Detecting Cancer in Gigapixel Medical Images

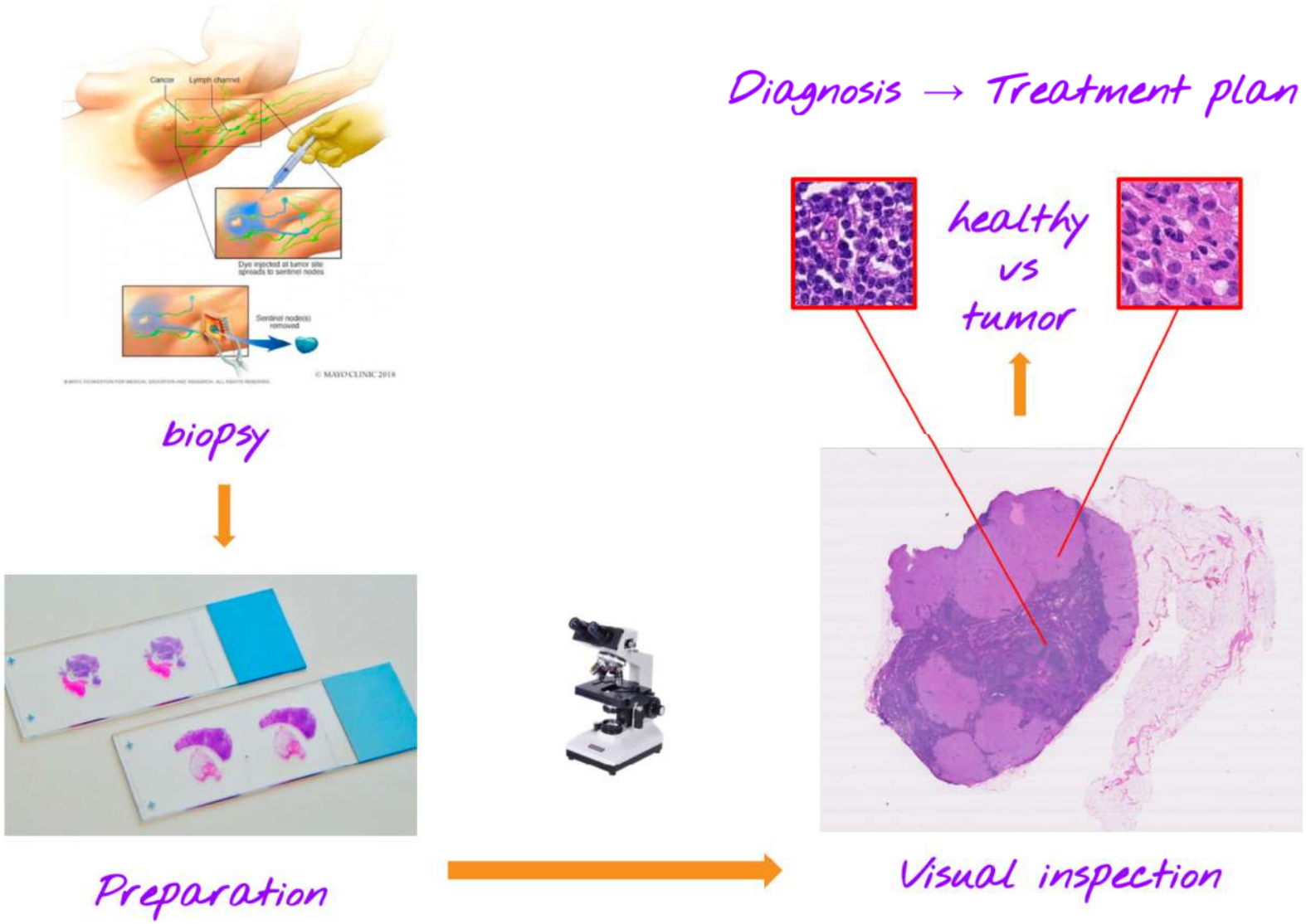
Final Project for Applied Deep Learning

Fall 2018

Aleksandra ‘Ola’ Hosa

Motivation

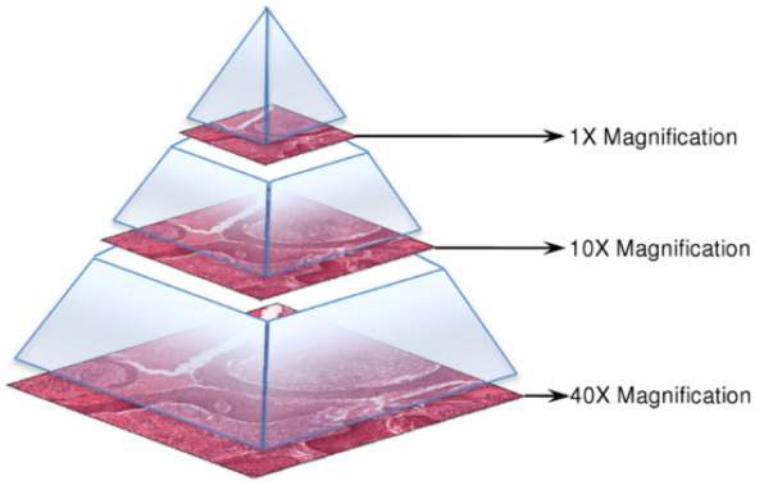
- Pathologists make mistakes when diagnosing biopsied tissue
- Assist and not automate and replace (!) human expertise
- Reduce misdiagnosis rate

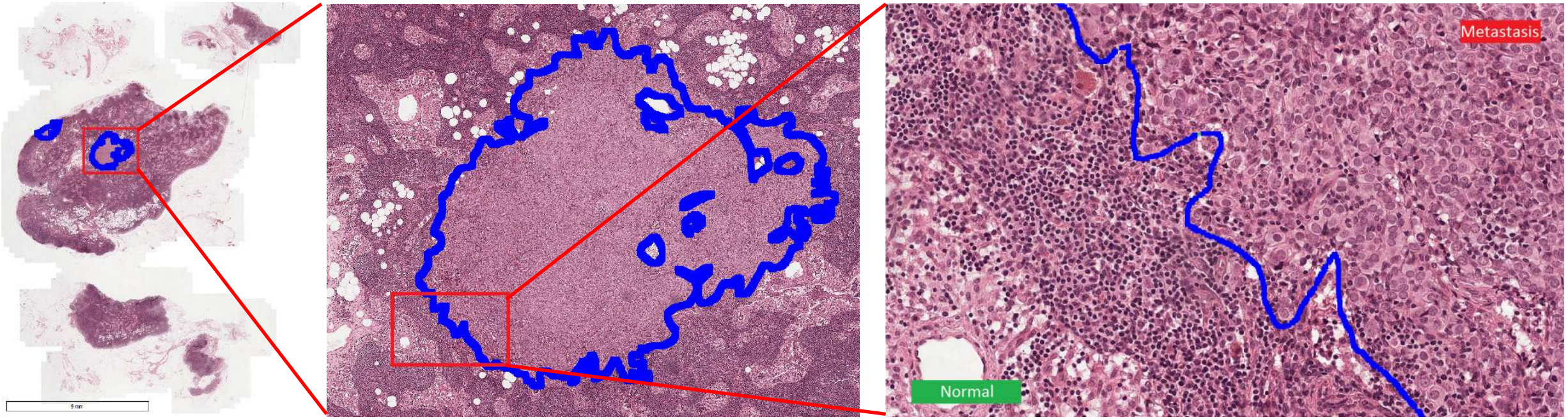


from ADL lecture 6 slides by Joshua Gordon

Data Source

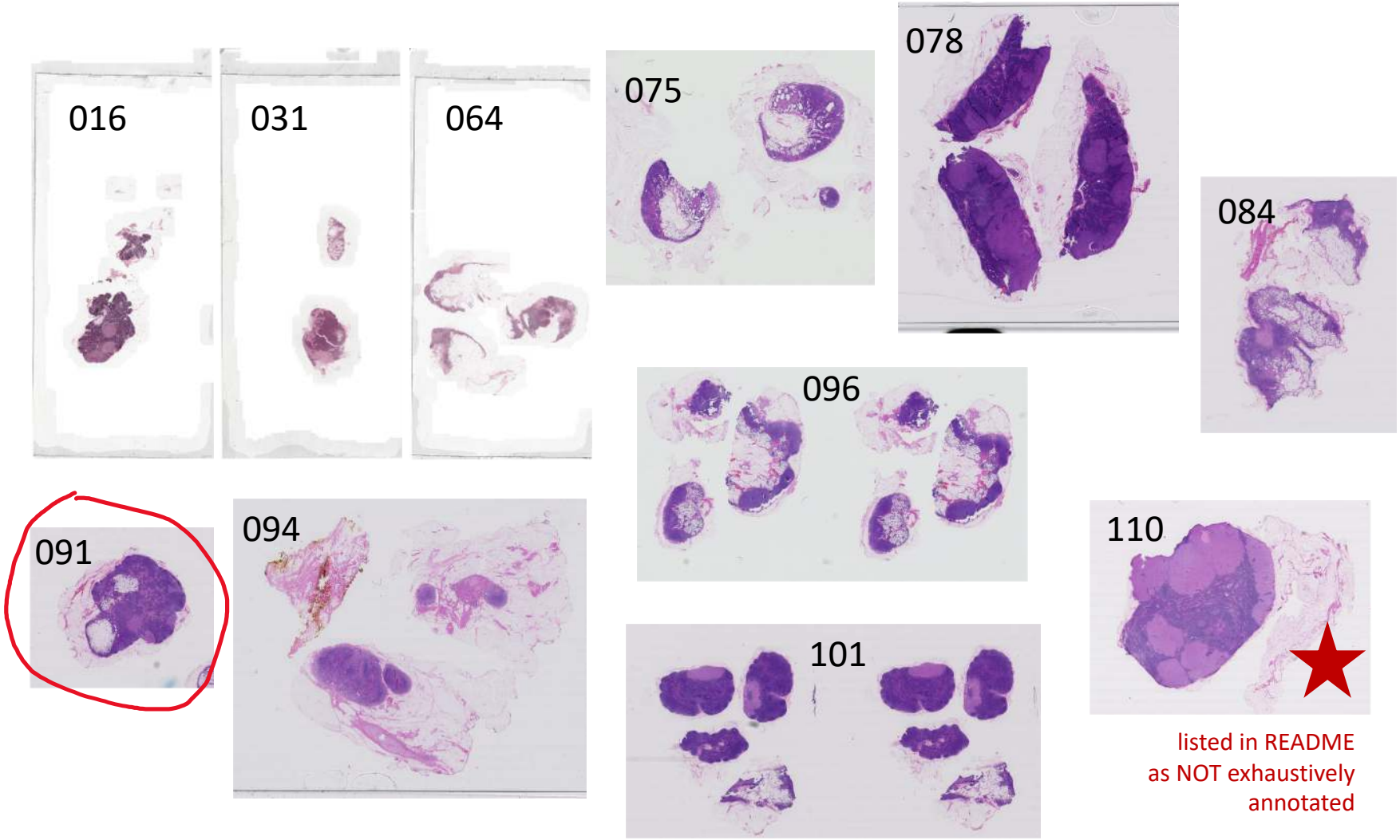
- CAMELYON16 challenge
- 400 images from 2 institutions (~600GB)
- 8 levels of magnification (0 to 7)





Chosen Images

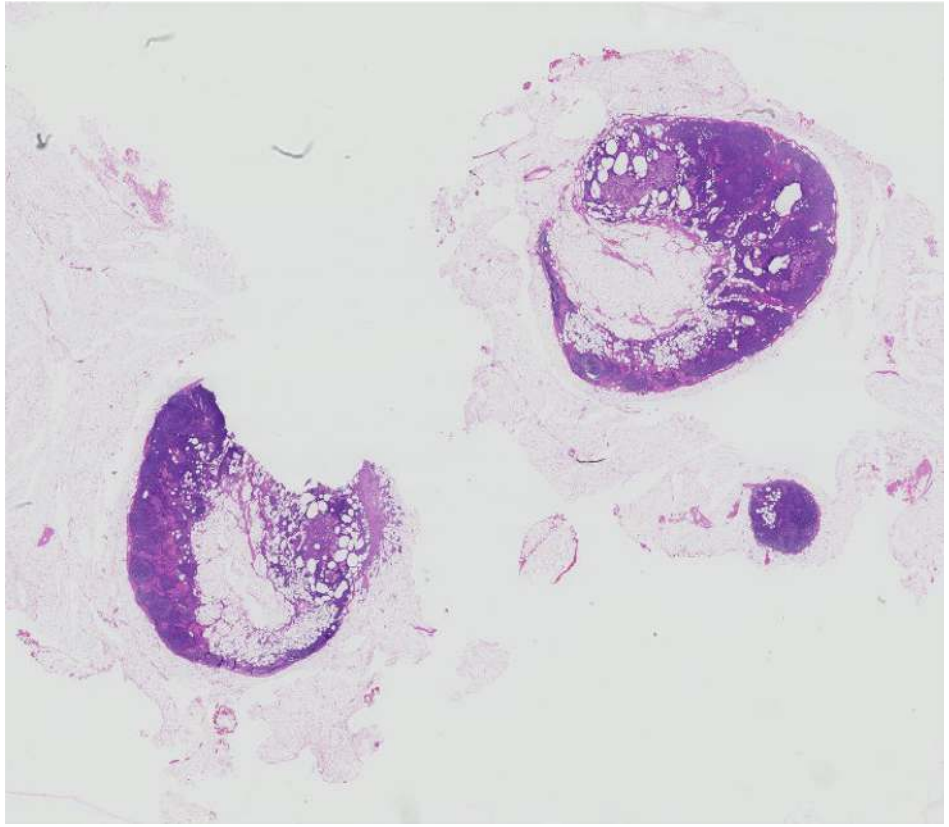
- Out of given 22 images I chose 11 with highest content of cancer
- 9 training: 016, 031, 064, 075, 078, 084, 094, 096, 101
- 1 test: 091
- Slide 110 (not exhaustively annotated) – predict heatmap



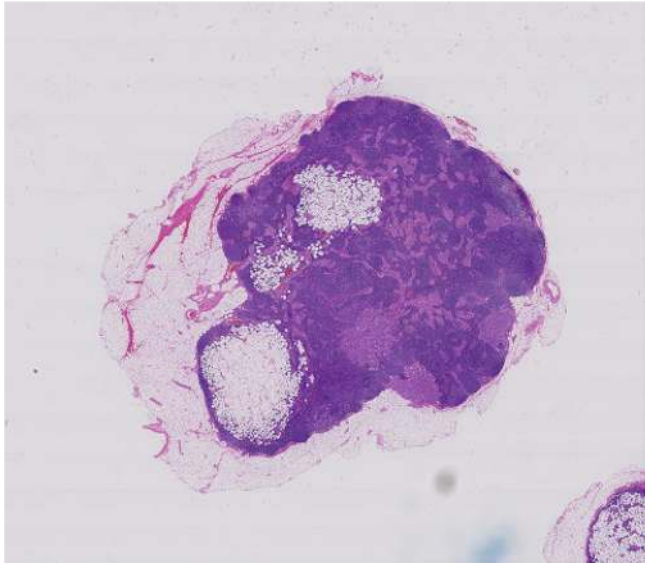
listed in README as NOT exhaustively annotated

Tumor Masks

Slide 075 (training)

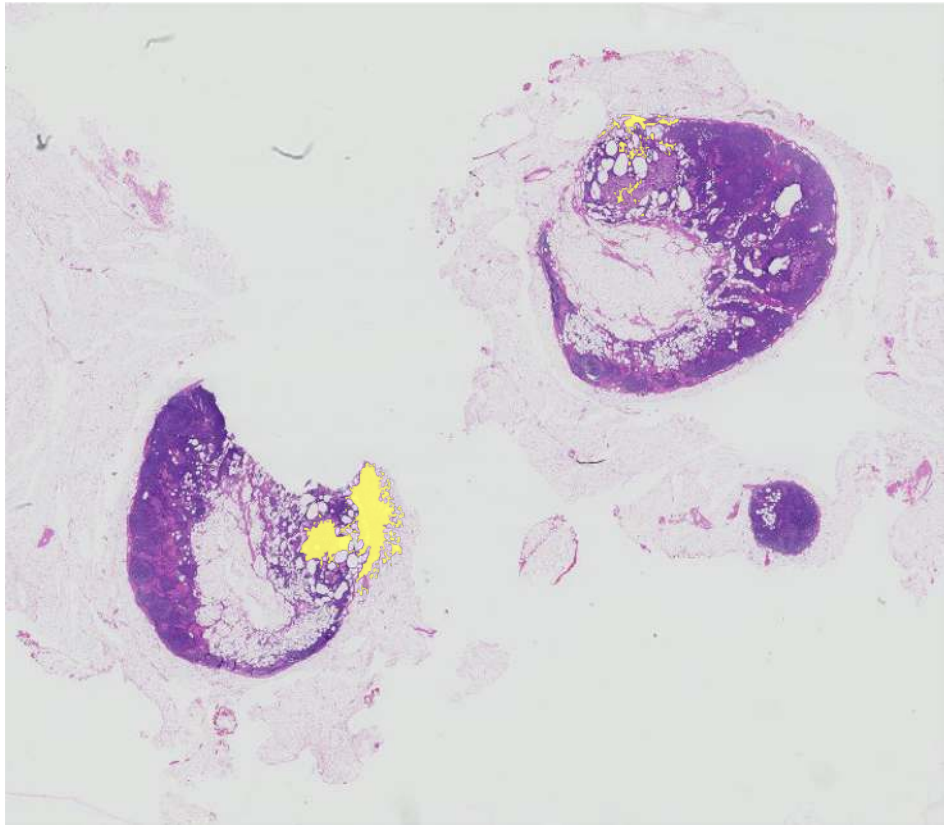


Slide 091 (test)

