Applied Deep Learning Project

Detecting Cancer Metastases on Gigapixel Pathology Images

Presenter : Yingxiang Chen

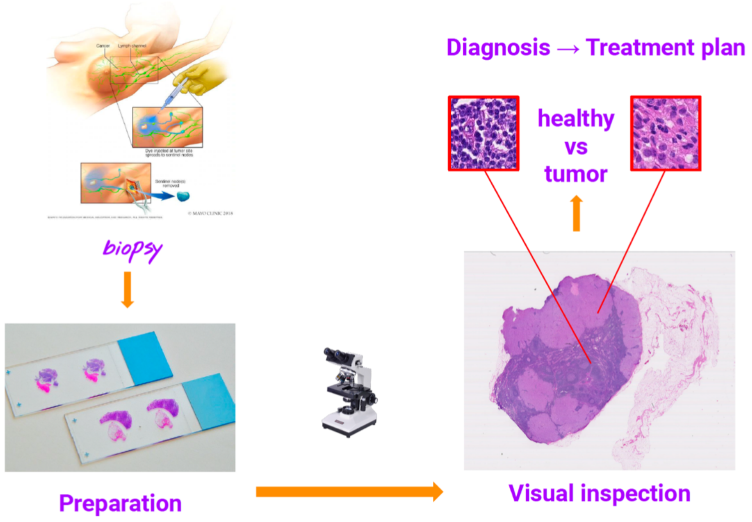
Columbia Uni : yc3526

Major : MSBA

Content

01- Motivation

01-Motivation



Source: Applied deep learning slides

- Microscopic examination of lymph nodes **is crucial** in breast cancer staging[1]
- Currently the manual process **requires highly skilled pathologists** [1]
- Fairly **time-consuming and error-prone**, particularly for lymph nodes with either no or small tumors [1]

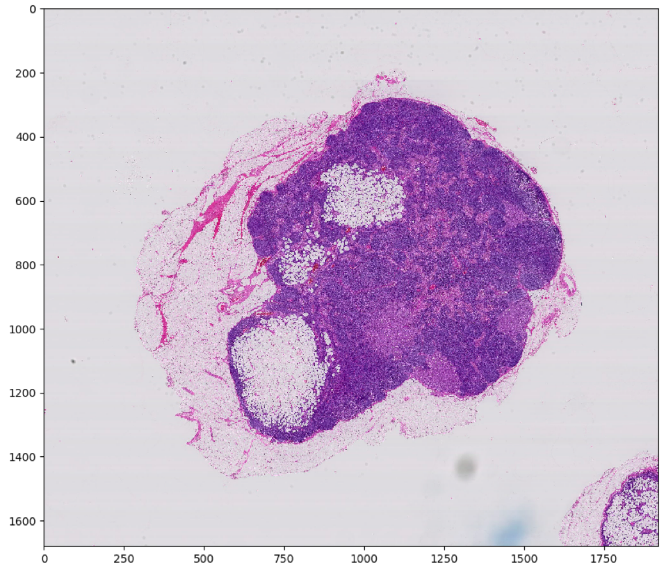
[1] Liu, Yun, et al. "Detecting cancer metastases on gigapixel pathology images." *arXiv preprint arXiv:1703.02442* (2017).

