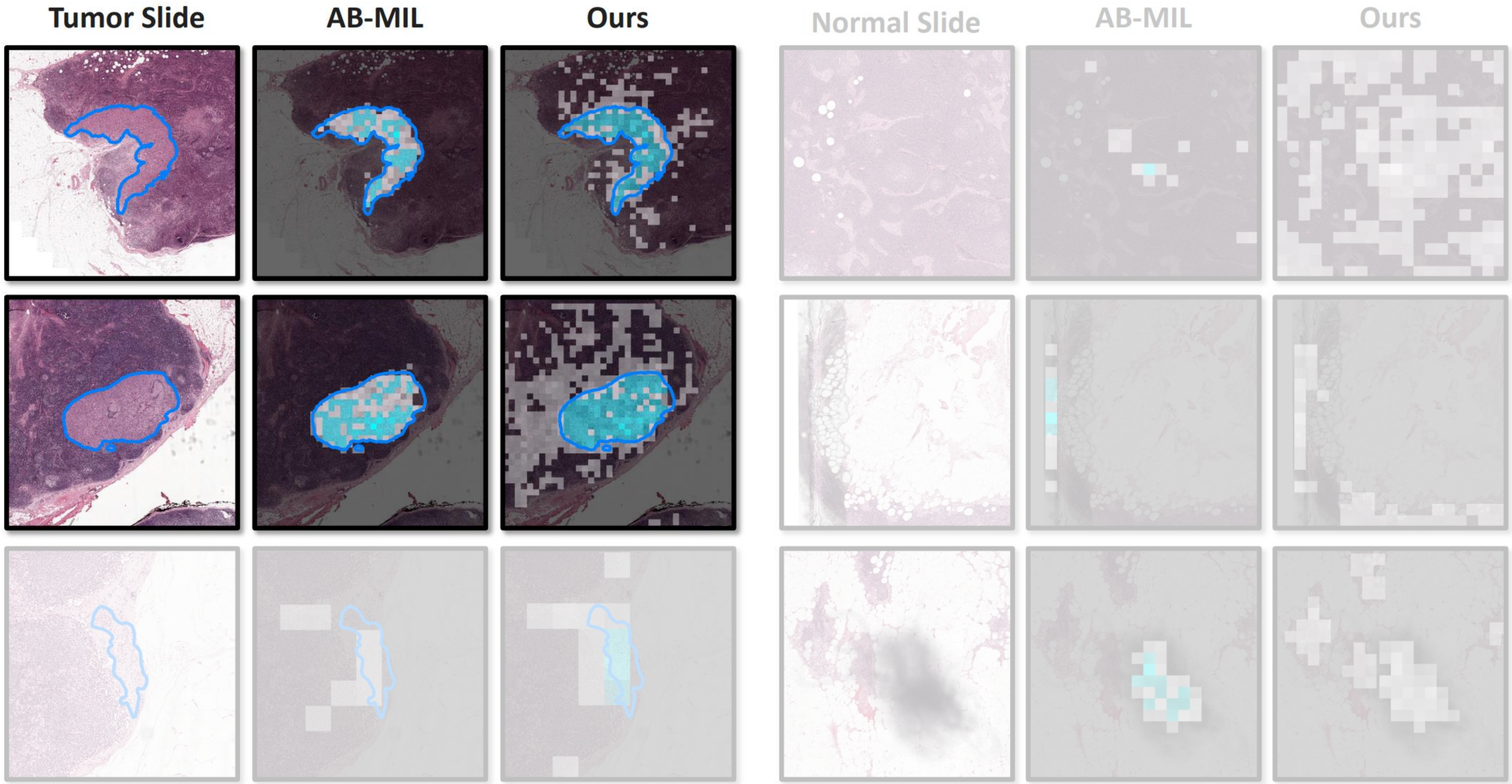


Intuitive: Only focus on salient area， but fail to detect complete tumor area, and fail to get better prediction

Counter-intuitive: Focus on more “useless area”， and make more complete prediction