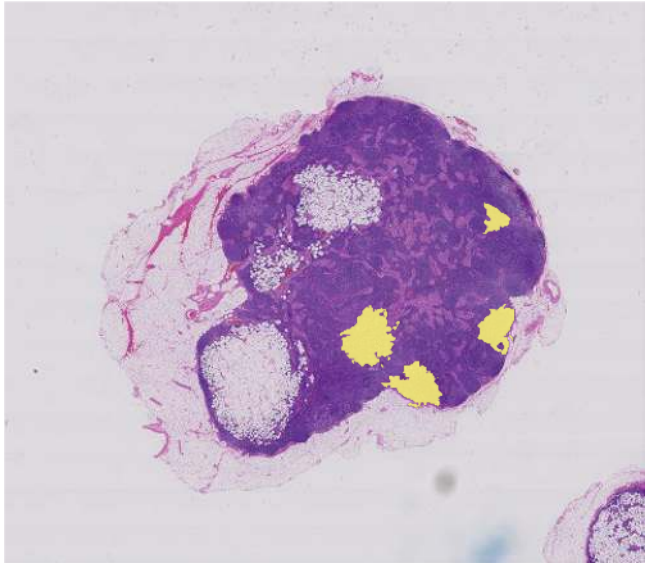


Tumor Masks

Slide 075 (training)



Slide 091 (test)



tumor shown in yellow

Methodology

Detecting Cancer Metastases on Gigapixel Pathology Images

Yun Liu[1*], Krishna Gadepalli[1], Mohammad Norouzi[1], George E. Dahl[1], Timo Kohlberger[1], Aleksey Boyko[1], Subhashini Venugopalan[2**], Aleksei Timofeev[2], Philip Q. Nelson[2], Greg S. Corrado[1], Jason D. Hipp[3], Lily Peng[1], and Martin C. Stumpe[1]

{liuyun,mnorouzi,gdahl,lhpeng,mstumpe}@google.com

[1]Google Brain, [2]Google Inc, [3]Verily Life Sciences, Mountain View, CA, USA

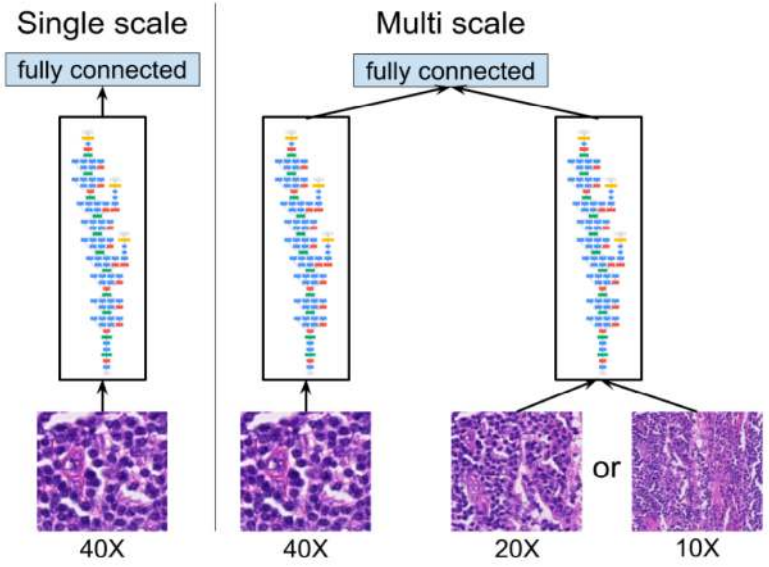


Fig. 3. The three colorful blocks represent Inception (V3) towers up to the second-last layer (PreLogit). *Single scale* utilizes one tower with input images at 40X magnification; *multi-scale* utilizes multiple (*e.g.*,2) input magnifications that are input to separate towers and merged.

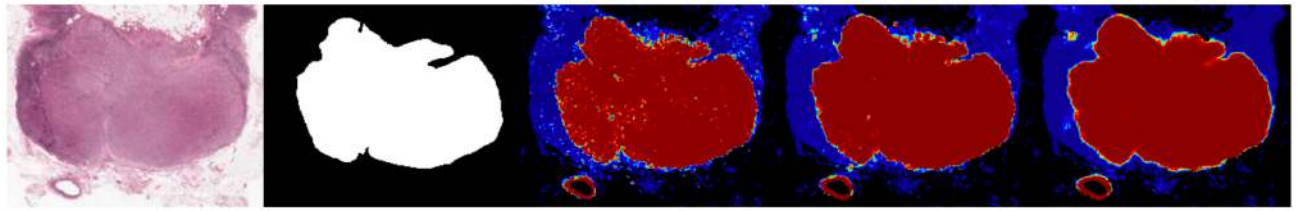
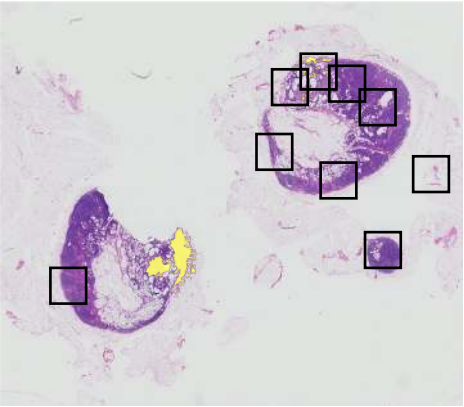


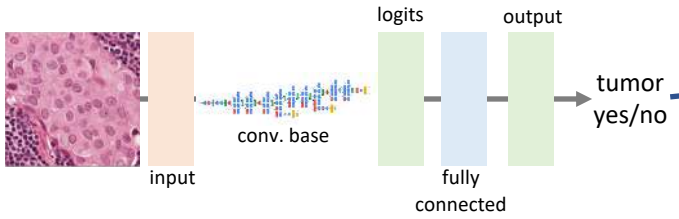
Fig. 4. Left to right: sample image, ground truth (tumor in white), and heatmap outputs (40X-ensemble-of-3, 40X+20X, and 40X+10X). Heatmaps of 40X and 40X-ensemble-of-3 look identical. The red circular regions at the bottom left quadrant of the heatmaps are unannotated tumor. Some of the speckles are either out of focus patches on the image or non-tumor patches within a large tumor.

Methodology Overview

Patch extraction for training and testing (random)

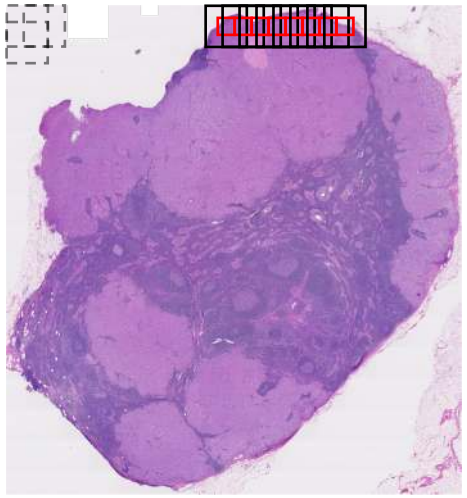


Training with a deep model

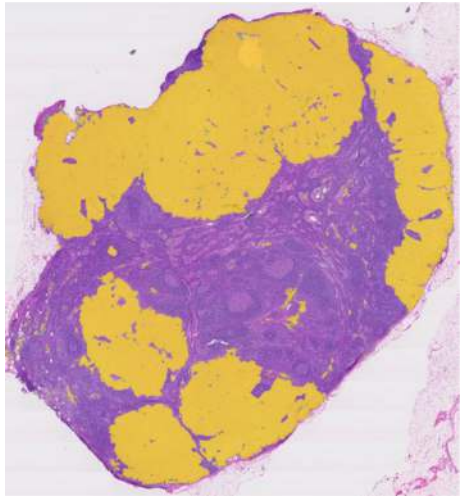


model evaluation on test patches (accuracy)

Patch extraction for heatmap (sliding)



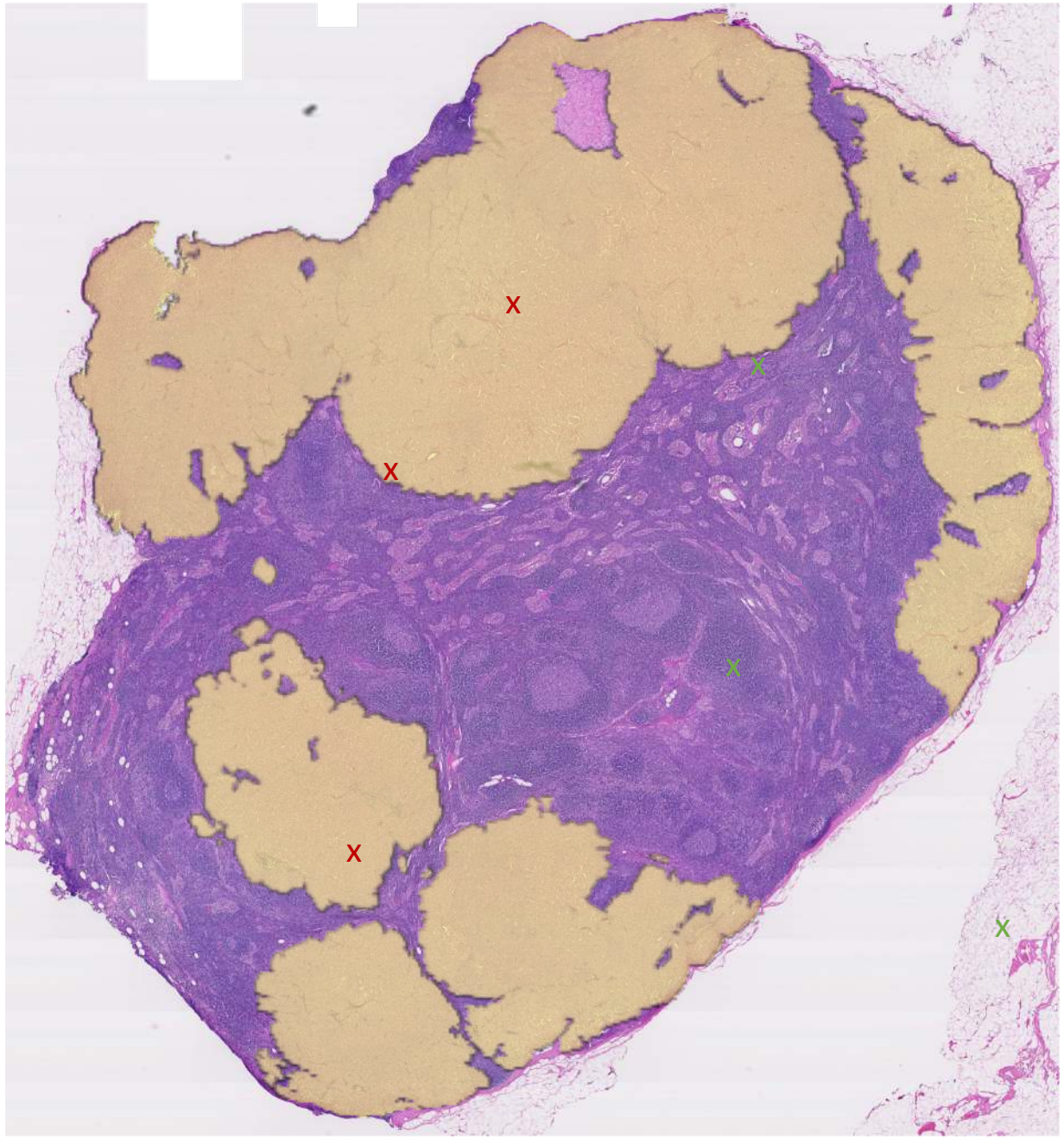
Heatmap generation



heatmap evaluation (precision, recall, f1)

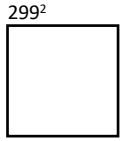
Extraction of Patches for Training

- Choose a **random** sample of tumor pixels and of healthy tissue pixels



slide 110, level 4

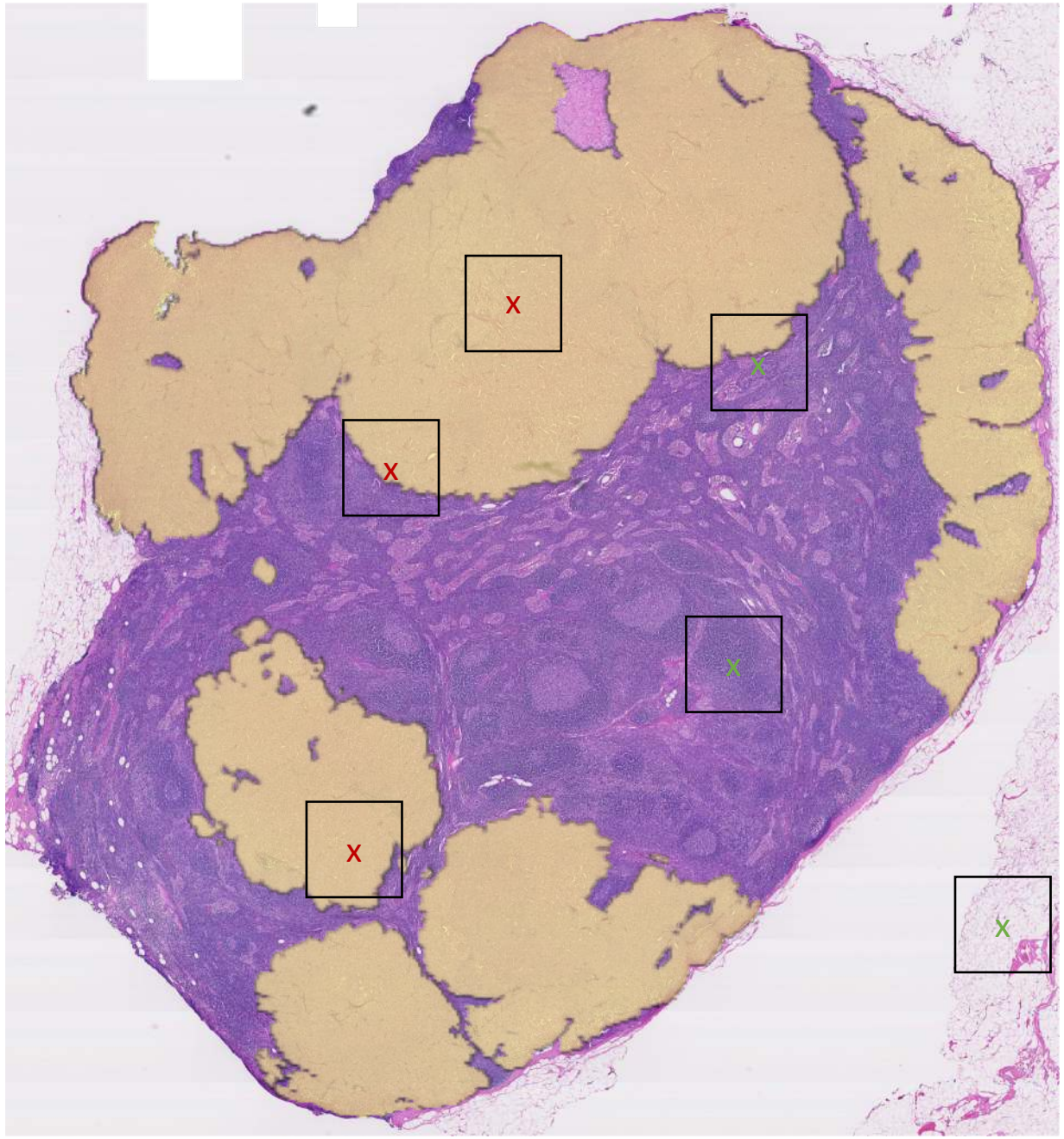
Extraction of Patches for Training



- Choose a **random** sample of tumor pixels and of healthy tissue pixels

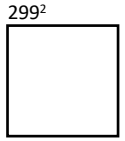
For pixels in the random sample do:

- Extract a 299x299 patch centered at a sampled pixel at a target magnification level (model input)



slide 110, level 4

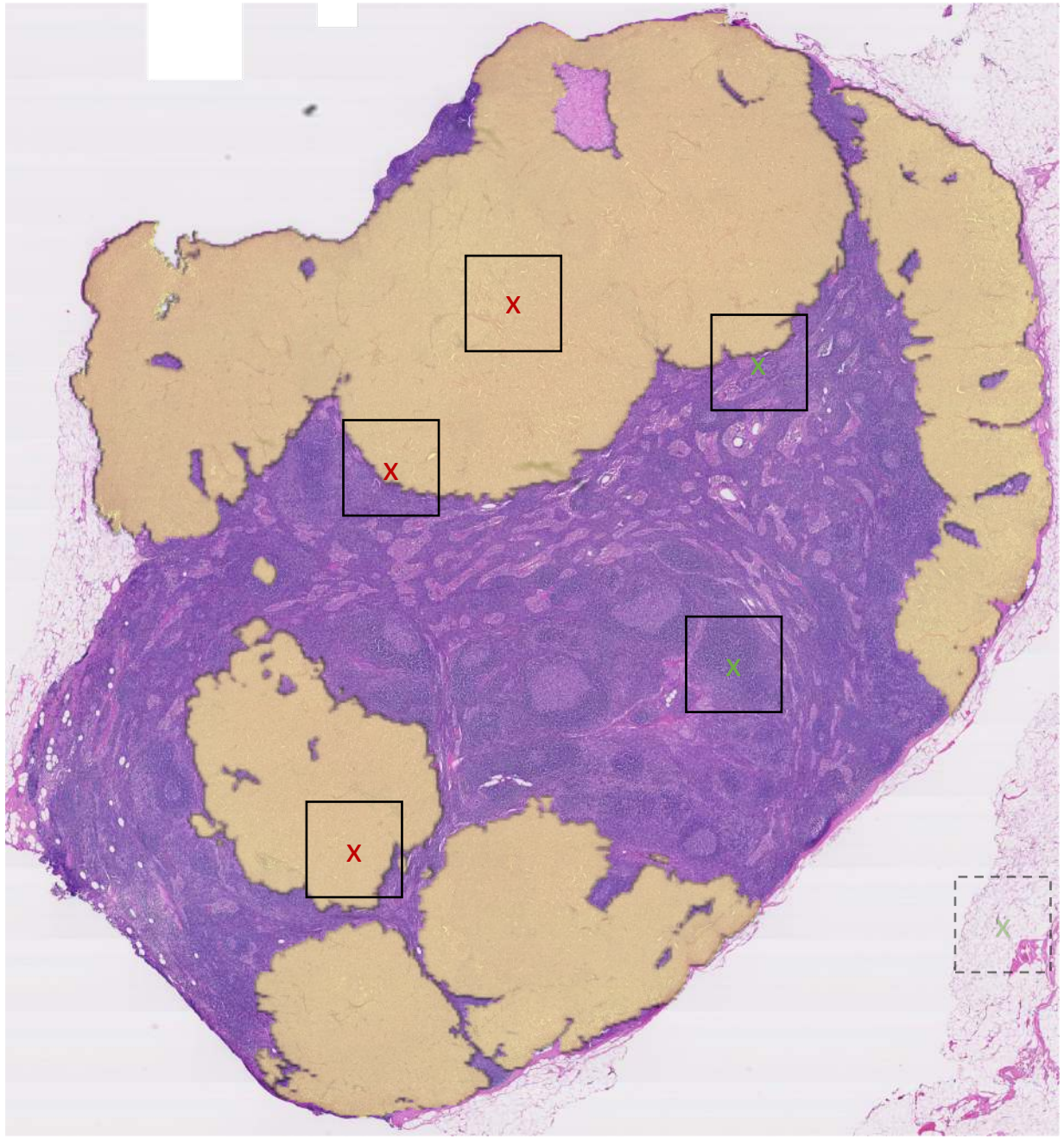
Extraction of Patches for Training



- Choose a **random** sample of tumor pixels and of healthy tissue pixels

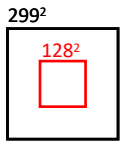
For pixels in the random sample do:

- Extract a 299x299 patch centered at a sampled pixel at a target magnification level (model input)
- Check if that patch has > 50% tissue (defined as > 0.8 grayscale value):
  if no:
  - reject the patch
  if yes:
  - continue



slide 110, level 4

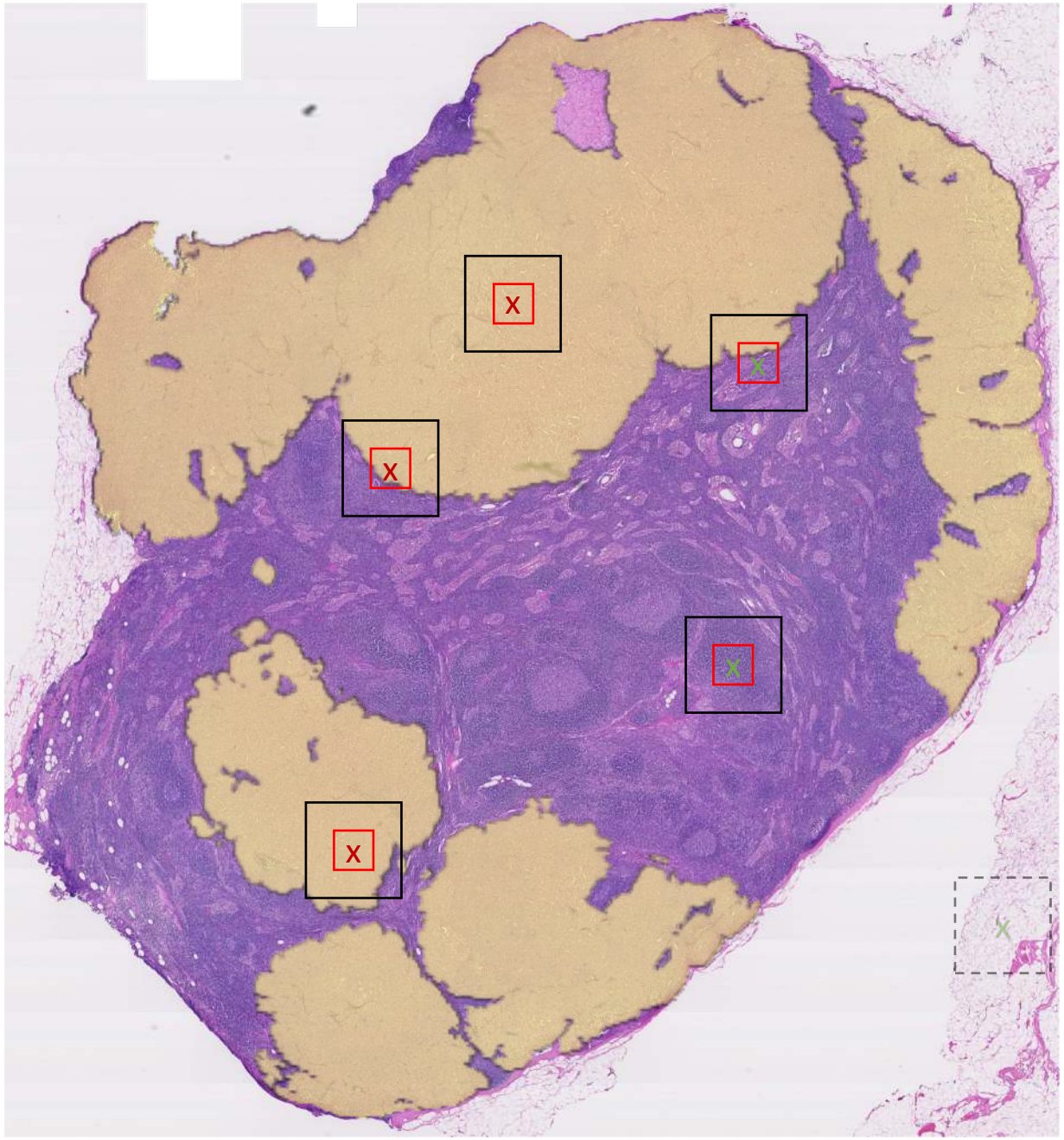
Extraction of Patches for Training



- Choose a **random** sample of tumor pixels and of healthy tissue pixels

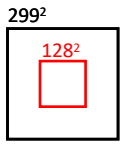
For pixels in the random sample do:

- Extract a 299x299 patch centered at a sampled pixel at a target magnification level (model input)
- Check if that patch has > 50% tissue (defined as > 0.8 grayscale value):
  if no:
  - reject the patch
  if yes:
  - continue
- Check if the 128x128 center includes any tumor cell:
  if yes:
  - for tumor pixel – accept and collect the patch with label '1'
  - for healthy pixel – reject the patch
  if no:
  - for healthy pixel – accept and collect the patch with label '0'



slide 110, level 4

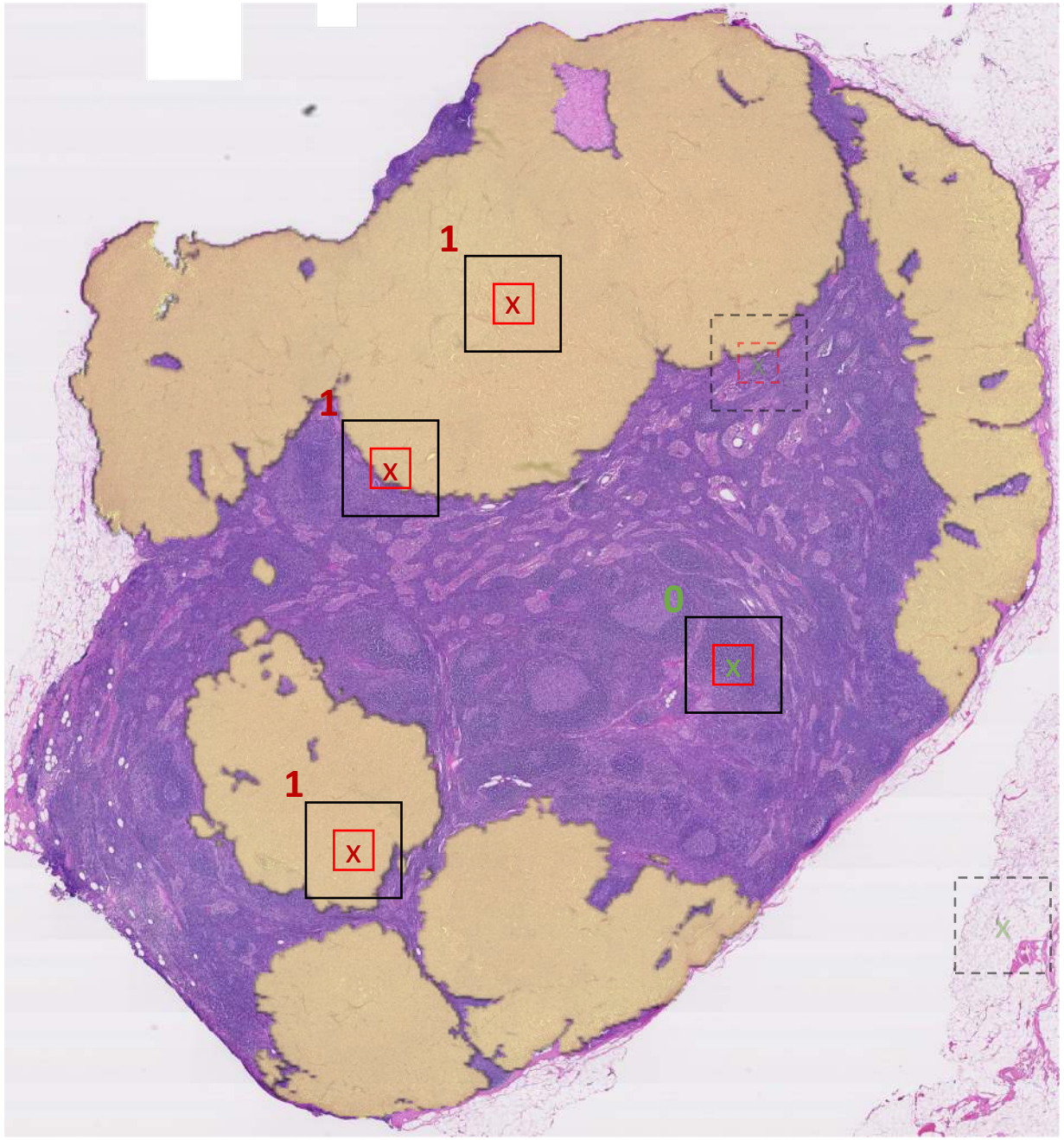
Extraction of Patches for Training



- Choose a **random** sample of tumor pixels and of healthy tissue pixels

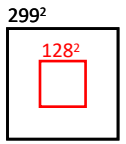
For pixels in the random sample do:

- Extract a 299x299 patch centered at a sampled pixel at a target magnification level (model input)
- Check if that patch has > 50% tissue (defined as > 0.8 grayscale value):
  if no:
  - reject the patch
  if yes:
  - continue
- Check if the 128x128 center includes any tumor cell:
  if yes:
  - for tumor pixel – accept and collect the patch with label '1'
  - for healthy pixel – reject the patch
  if no:
  - for healthy pixel – accept and collect the patch with label '0'



slide 110, level 4

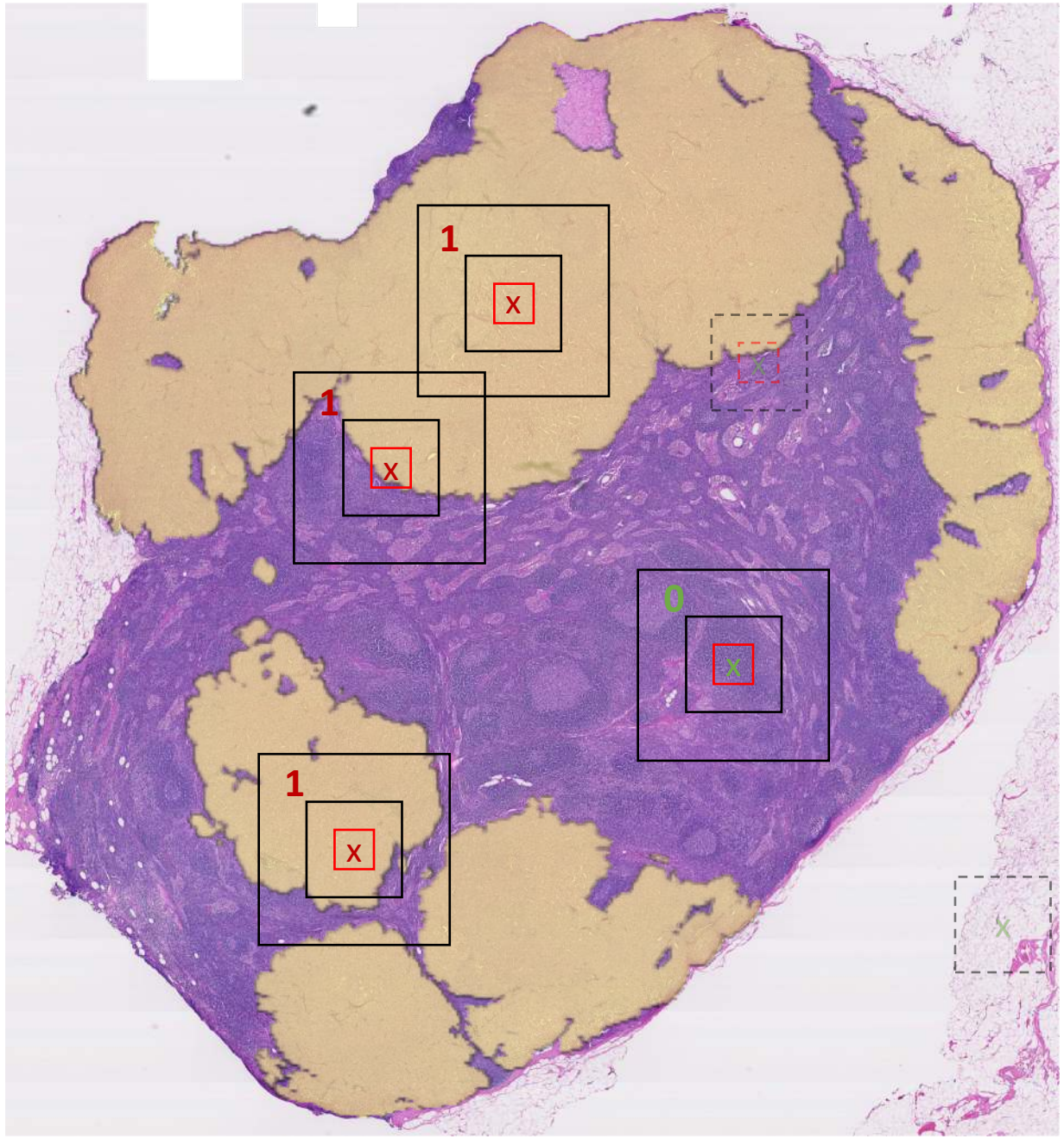
Extraction of Patches for Training



- Choose a **random** sample of tumor pixels and of healthy tissue pixels

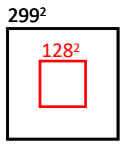
For pixels in the random sample do:

- Extract a 299x299 patch centered at a sampled pixel at a target magnification level (model input)
- Check if that patch has > 50% tissue (defined as > 0.8 grayscale value):
  if no:
  - reject the patch
  if yes:
  - continue
- Check if the 128x128 center includes any tumor cell:
  if yes:
  - for tumor pixel – accept and collect the patch with label '1'
  - for healthy pixel – reject the patch
  if no:
  - for healthy pixel – accept and collect the patch with label '0'
- In multiple level model, for every accepted patch, also collect a patch at a secondary (zoomed-out) level



slide 110, level 4

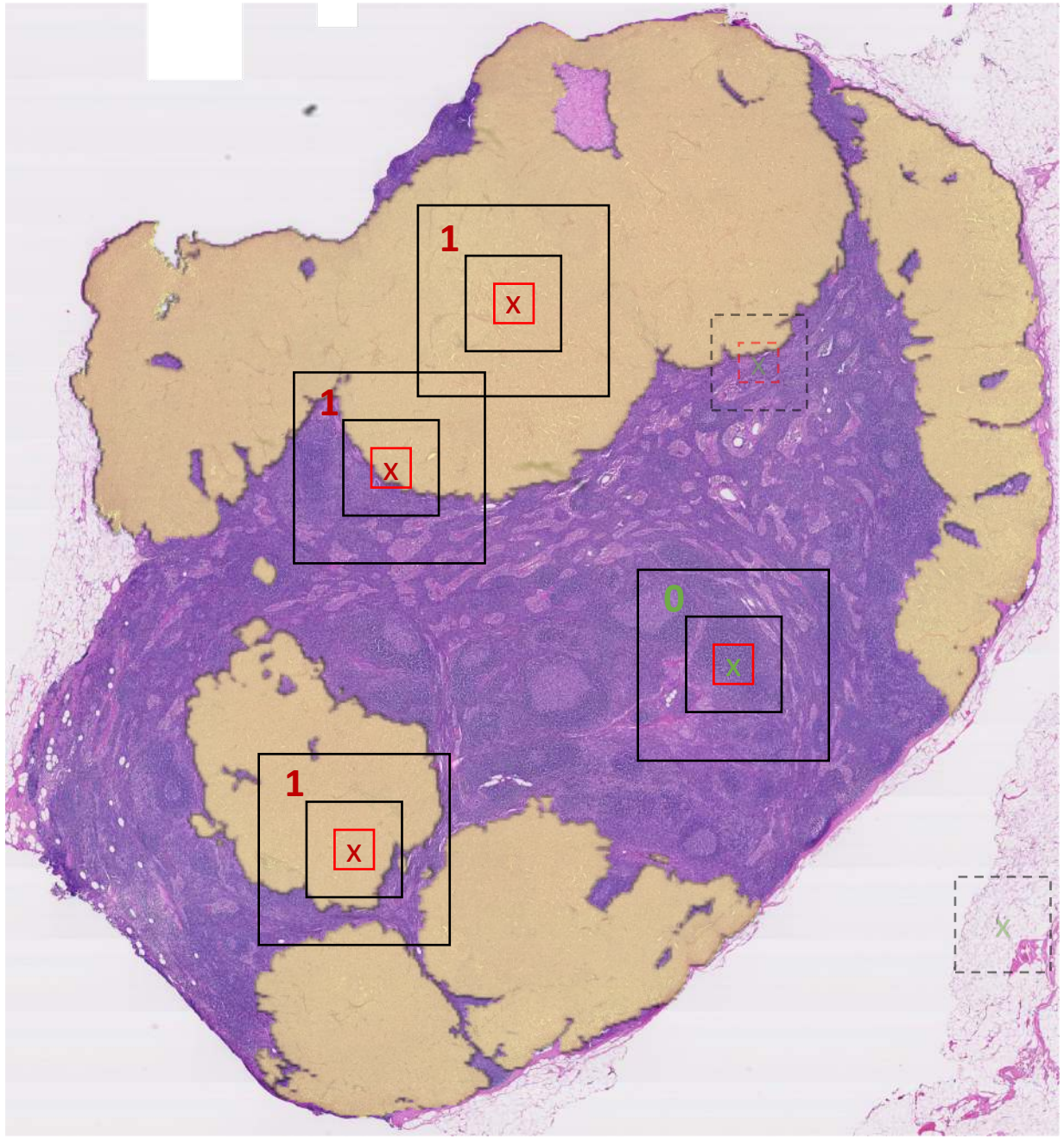
Extraction of Patches for Training



- Choose a **random** sample of tumor pixels and of healthy tissue pixels

For pixels in the random sample do:

- Extract a 299x299 patch centered at a sampled pixel at a target magnification level (model input)
- Check if that patch has > 50% tissue (defined as > 0.8 grayscale value):
  if no:
  - reject the patch
  if yes:
  - continue
- Check if the 128x128 center includes any tumor cell:
  if yes:
  - for tumor pixel – accept and collect the patch with label '1'
  - for healthy pixel – reject the patch
  if no:
  - for healthy pixel – accept and collect the patch with label '0'
- In multiple level model, for every accepted patch, also collect a patch at a secondary (zoomed-out) level
- Continue until 100 patches collected (both with and w/o tumor)

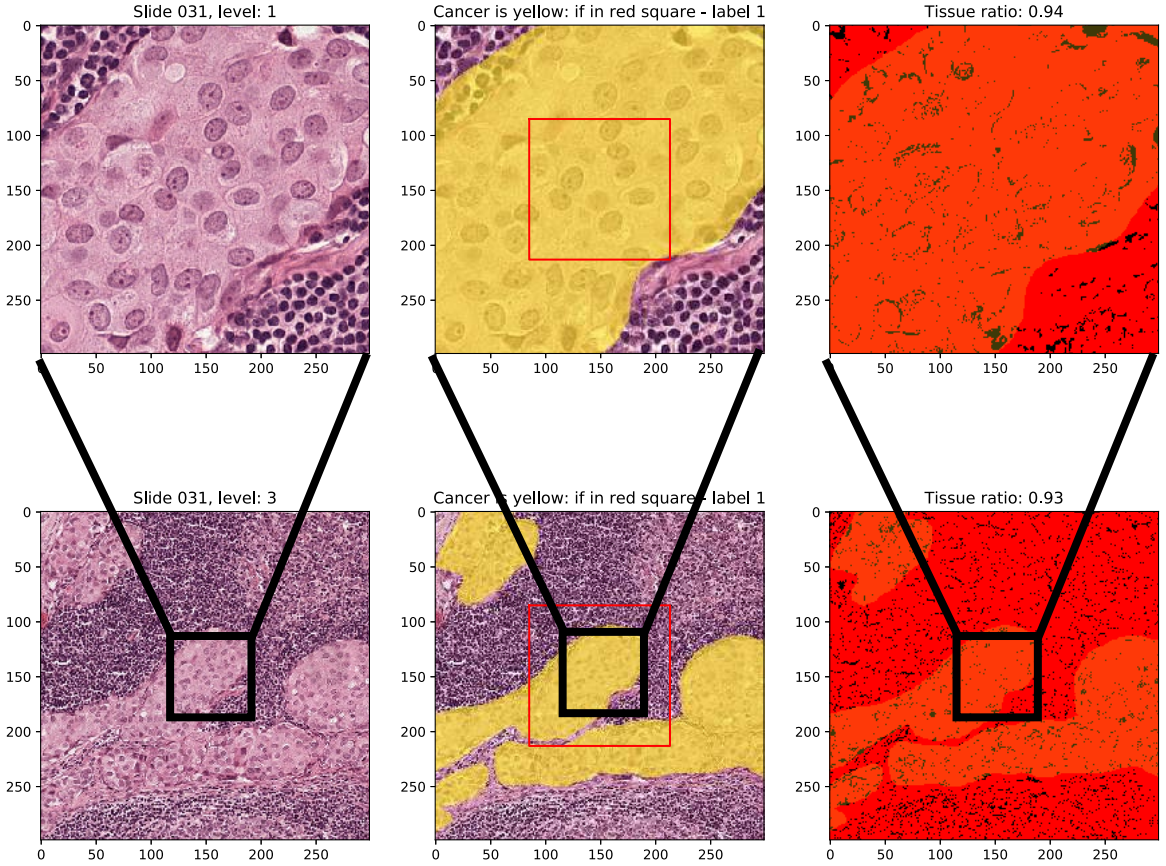


slide 110, level 4

Examples of Patches

Slide 031, tumor

level 1



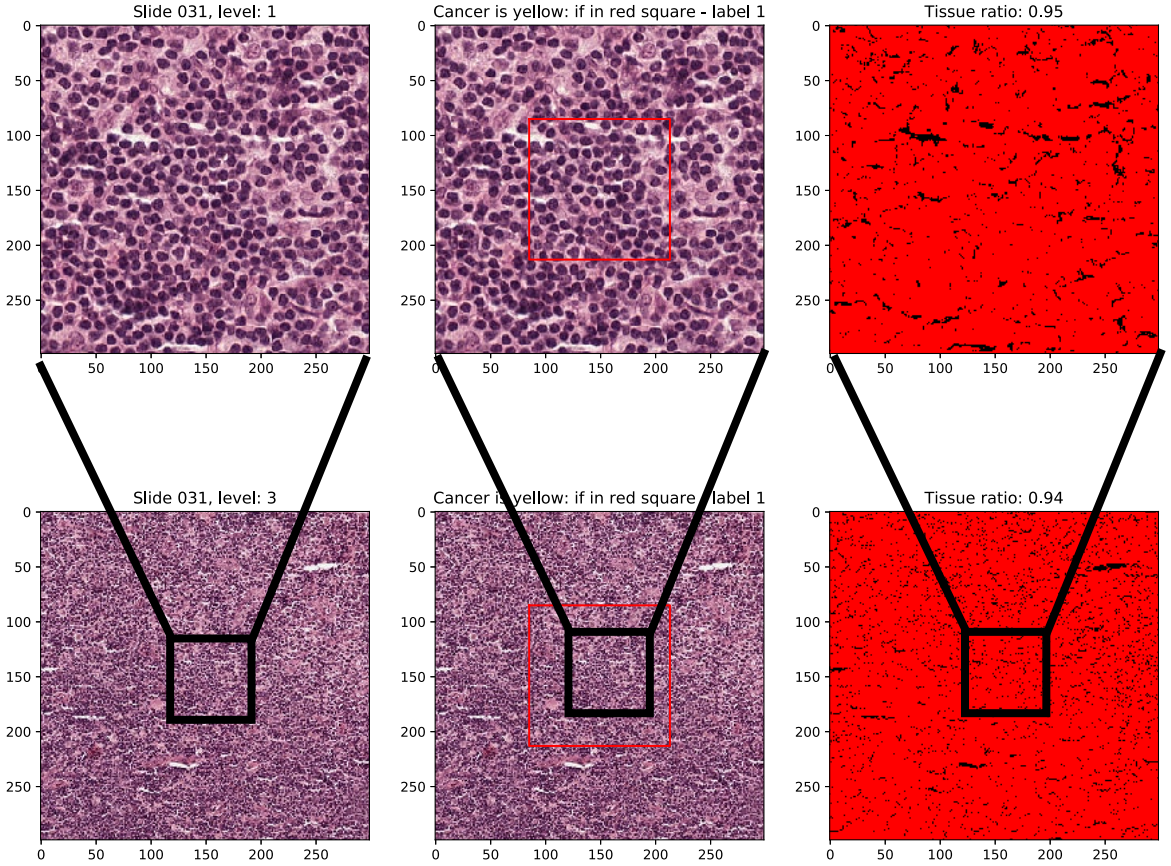
level 3

tumor shown in yellow

tissue shown in red

Slide 031, healthy cells

level 1



level 3

tumor shown in yellow

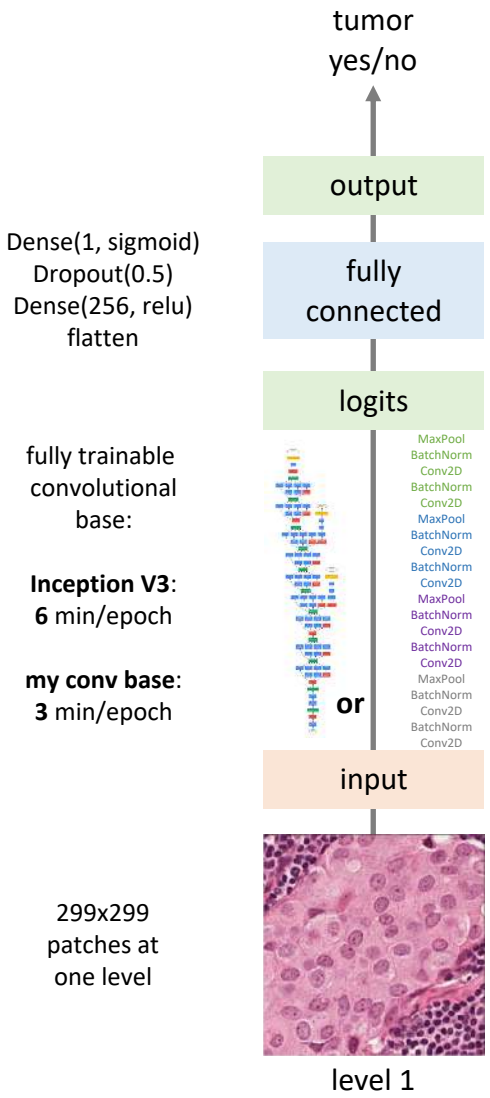
tissue shown in red

Models

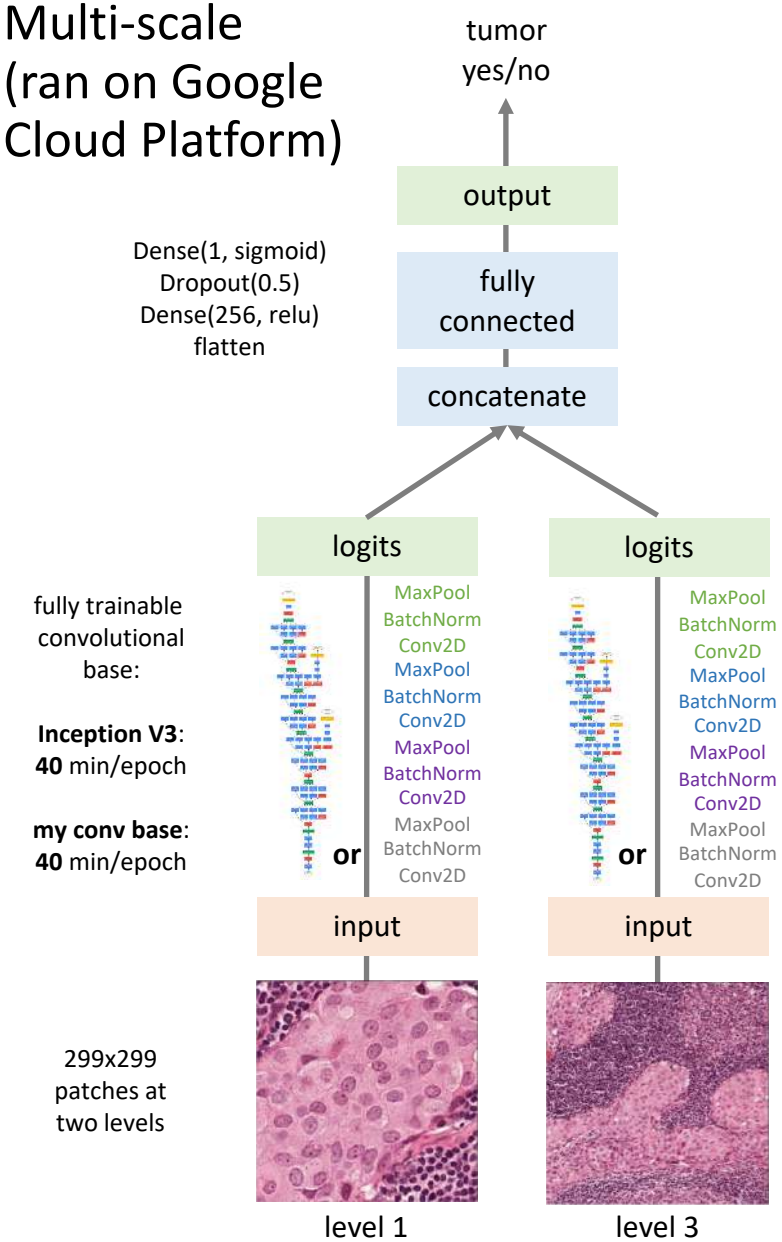
Model choices:

- Single level or multiple levels of input
- Target level of magnification: 0 to 7
- Convolutional base: Inception or smaller base defined by me

Single scale (ran on Colab)

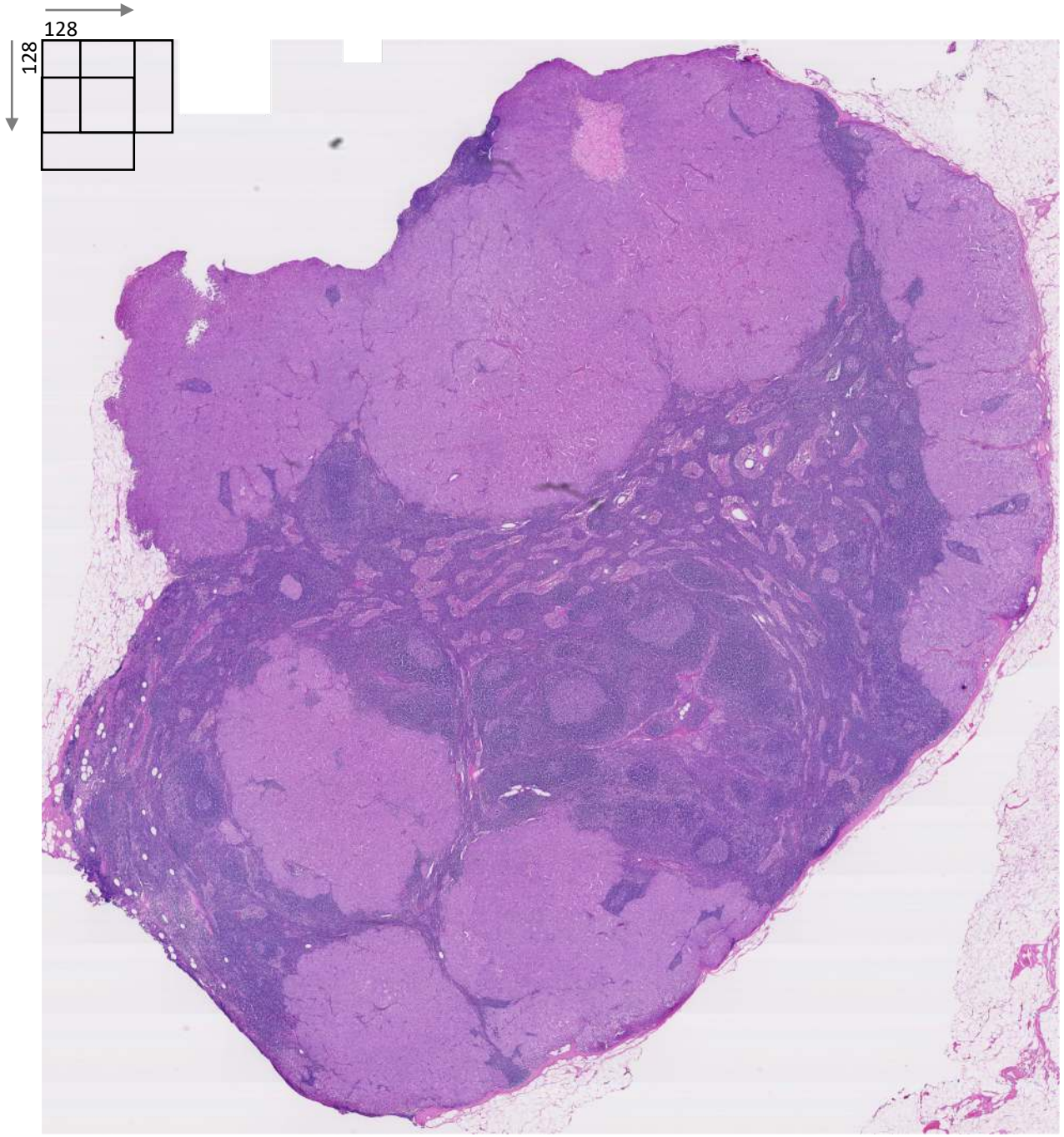


Multi-scale (ran on Google Cloud Platform)