02- Methods

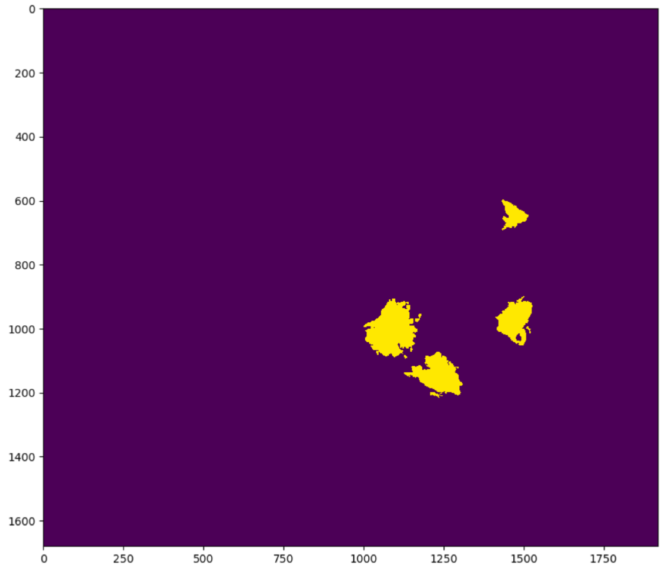
02-Methods

Data Preprocessing

- Raw Data: 22 Gigapixel Pathology Images, each has a tumor slide and a corresponding mask
- Sanity check: Remove slide 38 because it doesn’ t have correspond mask



Tumor Slide (Actual Images)