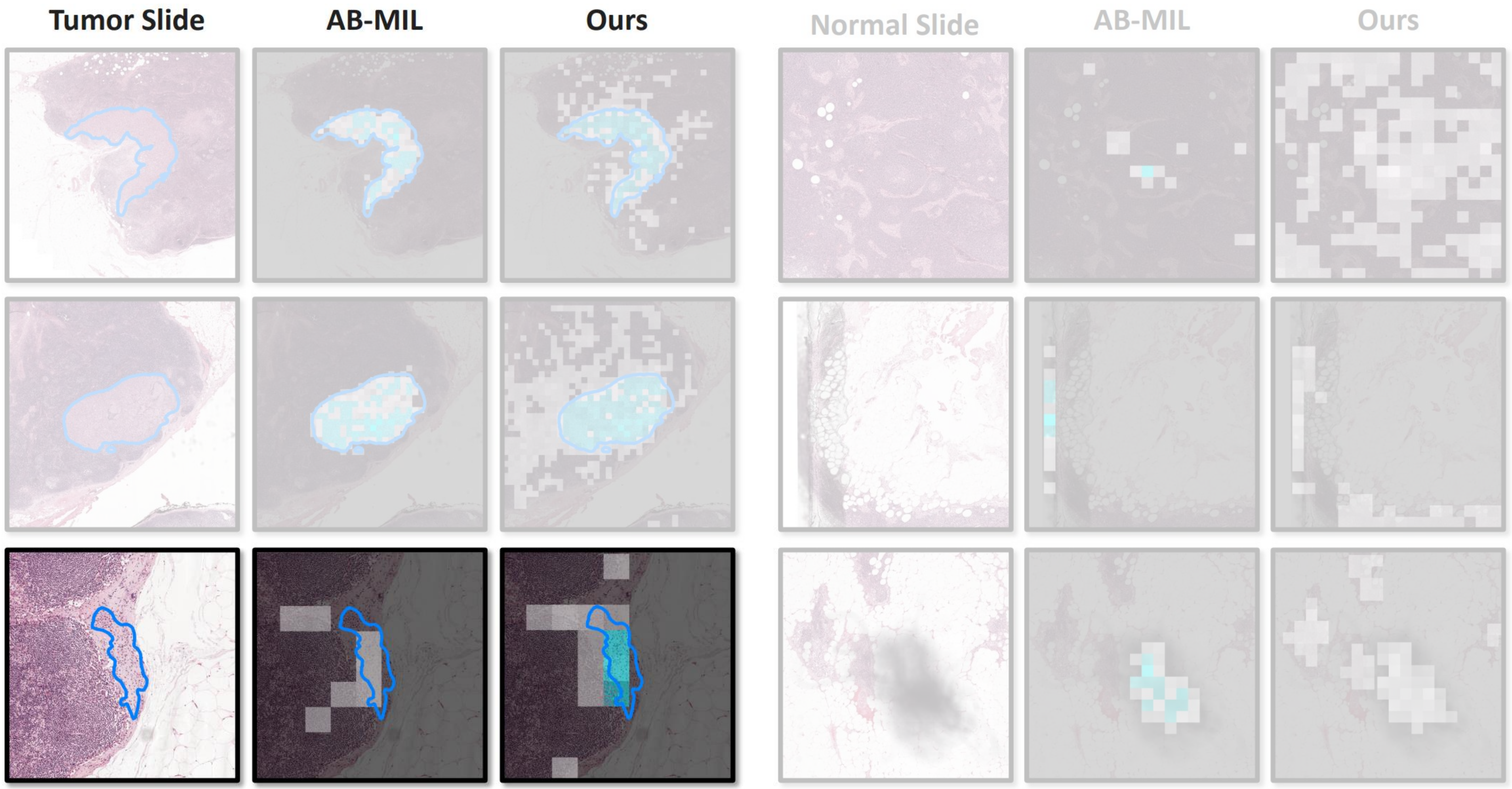


MHIM: More 'Useless', More Powerful



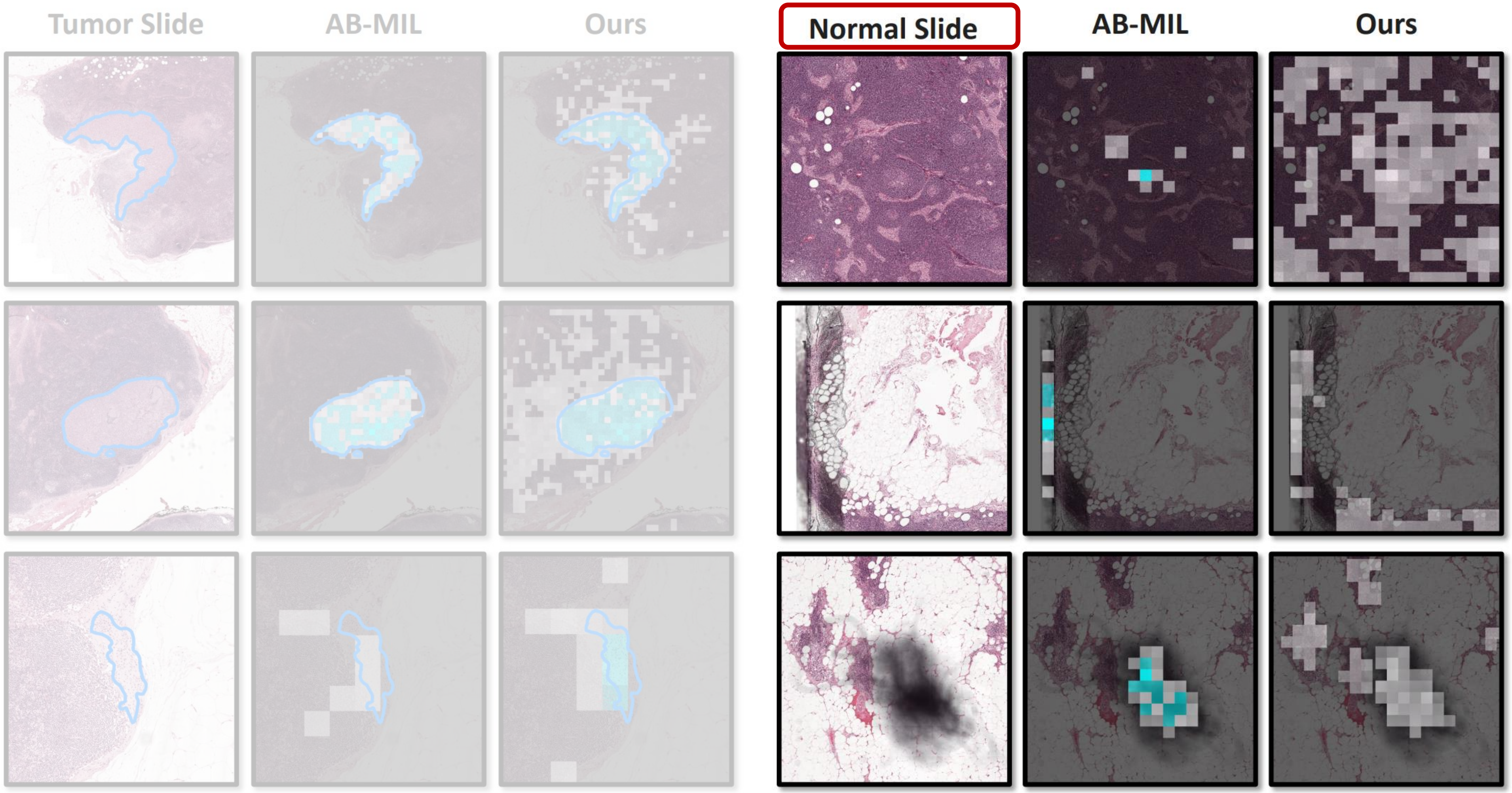
More Complete Prediction

MHIM: More 'Useless', More Powerful



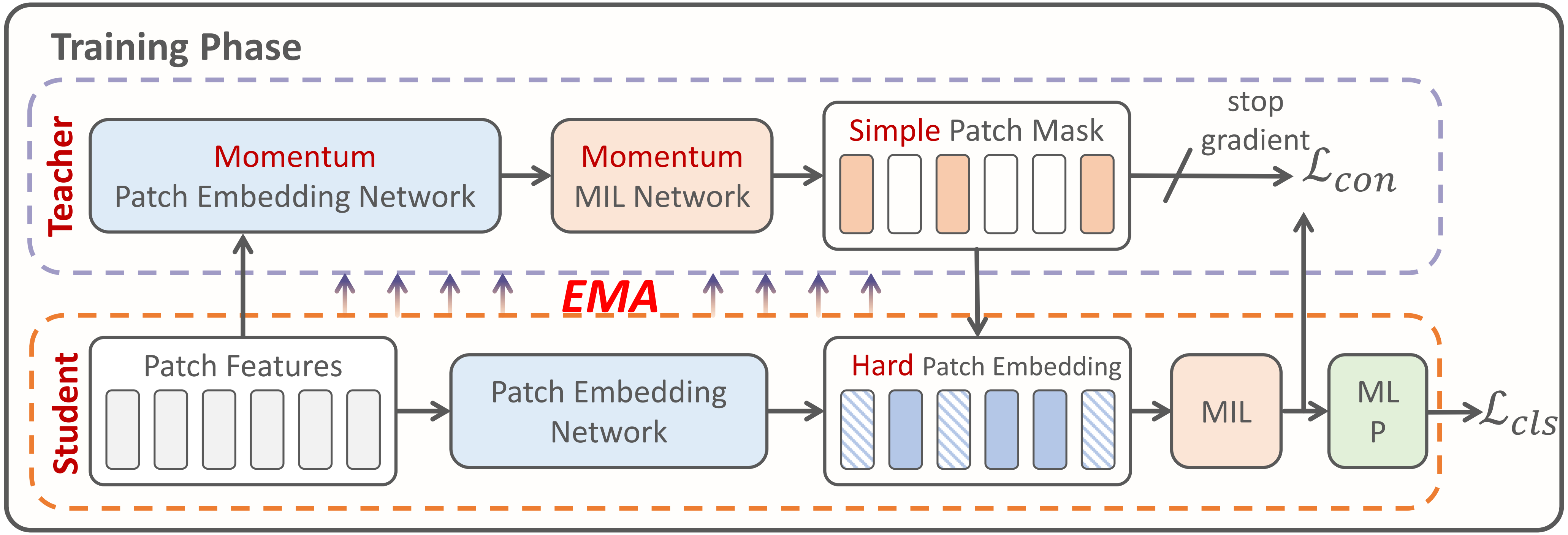
More Accurate Prediction