Heatmap Generation

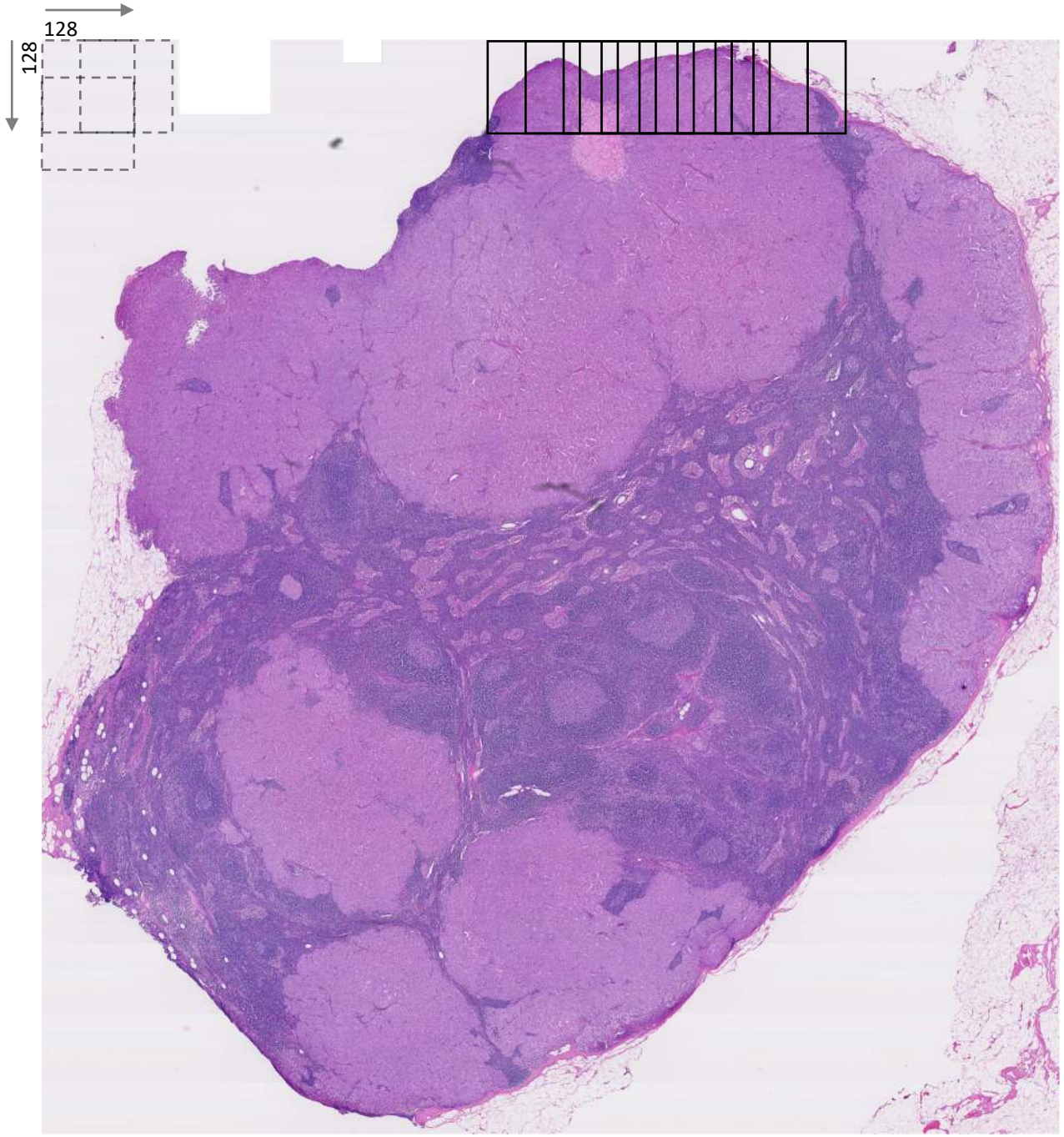
- Generate a heatmap at level 7 (maximally zoomed-out) – initialize with zeros
- Proceed through the slide by stride of 128 pixels in x and y, getting 299x299 patches at the target levels



slide 110, size of patches at level 4

Heatmap Generation

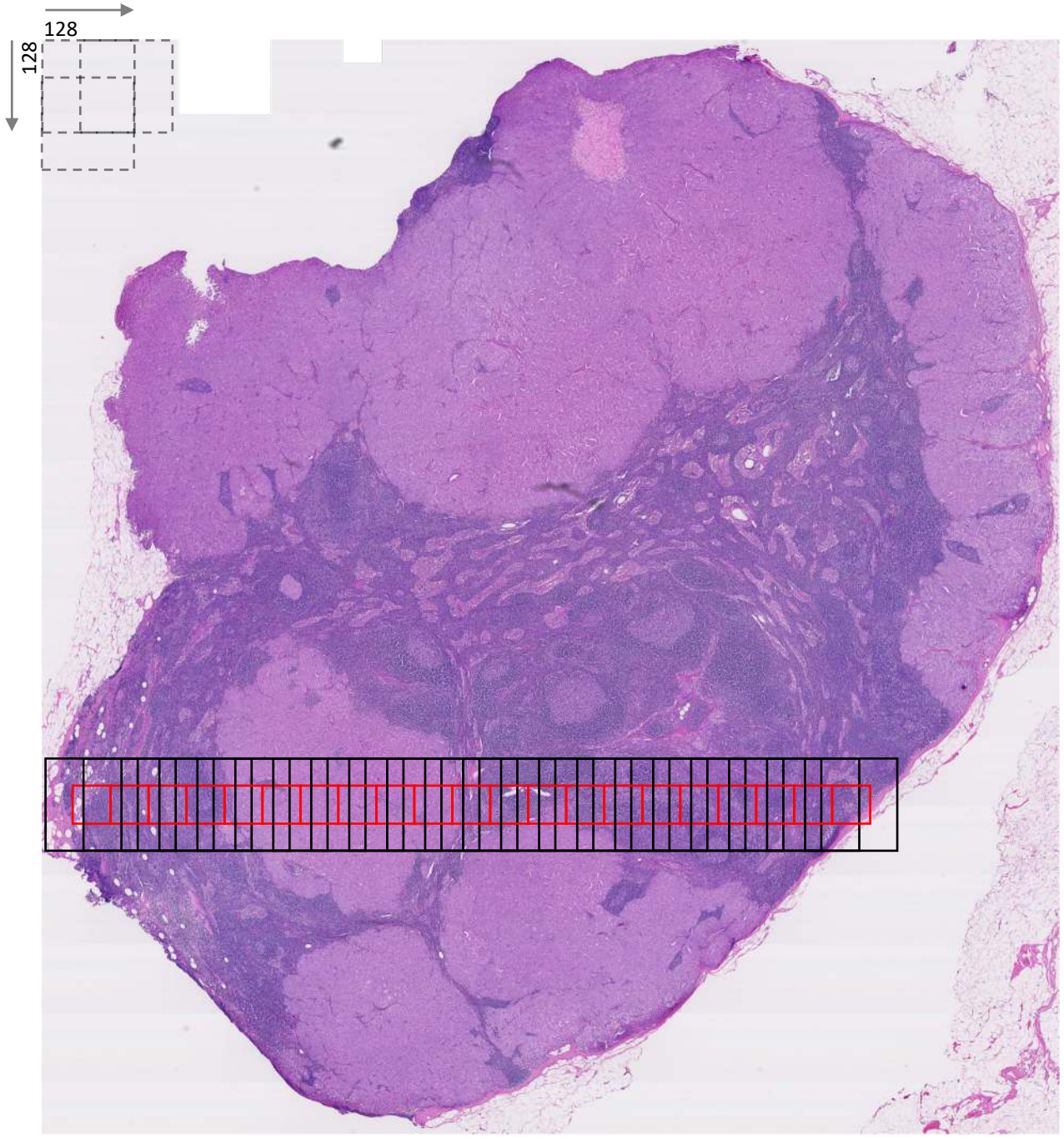
- Generate a heatmap at level 7 (maximally zoomed-out) – initialize with zeros
- Proceed through the slide by stride of 128 pixels in x and y, getting 299x299 patches at the target levels
- Skip patches with less than 50% tissue
- To speed-up, gather 1000 patches and do the prediction on that batch



slide 110, size of patches at level 4

Heatmap Generation

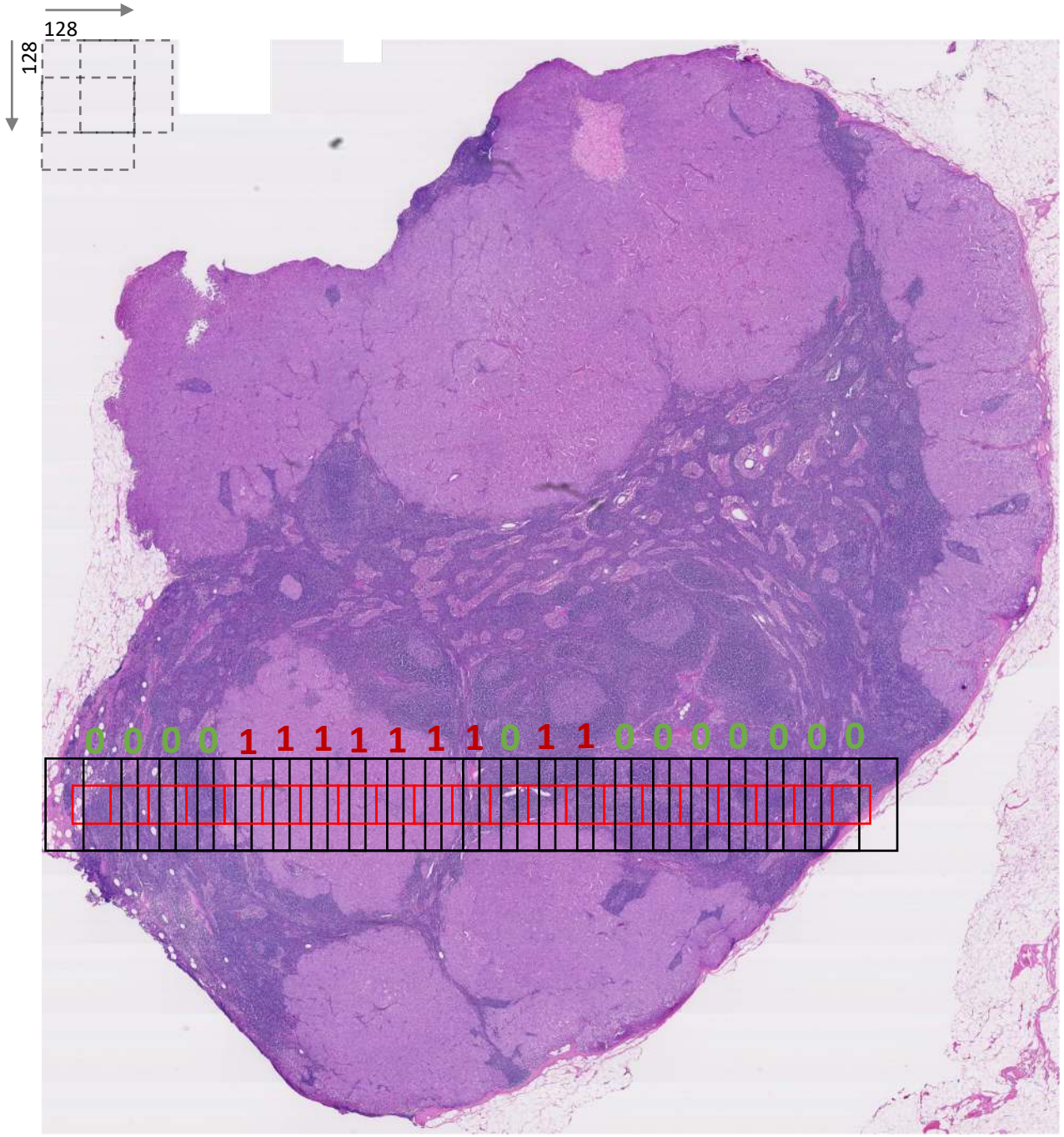
- Generate a heatmap at level 7 (maximally zoomed-out) – initialize with zeros
- Proceed through the slide by stride of 128 pixels in x and y, getting 299x299 patches at the target levels
- Skip patches with less than 50% tissue
- To speed-up, gather 1000 patches and do the prediction on that batch
- Get the true label for each patch by checking if there is any tumor pixel in the 128x128 center of the patch



slide 110, size of patches at level 4

Heatmap Generation

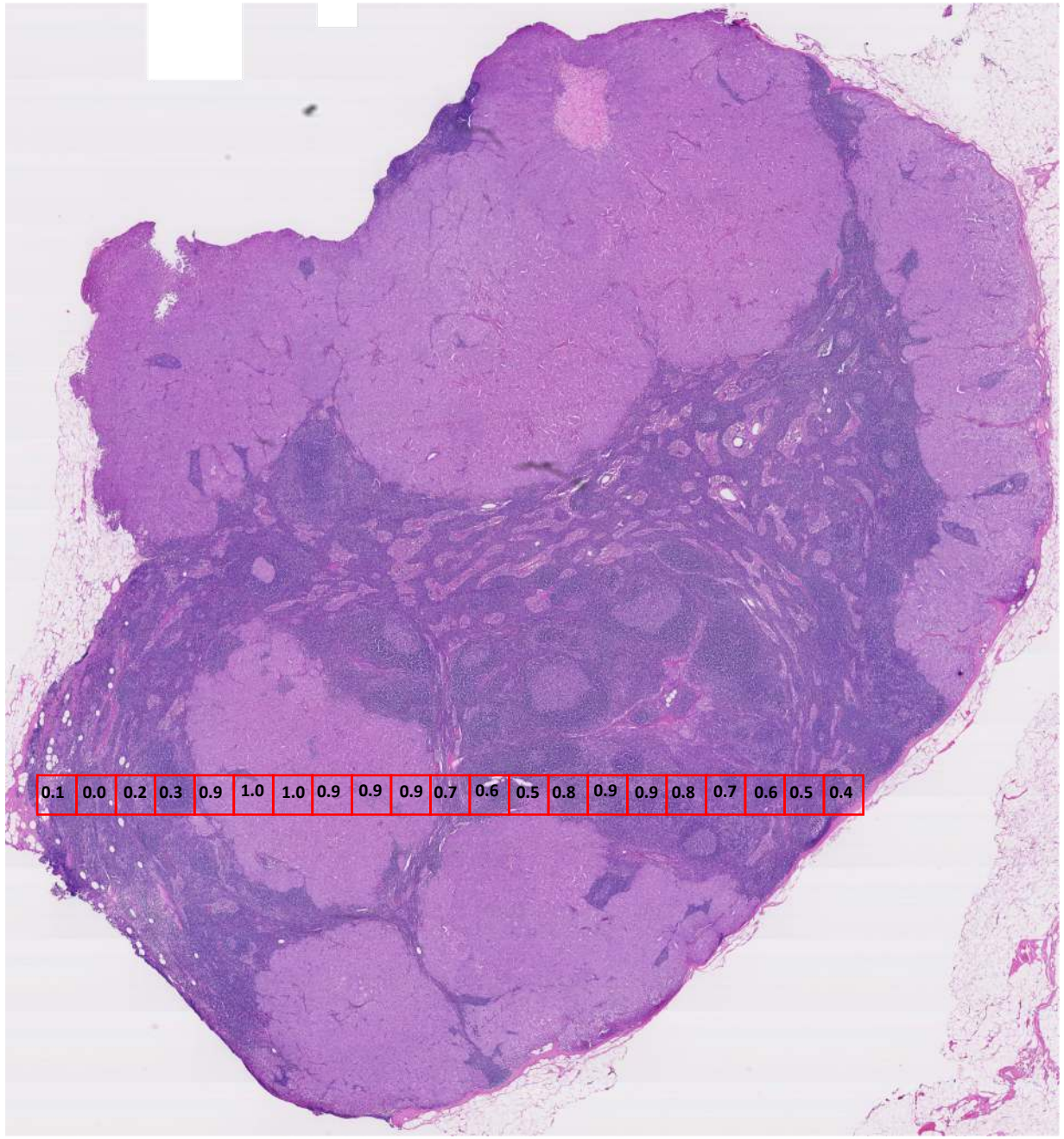
- Generate a heatmap at level 7 (maximally zoomed-out) – initialize with zeros
- Proceed through the slide by stride of 128 pixels in x and y, getting 299x299 patches at the target levels
- Skip patches with less than 50% tissue
- To speed-up, gather 1000 patches and do the prediction on that batch
- Get the true label for each patch by checking if there is any tumor pixel in the 128x128 center of the patch
- Get model prediction (sigmoid output) and predicted class (binary) for each patch



slide 110, size of patches at level 4

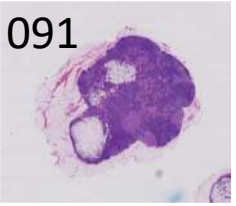
Heatmap Generation

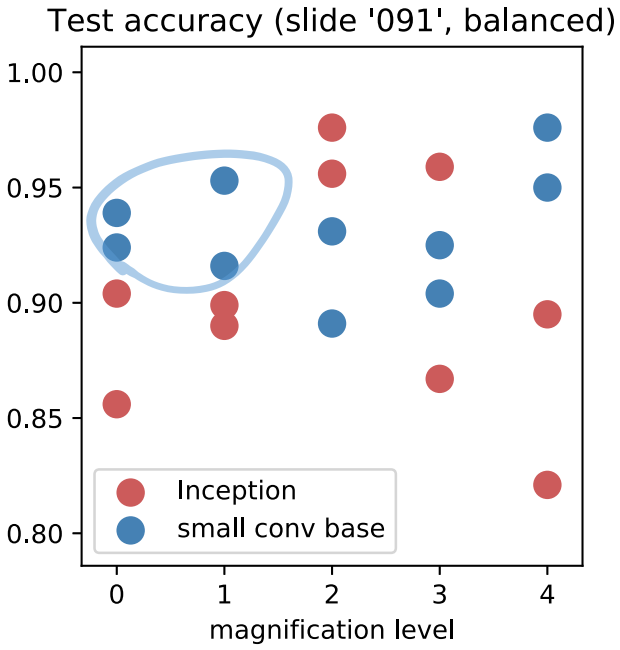
- Generate a heatmap at level 7 (maximally zoomed-out) – initialize with zeros
- Proceed through the slide by stride of 128 pixels in x and y, getting 299x299 patches at the target levels
- Skip patches with less than 50% tissue
- To speed-up, gather 1000 patches and do the prediction on that batch
- Get the true label for each patch by checking if there is any tumor pixel in the 128x128 center of the patch
- Get model prediction (sigmoid output) and predicted class (binary) for each patch
- Put values of prediction into the heatmap for pixels corresponding to the 128x128 center of the patch
- Calculate imbalanced classification metrics using the true and predicted classes