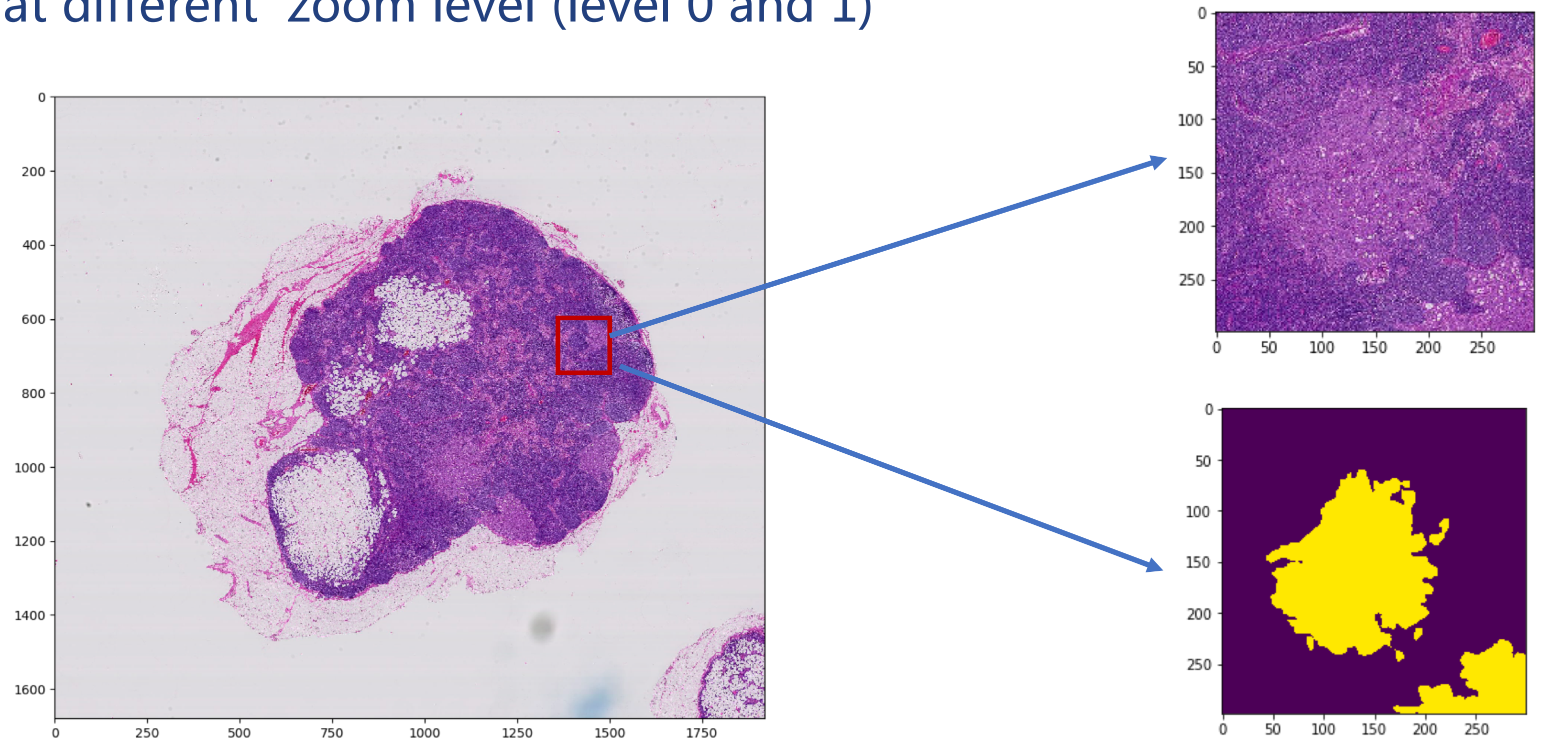


Tumor Mask (Tumor Location)

02-Methods

Data Preprocessing

- Train-Validation-Test Split: 16 Images for train, 2 Images for validation, 3 Images for test
- Create train/validation set: Sample images with size (299, 299) from raw Gigapixel Pathology Images at different zoom level (level 0 and 1)



Tumor Slide

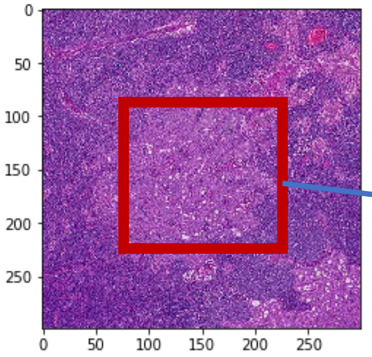
Tumor Mask

02-Methods

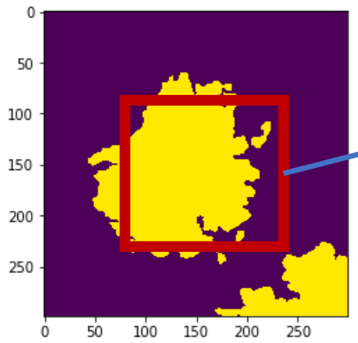
Data Preprocessing

- Balanced dataset: Select "normal" or "tumor" with equal probability.