MHIM: More 'Useless', More Powerful



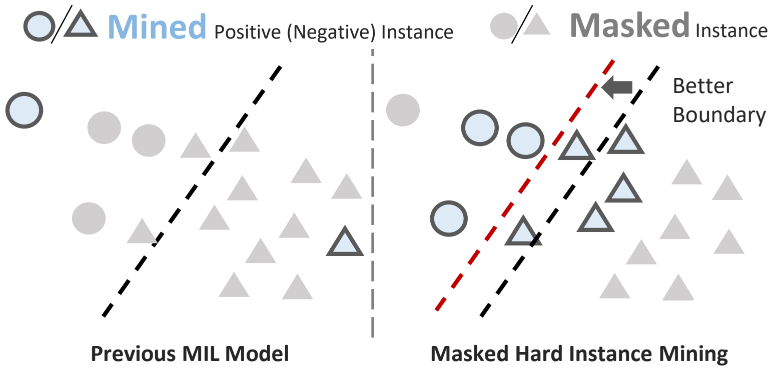
More Robust Prediction

MHIM: More 'Useless' and More Powerful

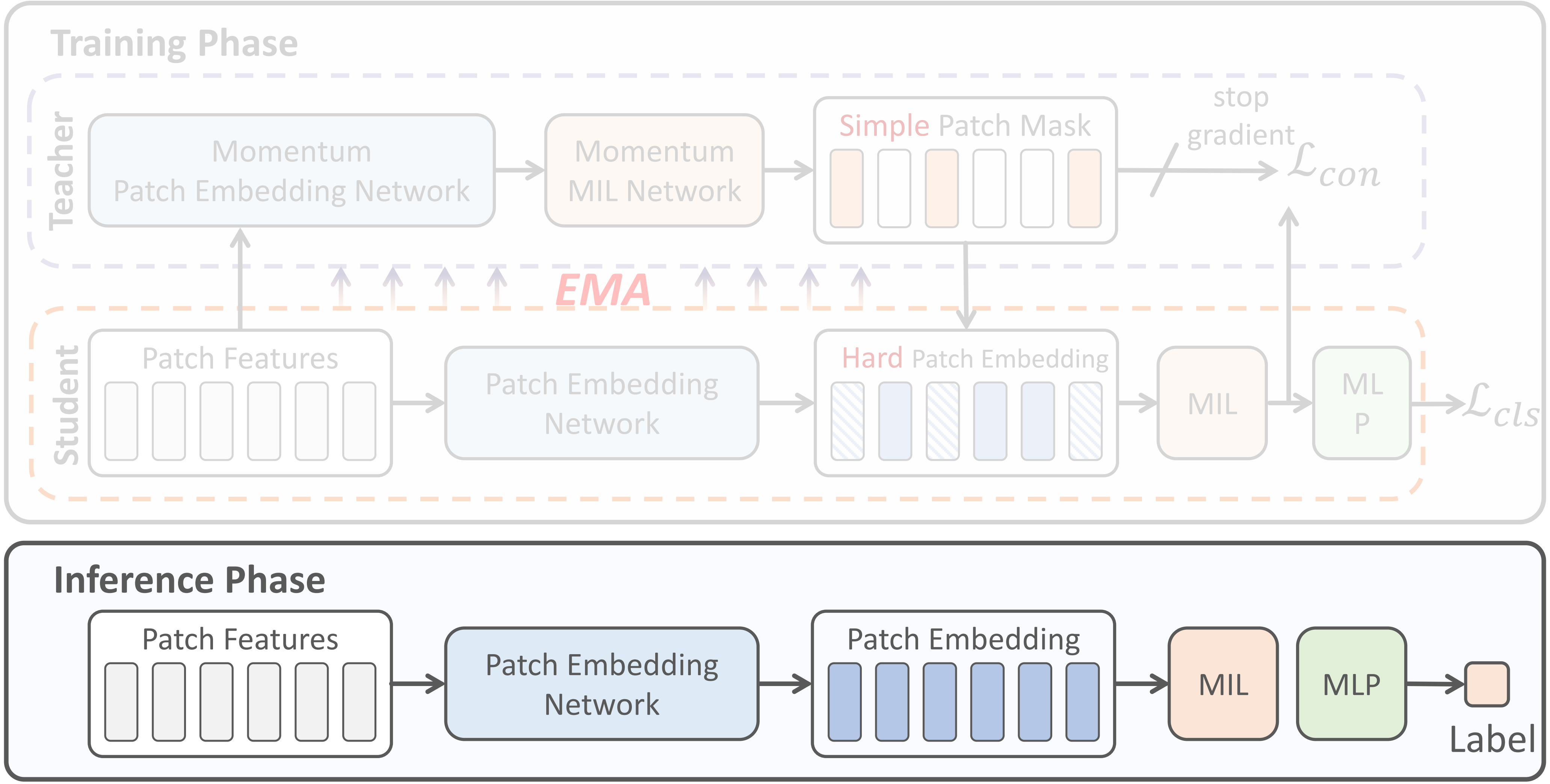


How to **effectively** mine **hard** instances **without** instance label ?

Find **Simple** First, **Mask** it Then, Remaining is **Hard**.



MHIM: More 'Useless' and More Powerful



Plain Inference Pipeline Enhanced Efficiency

MHIM: More 'Useless' and More Powerful

ICCV23 PARIS

Model	C16	TCGA	Para.	Time	Mem.
AB-MIL	94.00	93.17	**657K**	**4.0s**	2.4G
CLAM-MB	94.70	93.69	789K	4.3s	2.7G
DTFD-MIL	95.15	93.83	987K	5.2s	**2.1G**
MHIM-MIL	**96.14**	**94.97**	**657K**	4.3s	2.3G
TransMIL	93.51	92.51	2.67M	13.1s	10.6G
MHIM-MIL	**96.49**	**94.87**	**2.67M**	**10.1s**	**5.5G**