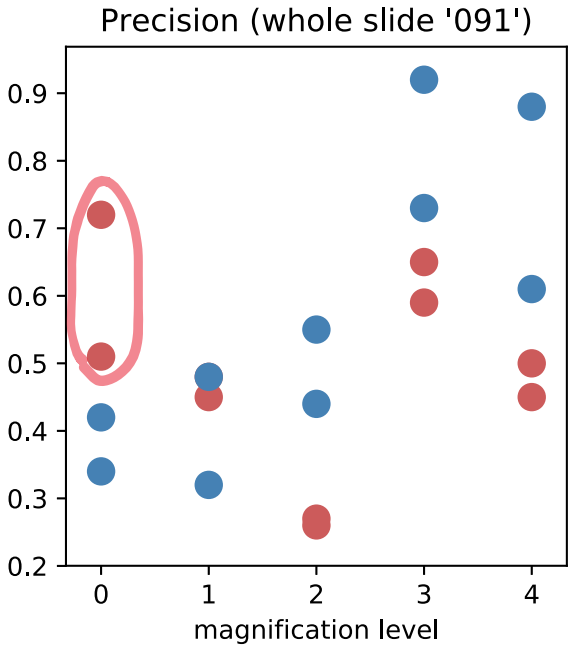


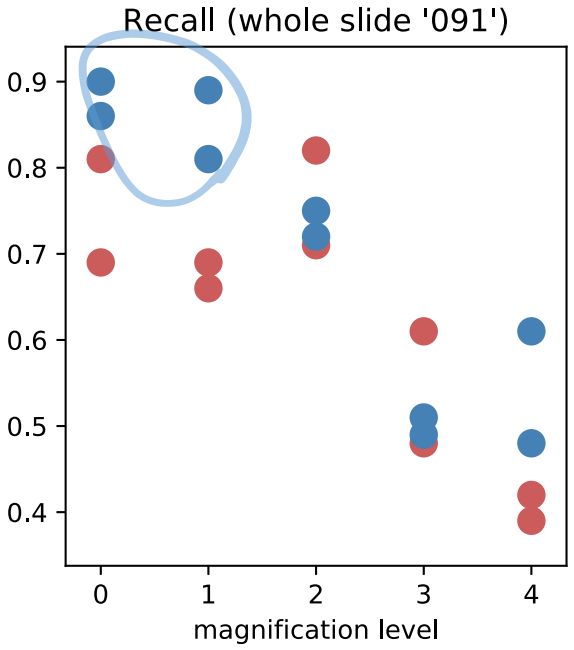
slide 110, size of patches at level 4

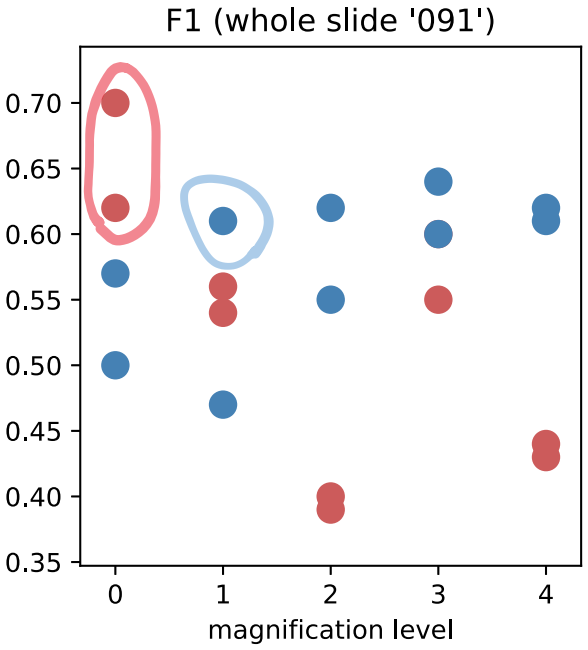
Single Scale Experiments











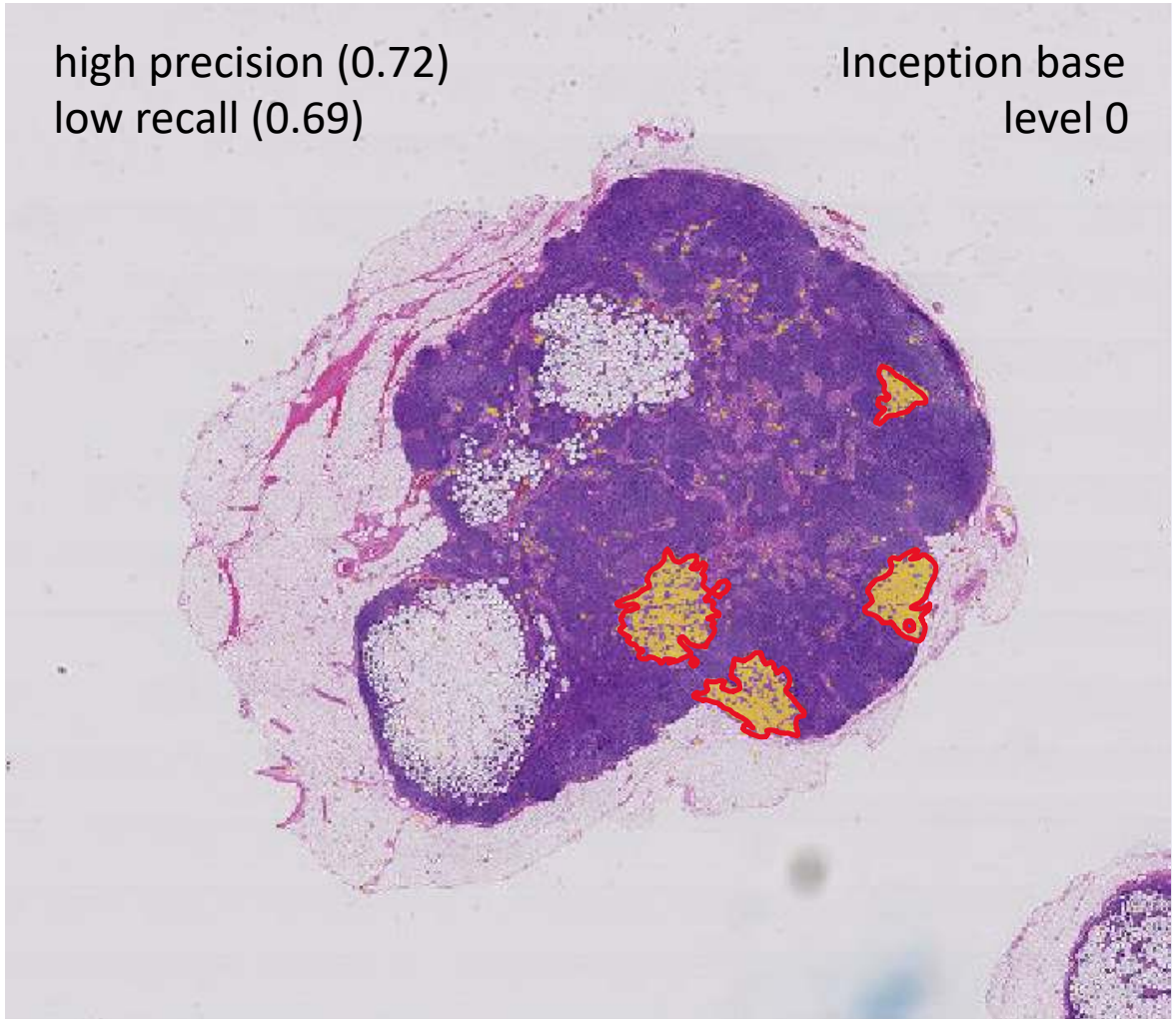
$$precision = \frac{tp}{tp + fp}$$

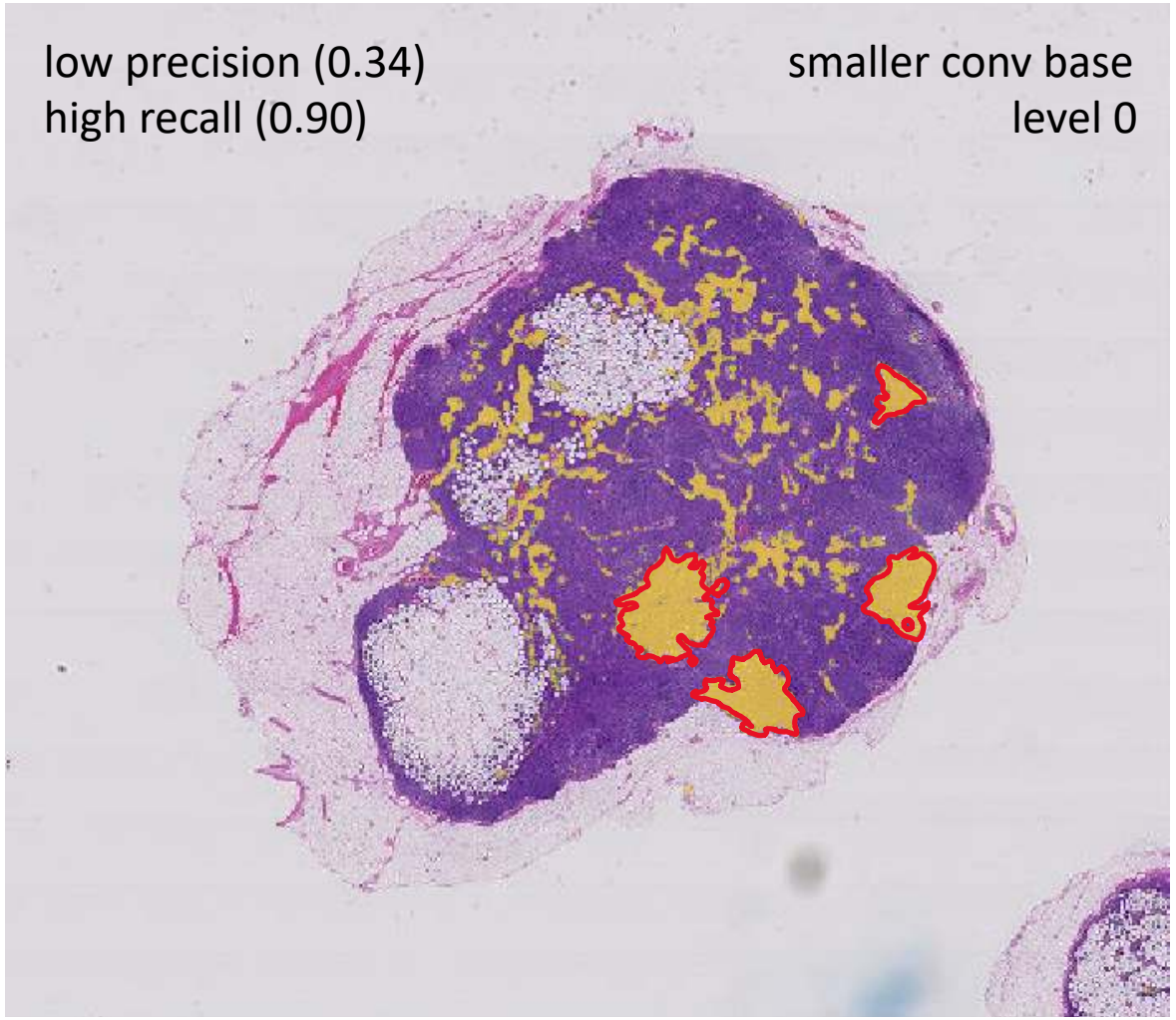
$$recall = \frac{tp}{tp + fn}$$

$$f1 = \frac{2tp}{2tp + fp + fn}$$

		predicted	
		0	1
true	0	*tn*	*fp*
	1	*fn*	*tp*

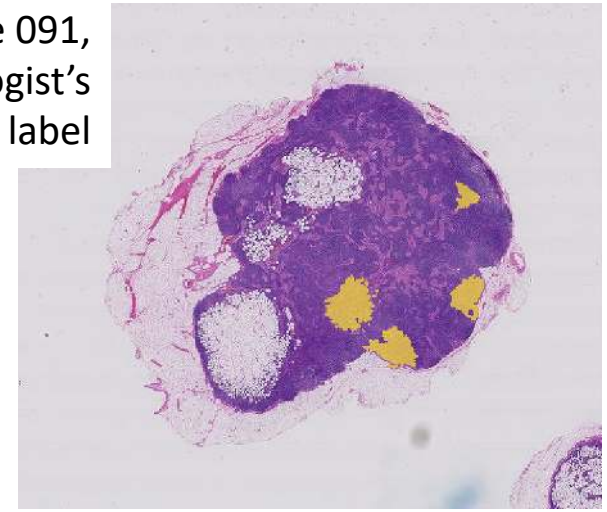
Precision-Recall Tradeoff



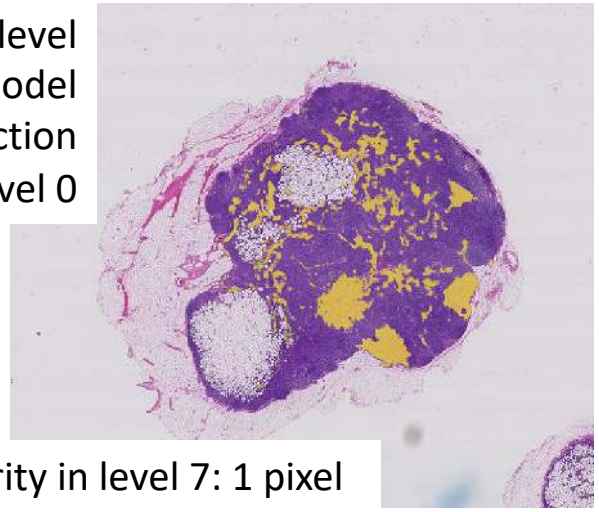


Single Scale Model with Small Conv Base

Slide 091, pathologist's label

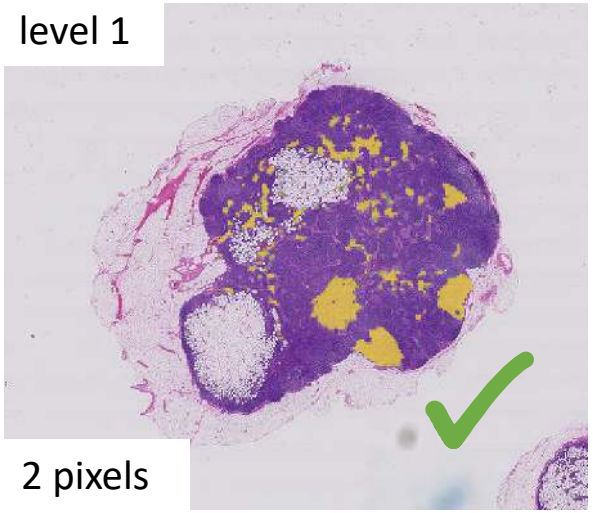


single level model prediction - level 0



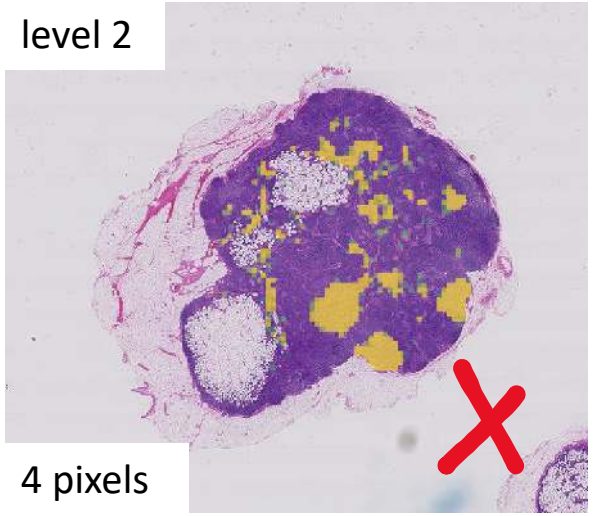
granularity in level 7: 1 pixel

level 1



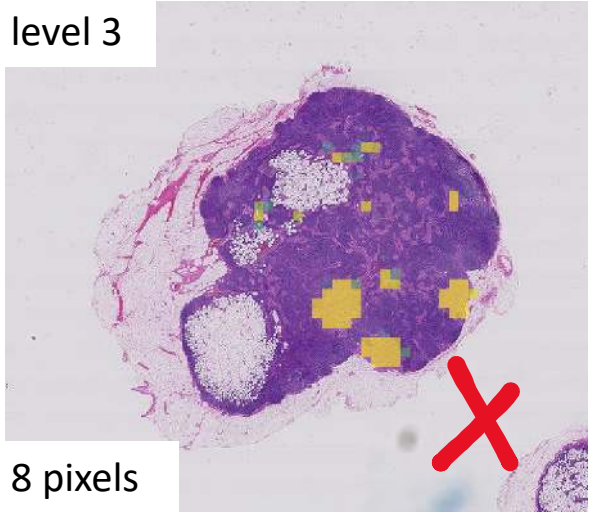
2 pixels

level 2



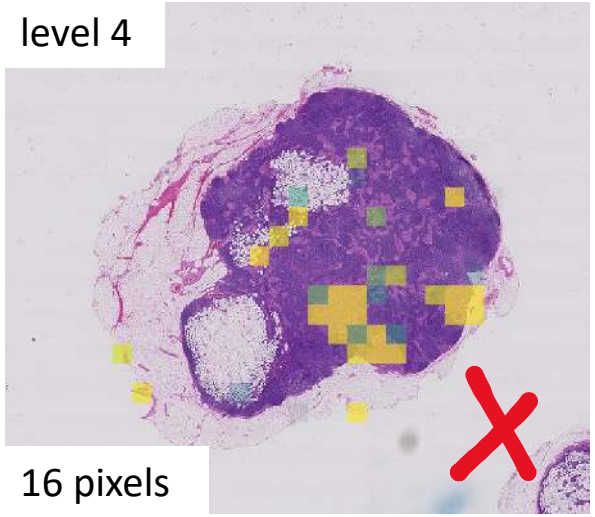
4 pixels

level 3



8 pixels

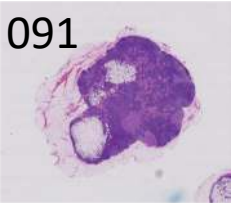
level 4



16 pixels

Slide '091'
level 7