- Enhanced dataset: Only contains images contains at least 30% tissue.
- Simplified target: If tumor exist in the center 199x199 area, then label it tumor else normal

Tumor Slide



Tumor Mask



If tumor exist, then label it tumor
Else, label it normal

02-Methods

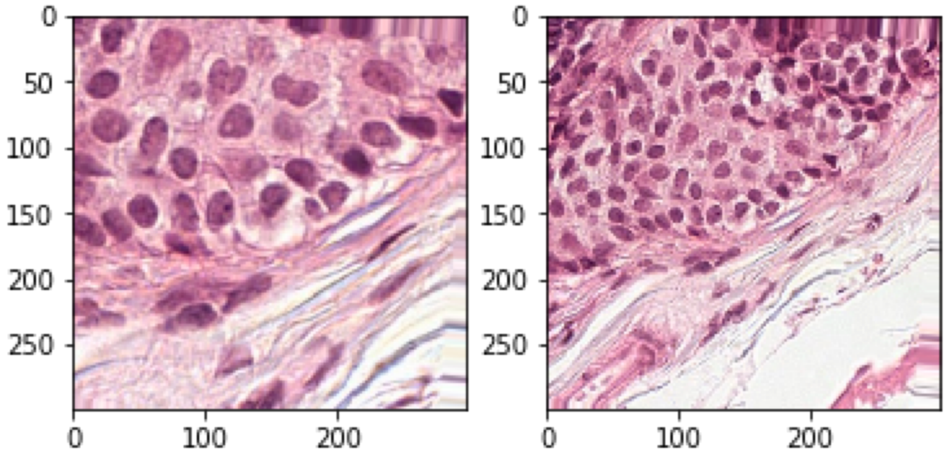
Data Preprocessing

	Raw Data	Sample Data	
Trainset	16 Tumor Slide	4000 299x299 images in zoom level 1 4000 299x299 images in zoom level 2	4000 Labels
Validation Set	2 Tumor Slide	800 299x299 images in zoom level 1 800 299x299 images in zoom level 2	800 Labels