Module	CAMELYON-16		TCGA	
	AB.	Trans.	AB.	Trans.
Baseline	94.00	93.51	93.17	92.51
+MHIM	95.86	96.06	94.14	93.75
+MHIM+Siam.	95.82	96.24	94.55	94.13
+MHIM+Siam.+Con.	**96.14**	**96.49**	**94.97**	**94.87**

	C16	TCGA
DSMIL	94.57±0.40	93.71±1.82
MHIM	96.22±0.28 (+1.65)	95.27±1.66 (+1.56)
MHIM[‡]	96.49±0.65 (+1.92)	95.53±1.74 (+1.82)

Strategy	CAMELYON-16		TCGA	
	AB.	Trans.	AB.	Trans.
Baseline	94.00	93.51	93.17	92.51
HAM	95.68	95.90	93.83	94.54
R-HAM	**96.14**	95.88	94.79	94.60
L-HAM	95.81	**96.49**	94.33	94.67
LR-HAM	95.92	96.33	**94.97**	**94.87**

Adapt to All Major Baselines *(AB-MIL, TransMIL, DSMIL)*

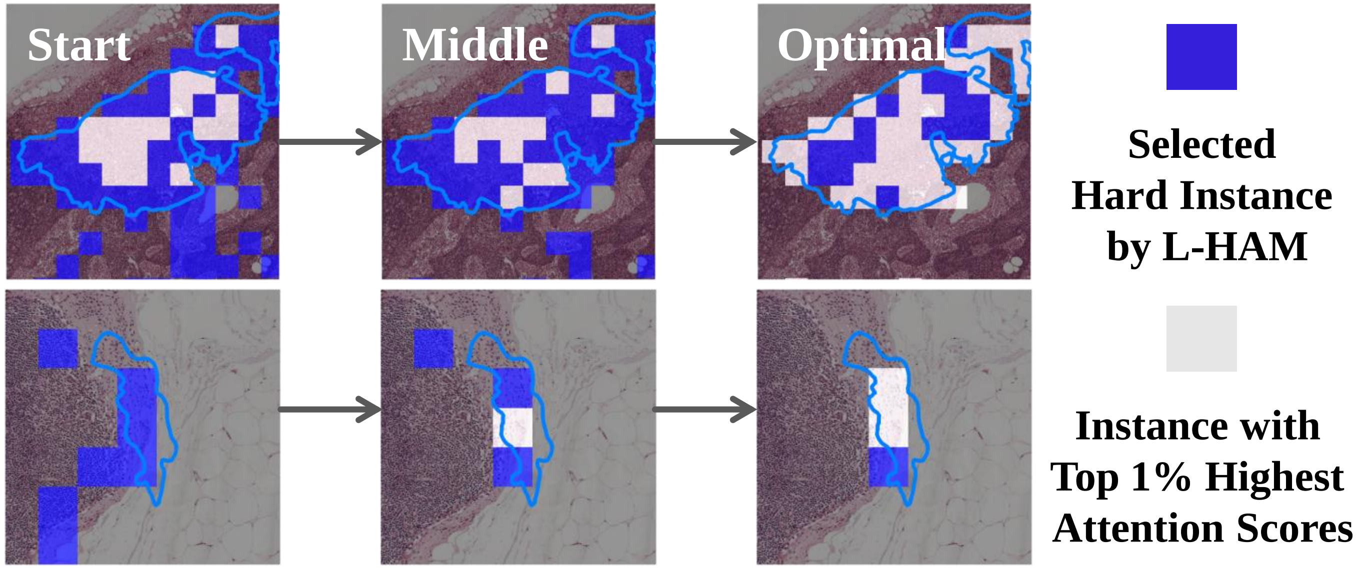
More Powerful and More Efficient *(+1.7% AUC on TCGA, -48% Mem.)*