480x420
pixels

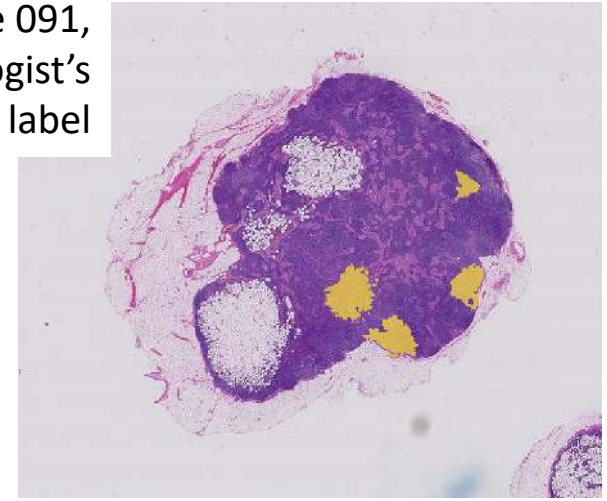
Multi Scale Experiments



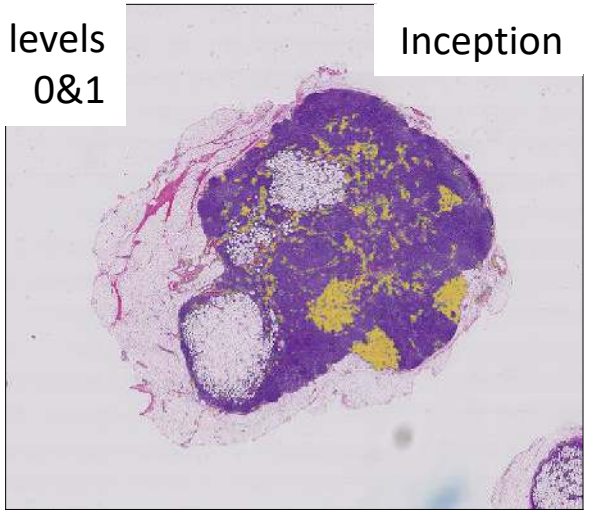
Model Type	Conv Base	Zoom Level	Test Accuracy (balanced)	Precision (entire 091)	Recall (entire 091)	F1 (entire 091)
Multi level	Inception	0 & 1	0.846	0.3	0.87	0.44
		1 & 2	0.864	0.26	0.81	0.4
	small (4xConv2D)	0 & 1	0.924	0.5	0.79	0.62
		0 & 2	0.917	0.3	0.91	0.46
		1 & 2	0.883	0.31	0.86	0.45
		1 & 3	0.869	0.3	0.88	0.44
		2 & 4	0.921	0.38	0.78	0.51

Multi Scale Model Heatmaps

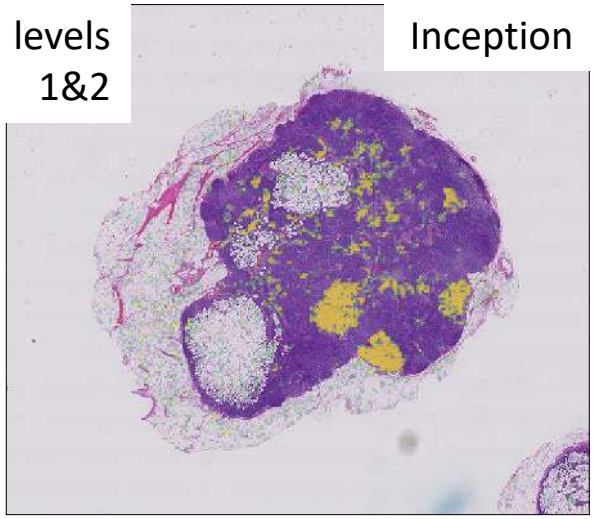
Slide 091, pathologist's label



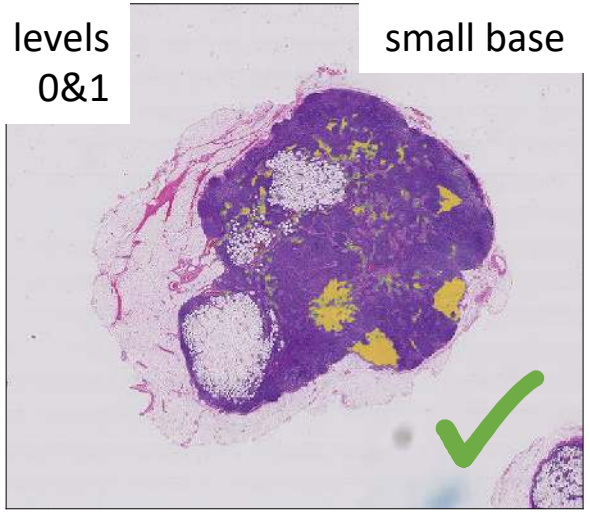
levels 0&1 Inception



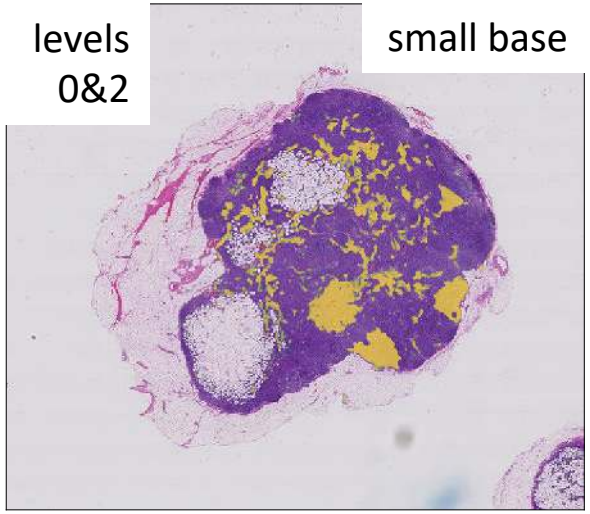
levels 1&2 Inception



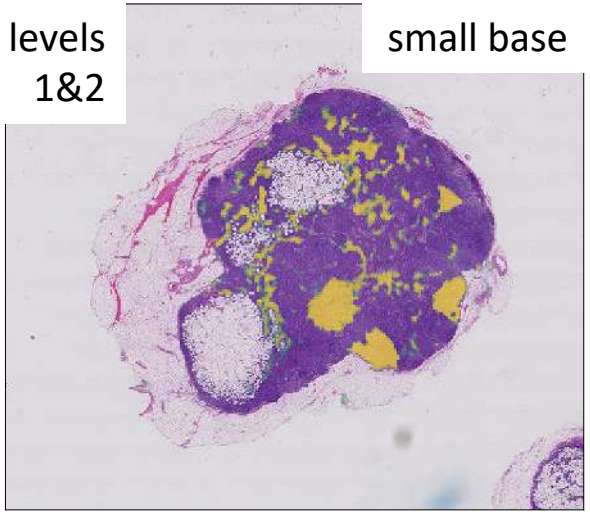
levels 0&1 small base



levels 0&2 small base

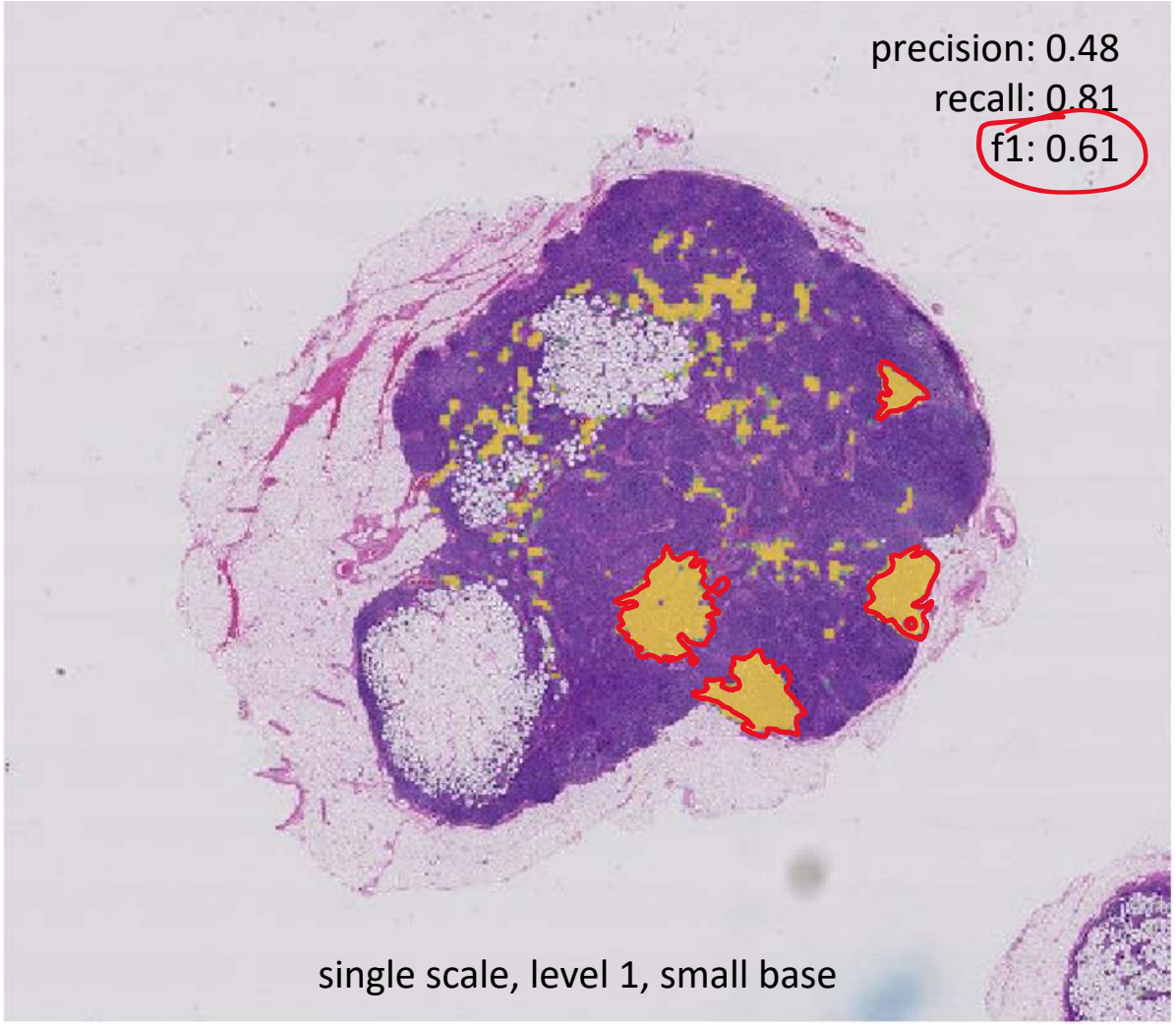


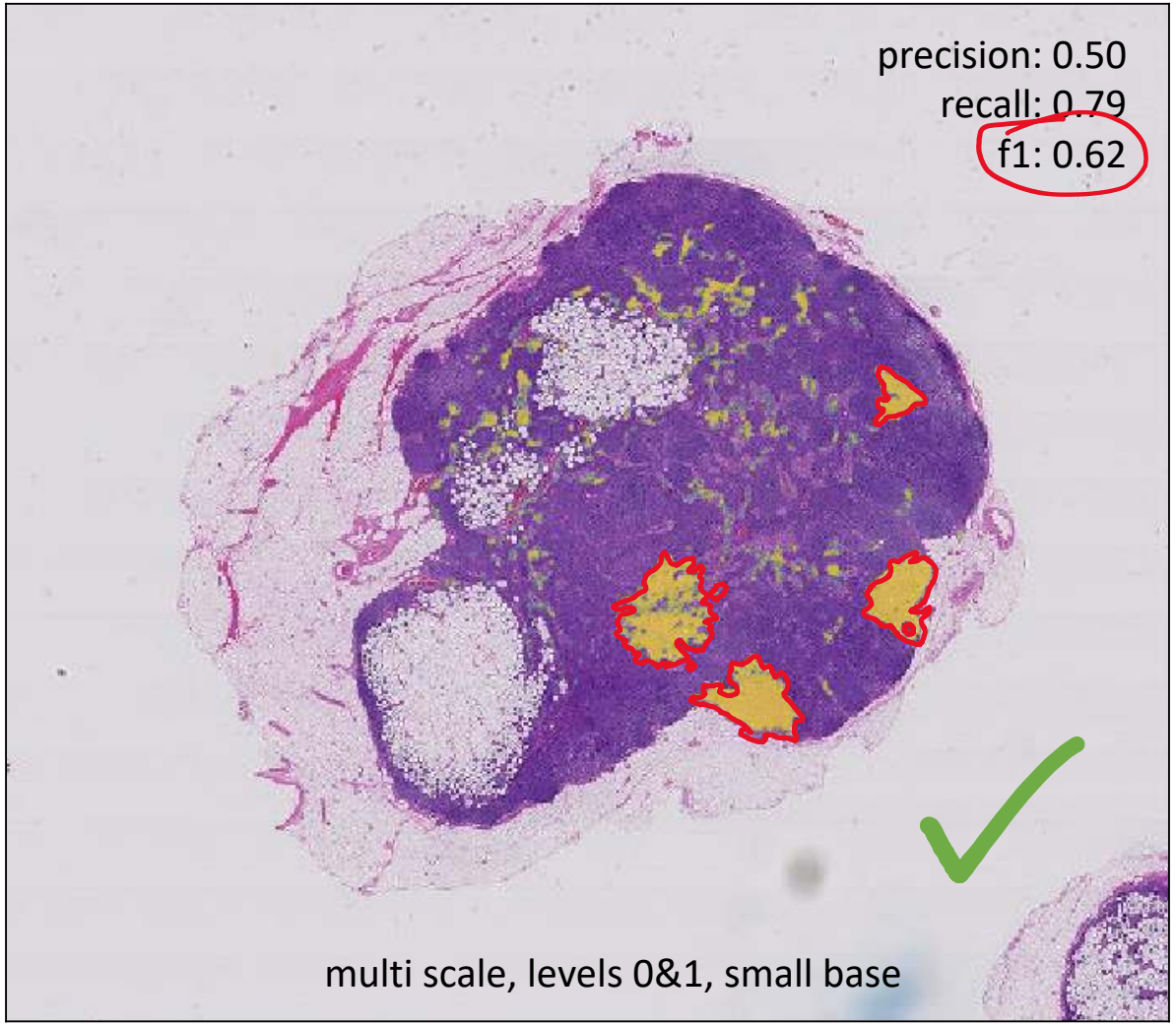
levels 1&2 small base



Slide '091'
level 7
480x420
pixels

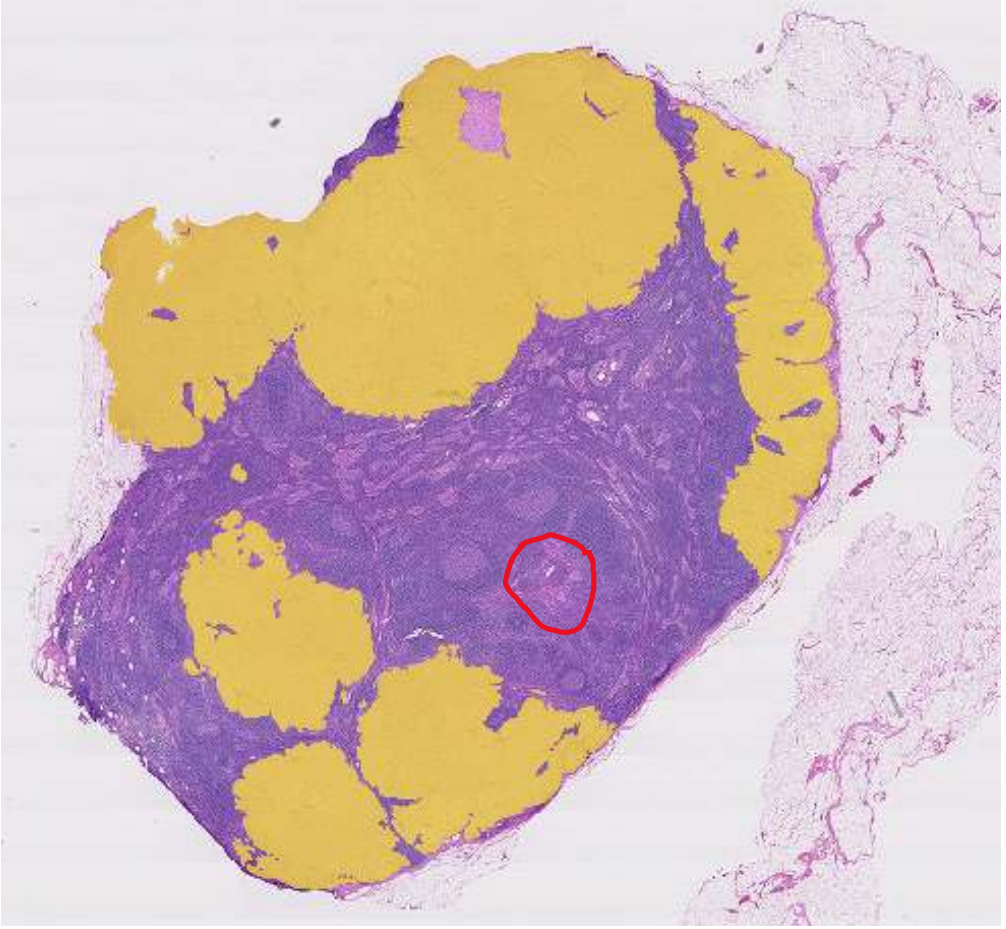
Model Comparison – Slide ‘091’