02-Methods

Data Augmentation

- ◆ Use Keras ImageDataGenerator to augment data
  - ◆ Horizontal_flip
  - ◆ Vertical_flip
  - ◆ Rescale
  - ◆ Width_shift
  - ◆ Height_shift
  - ◆ Rotation
- ◆ Use TensorFlow image random to augment data
  - ◆ Random brightness
  - ◆ Random saturation
  - ◆ Random hue
  - ◆ Random contrast

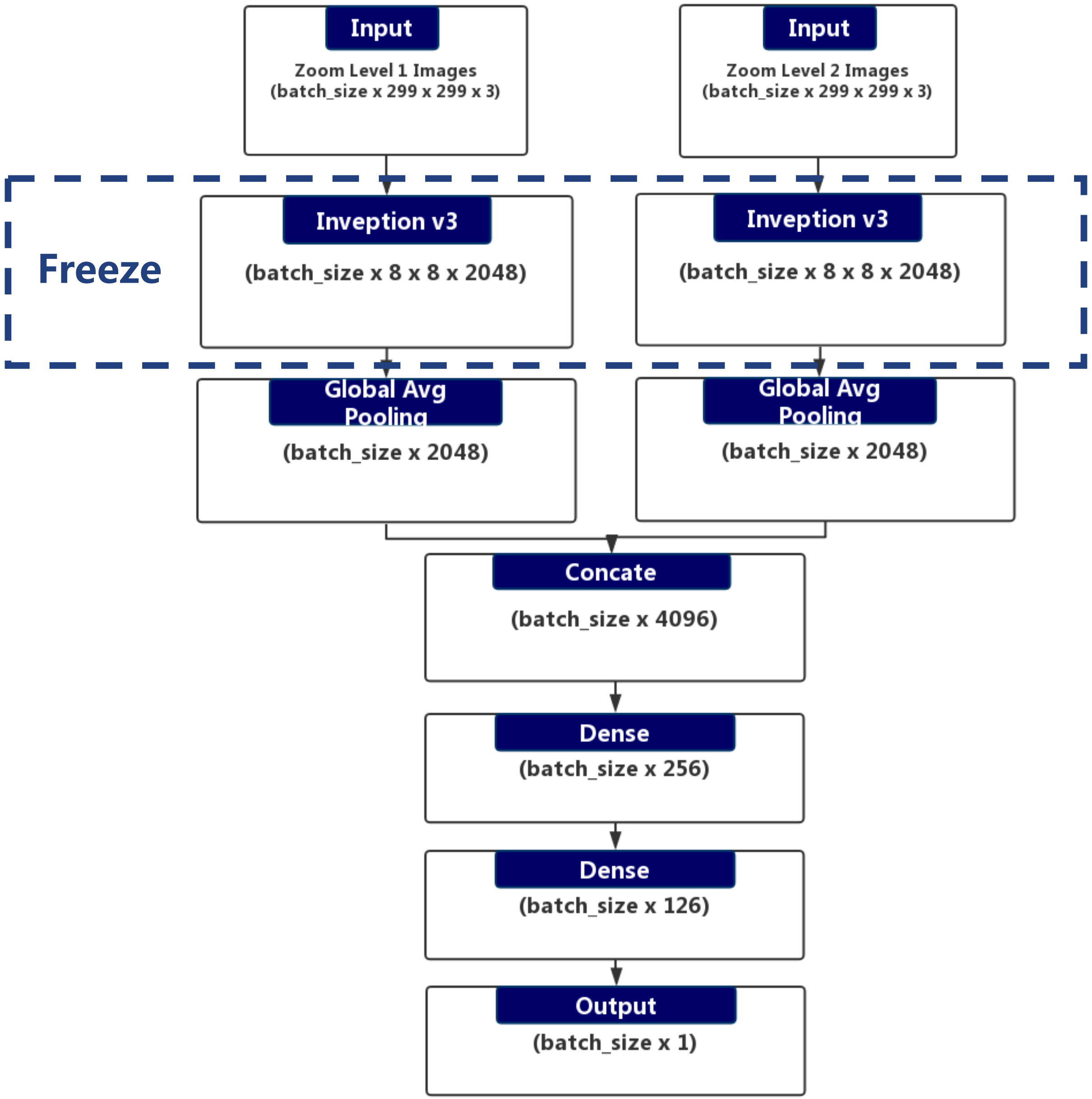


Sample augmentation result

02-Methods

Modeling

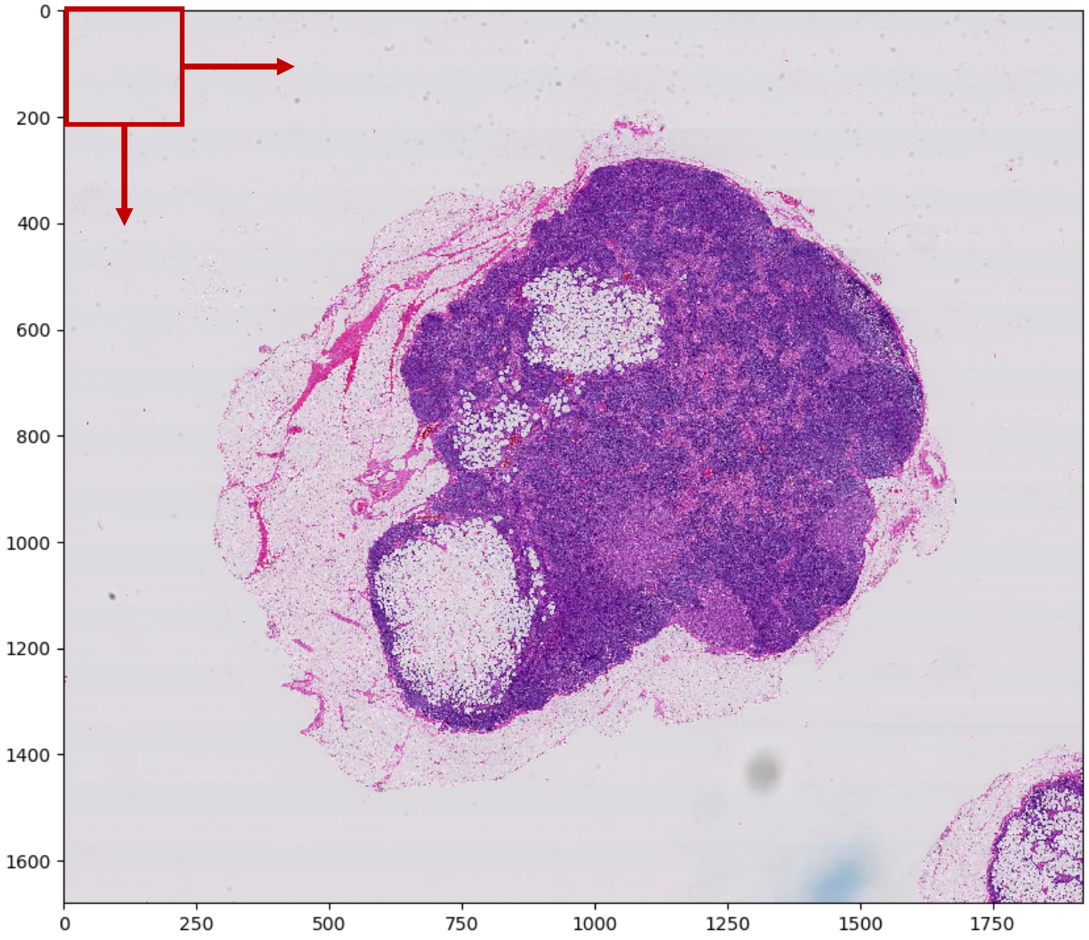
- Use Pretrained inception v3 model on Image
- Use Global Avg Pooling layer to reduce para
- Concate activations from processing two zoo
  images to make joint prediction