MHIM: More 'Useless' and More Powerful

ICCV23 PARIS

Model	C16	TCGA	Para.	Time	Mem.
AB-MIL	94.00	93.17	**657K**	**4.0s**	2.4G
CLAM-MB	94.70	93.69	789K	4.3s	2.7G
DTFD-MIL	95.15	93.83	987K	5.2s	**2.1G**
MHIM-MIL	**96.14**	**94.97**	**657K**	4.3s	2.3G
TransMIL	93.51	92.51	2.67M	13.1s	10.6G
MHIM-MIL	**96.49**	**94.87**	**2.67M**	**10.1s**	**5.5G**

	C16	TCGA
DSMIL	94.57±0.40	93.71±1.82
MHIM	96.22±0.28 (+1.65)	95.27±1.66 (+1.56)
MHIM‡	96.49±0.65 (+1.92)	95.53±1.74 (+1.82)



Can teacher really provide hard instances through Mask strategies?

Take Home Message

Background: **WSI Classification**

- Gigapixel Resolution, Low-Data Dataset
- Offline Feature Extractor
- Redundancy

Previous Work: **Only for Salient Area**

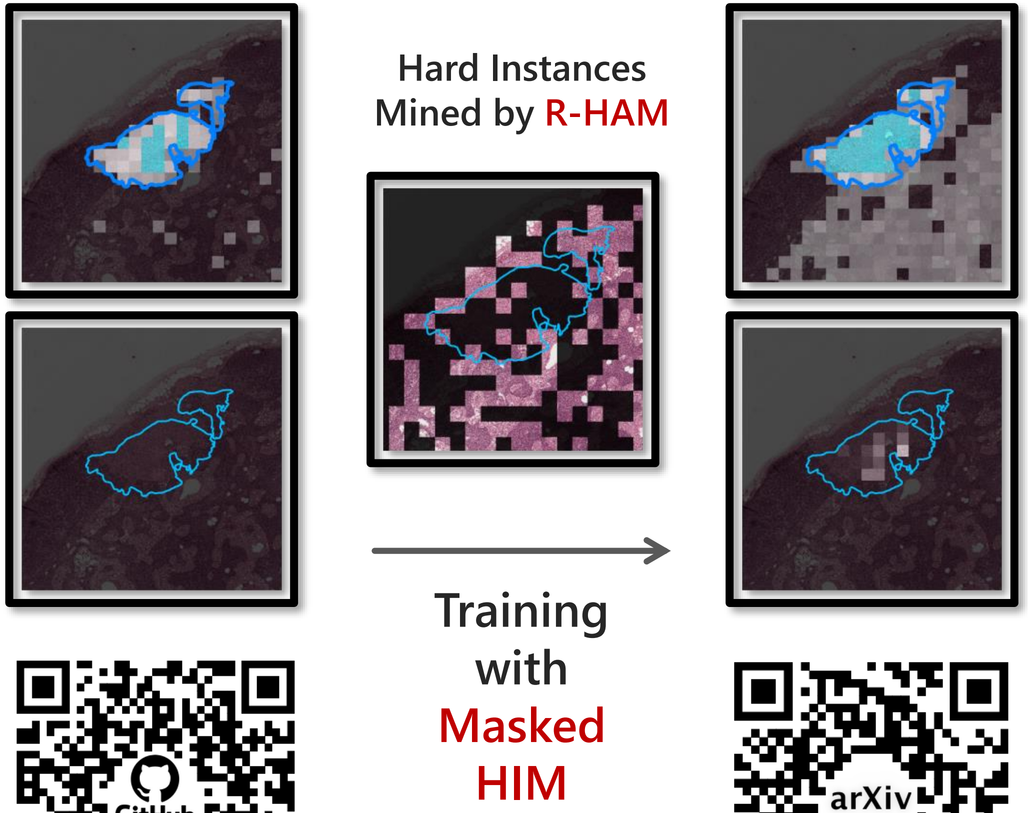
- Not Perfect in Tumor Probabilities
- Not Perfect in Final Prediction

MHIM: **More 'Useless', More Powerful**

- More Complete Prediction
- More Accurate Prediction
- More Robust Prediction

Welcome to MHIM: **Look for More about**

- Masked Strategy, Experiment, Visualization
- Dataset, Pre-process, Implement Detail



Contact me: whtang@cqu.edu.cn / (+086) 17784750069 (WeChat)