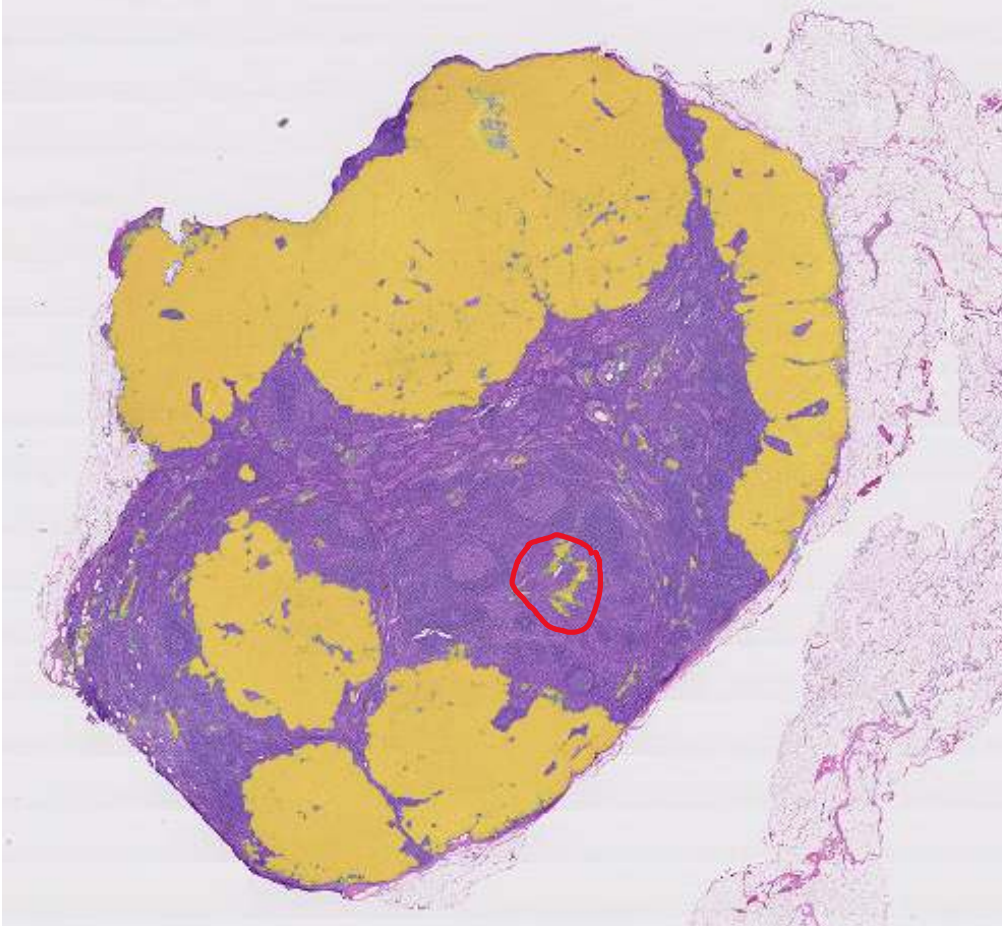


Best Model – Slide ‘110’ Prediction

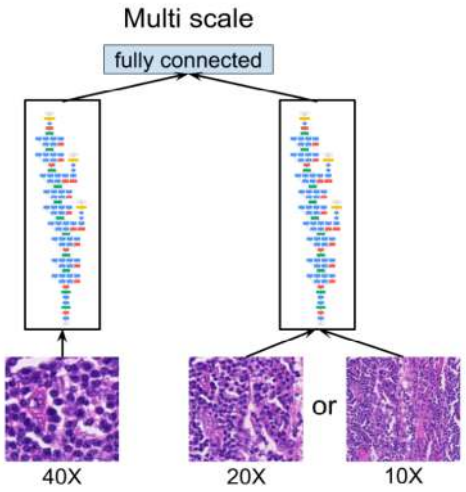
true annotation (non-exhaustive)



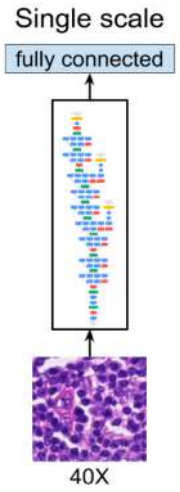
multi-level, small base, lev 0&1



Conclusions



multi scale > single scale



MaxPool
BatchNorm
Conv2D
MaxPool
BatchNorm
Conv2D
MaxPool
BatchNorm
Conv2D
MaxPool
BatchNorm
Conv2D

small convolutional base > Inception

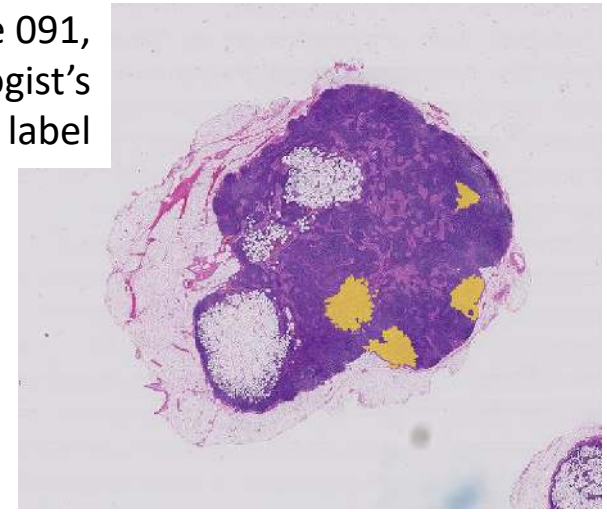
Appendix

Model Comparison

	Single Level		Multiple Level	
device	Colab: GPU, 12GB		GC compute instance: CPU, 60GB	
set-up	runs end-to-end in Colab (packages installed with !apt-get and !pip)		packages and libraries installed externally (careful to install openslide 3.4.1!)	
slides source	copied from mounted Google Drive		copied to GC instance using gcloud utilities	
# examples	100 per image per class, train: 1140, validation: 360, test: 200			
loss	binary cross-entropy			
Convolutional base	Inception	smaller	Inception	smallest
steps per epoch	288	288	96	288
training time per epoch	~6 min	~3 min	~40 min	~40 min
optimizer	RMSprop(lr=2e-5)		RMSprop(lr=1e-6)	

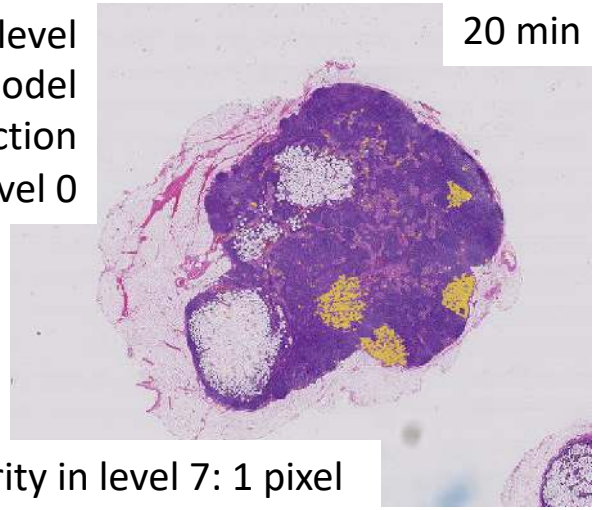
Single Level Model with Inception

Slide 091, pathologist's label



single level model prediction - level 0

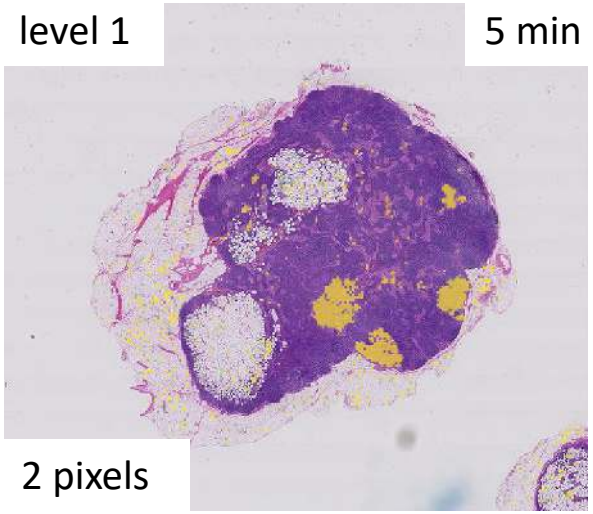
20 min



granularity in level 7: 1 pixel

level 1

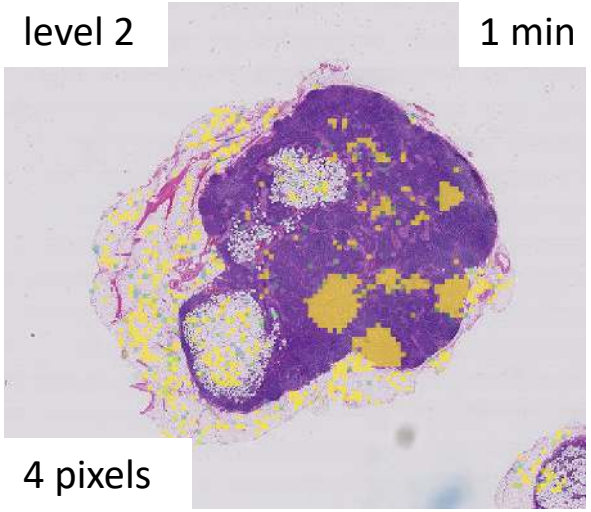
5 min



2 pixels

level 2

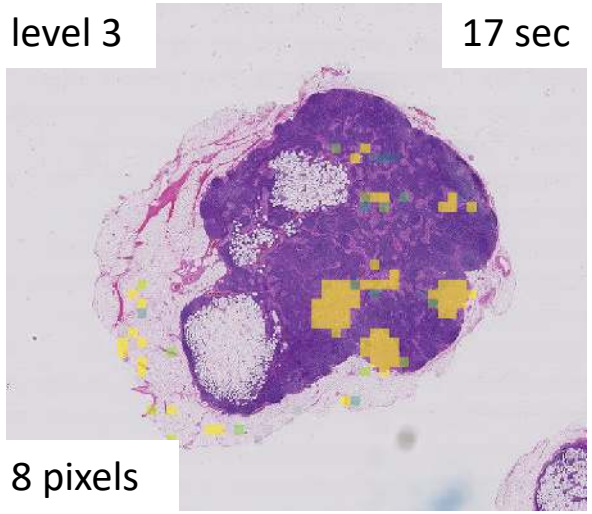
1 min



4 pixels

level 3

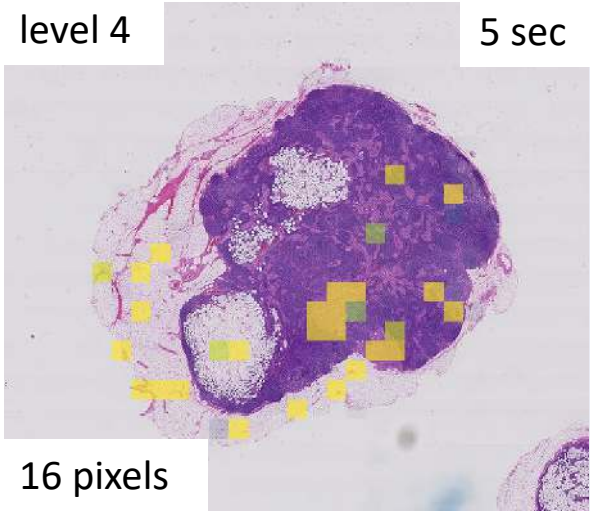
17 sec



8 pixels

level 4

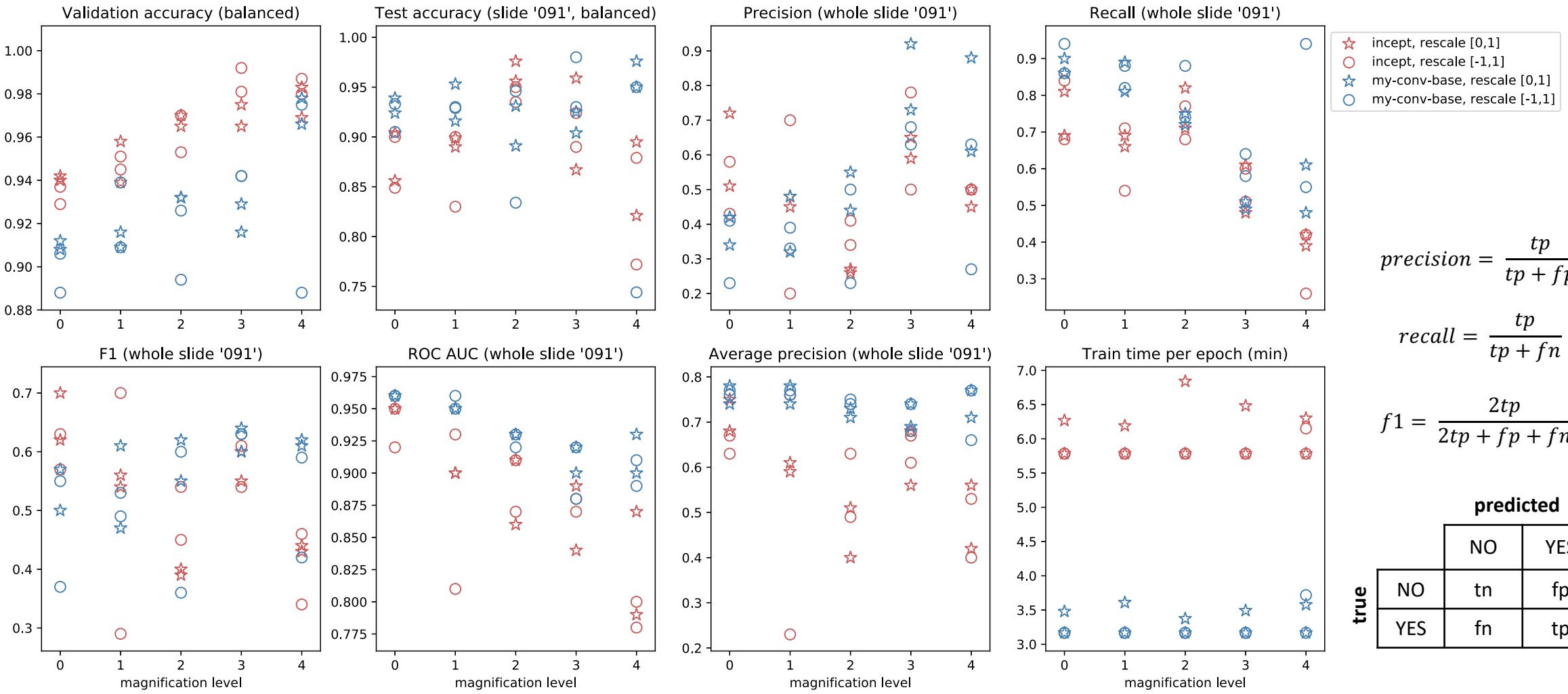
5 sec



16 pixels

Slide '091'
level 7
480x420
pixels

Single Level Model Experiments



$$precision = \frac{tp}{tp + fp}$$

$$recall = \frac{tp}{tp + fn}$$

$$f1 = \frac{2tp}{2tp + fp + fn}$$

		predicted	
		NO	YES
true	NO	tn	fp
	YES	fn	tp

Multi Level Model Experiments

Model Type	Conv Base	Zoom Level	Test Accuracy (balanced)	Precision (entire 091)	Recall (entire 091)	F1 (entire 091)	ROC AUC (entire 091)	Average Precision (entire 091)
Multi level	Inception	0 & 1	0.846	0.3	0.87	0.44	0.94	0.58
		1 & 2	0.864	0.26	0.81	0.4	0.91	0.63
	small (4xConv2D)	0 & 1	0.924	0.5	0.79	0.62	0.96	0.76
		0 & 2	0.917	0.3	0.91	0.46	0.96	0.75
		1 & 2	0.883	0.31	0.86	0.45	0.94	0.76
		1 & 3	0.869	0.3	0.88	0.44	0.95	0.78
		2 & 4	0.921	0.38	0.78	0.51	0.91	0.71
Single level	Inception	0	0.856	0.72	0.69	0.7	0.95	0.75
	small	0	0.924	0.34	0.9	0.5	0.96	0.74