02-Methods

Prediction

- ◆ Generate 299 x 299 patches from the test set images rows by rows, columns by columns
- ◆ Make prediction on each patches
- ◆ Set a threshold and make predictions



03- Results

03-Results

Evaluation Matrices

Precision

$$Precision = \frac{true\ positive}{true\ positive + false\ positive}$$

Recall

$$Recall = \frac{true\ positive}{true\ positive + false\ negative}$$

F1 - Score

$$F1\ Score = 2 * \frac{precision * recall}{precision + recal}$$

AUC

Prediction

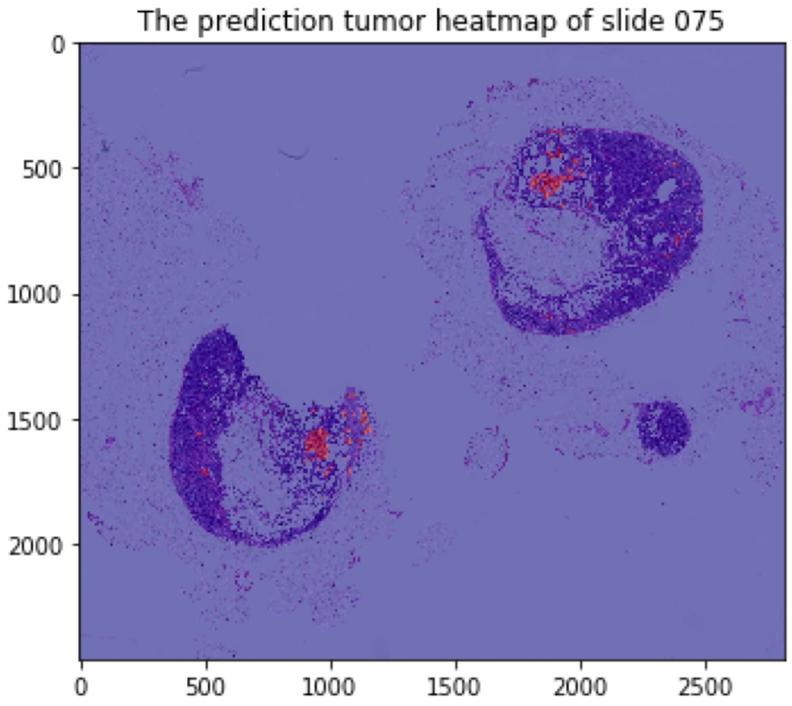
Actual

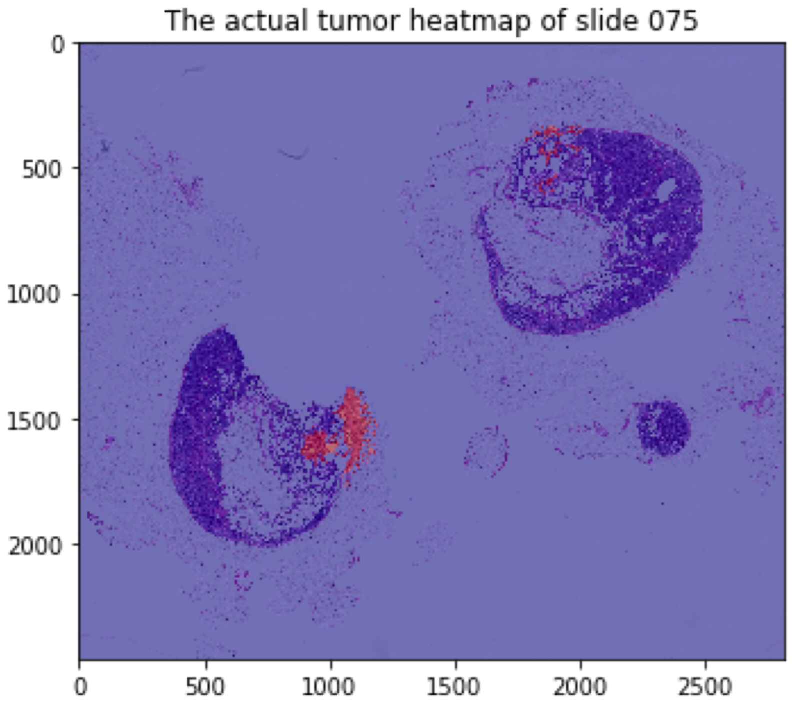
	Tumor	Normal
Tumor	True Positive	False Negative
Normal	False Positive	True Negative

03-Results

Test set evaluation

Matrix	Result
AUC	0.9713
Precision	0.4678
Recall	0.3855
F1 Score	0.4227

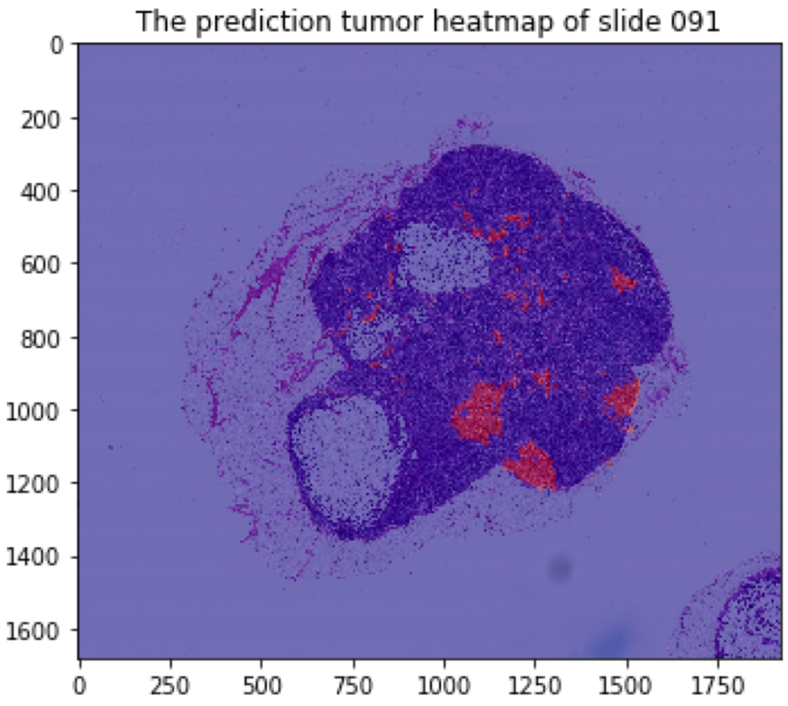


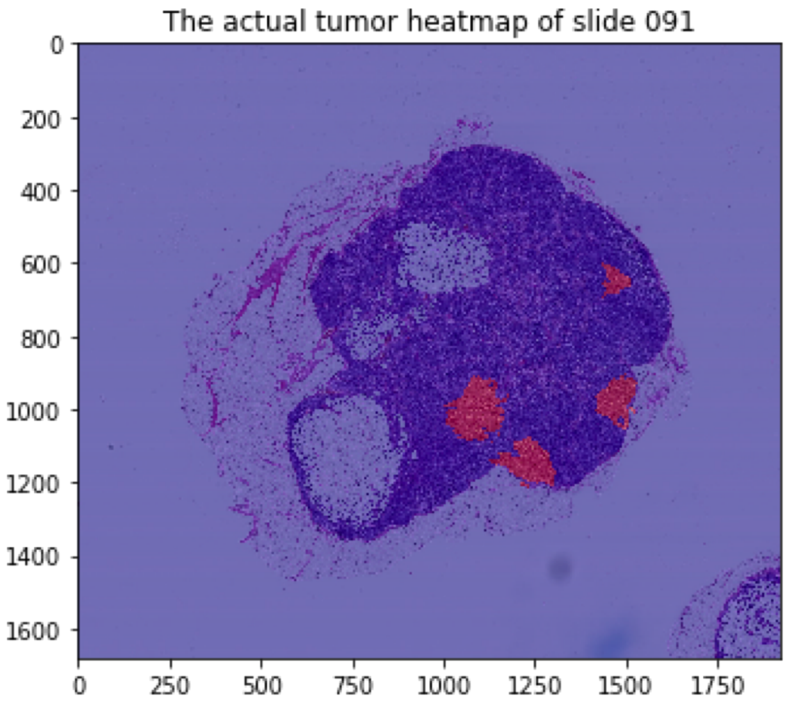


03-Results

Test set evaluation

Matrix	Result
AUC	0.9884
Precision	0.5126
Recall	0.7288
F1 Score	0.6019

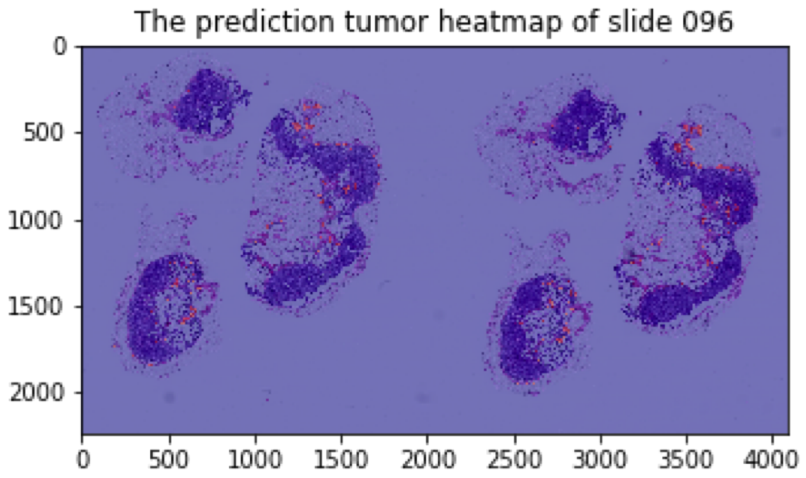


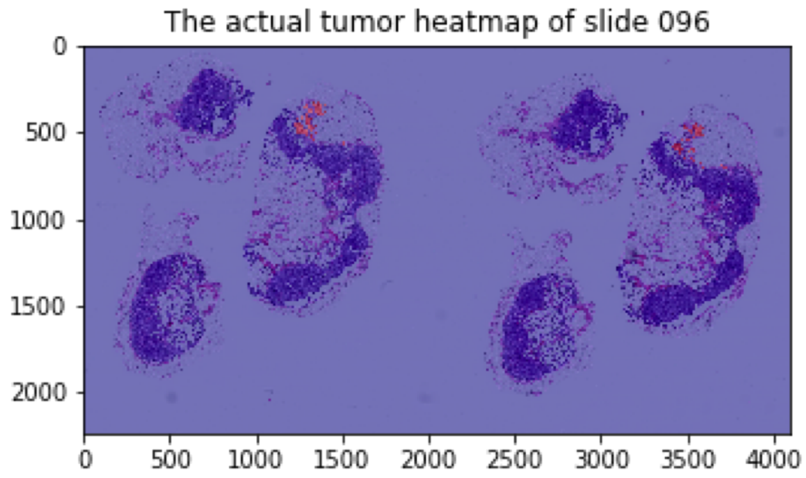


03-Results

Test set evaluation

Matrix	Result
AUC	0.9719
Precision	0.1826
Recall	0.4572
F1 Score	0.2609





04- Conclusions

04-Conclusion

Our prototype works quite, but still have a lot to improve

- Train a new model without using pretrained parameters
- Include more dataset and sample more patches
- Buy more GPUs and RAM! And increase batch size

THANKS