

Applied Deep Learning

Lecture 5 • Feb 21st, 2019

News / Administrative stuff

News

- [Open AI GPT-2](#)

Administrative stuff

- Pushed HW3 back a bit.
- **Cloud credits** (just a small amount, please fill out [this survey](#) if you'd like some).

It'll be helpful to host files on cloud storage for your project.

Questions from Office Hours

**What do we mean
by small models?**

	Top-1	Top-5	10-5	Size	Stem	References
VGG16	28.732	9.950	8.834	138.4M	14.7M	[paper] [tf-models]
VGG19	28.744	10.012	8.774	143.7M	20.0M	[paper] [tf-models]
ResNet50	25.072	7.940	6.828	25.6M	23.6M	[paper] [tf-models] [torch] [caffe]
ResNet101	23.580	7.214	6.092	44.7M	42.7M	[paper] [tf-models] [torch] [caffe]
ResNet152	23.396	6.882	5.908	60.4M	58.4M	[paper] [tf-models] [torch] [caffe]
ResNet50V2	24.040	6.966	5.896	25.6M	23.6M	[paper] [tf-models] [torch]
ResNet101V2	22.766	6.184	5.158	44.7M	42.6M	[paper] [tf-models] [torch]
ResNet152V2	21.968	5.838	4.900	60.4M	58.3M	[paper] [tf-models] [torch]
ResNeXt50	22.260	6.190	5.410	25.1M	23.0M	[paper] [torch]
ResNeXt101	21.270	5.706	4.842	44.3M	42.3M	[paper] [torch]
InceptionV3	22.102	6.280	5.038	23.9M	21.8M	[paper] [tf-models]
InceptionResNetV2	19.744	4.748	3.962	55.9M	54.3M	[paper] [tf-models]
Xception	20.994	5.548	4.738	22.9M	20.9M	[paper]
MobileNet(alpha=0.25)	48.418	24.208	21.196	0.5M	0.2M	[paper] [tf-models]

<https://github.com/keras-team/keras-applications>

Storing files on the Cloud

To create a public file:

- Add a user named “allUsers” with Reader access.

butler.jpg permissions

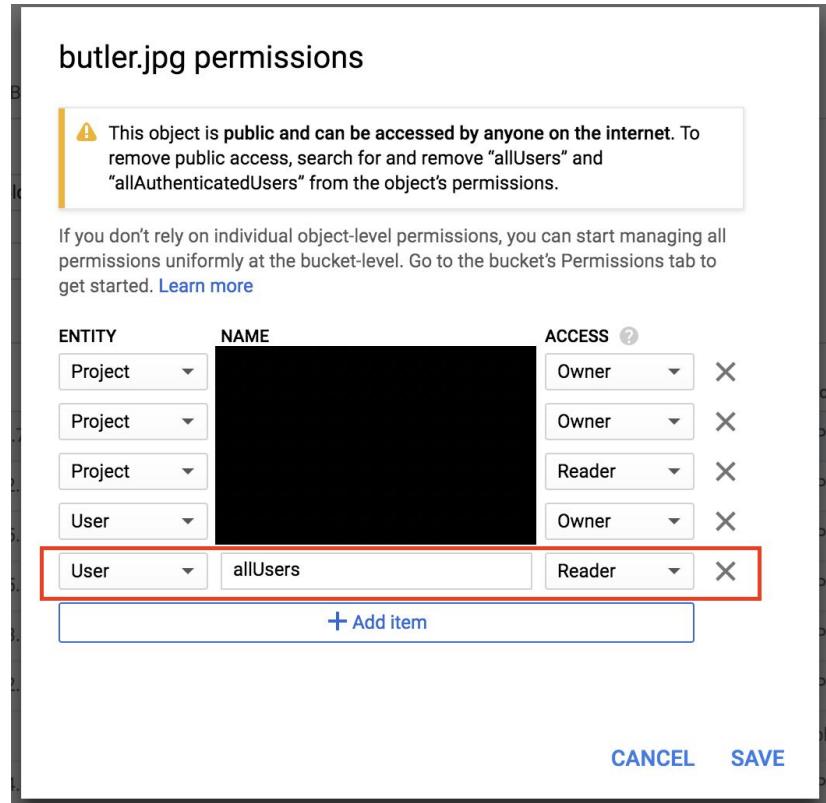
⚠ This object is public and can be accessed by anyone on the internet. To remove public access, search for and remove “allUsers” and “allAuthenticatedUsers” from the object’s permissions.

If you don't rely on individual object-level permissions, you can start managing all permissions uniformly at the bucket-level. Go to the bucket's Permissions tab to get started. [Learn more](#)

ENTITY	NAME	ACCESS	X
Project		Owner	X
Project		Owner	X
Project		Reader	X
User		Owner	X
User	allUsers	Reader	X

[+ Add item](#)

CANCEL SAVE



<https://cloud.google.com/storage/docs/creating-buckets>

Is this text real or fake?

In a shocking finding, scientist discovered a herd of unicorns living in a remote, previously unexplored valley, in the Andes Mountains. Even more surprising to the researchers was the fact that the unicorns spoke perfect English.

The scientist named the population, after their distinctive horn, Ovid's Unicorn. These four-horned, silver-white unicorns were previously unknown to science.

Now, after almost two centuries, the mystery of what sparked this odd phenomenon is finally solved.

Dr. Jorge Pérez, an evolutionary biologist from the University of La Paz, and several companions, were exploring the Andes Mountains when ...

Try it (5 mins)

Warning: this model is not filtered at all, may generate offensive text

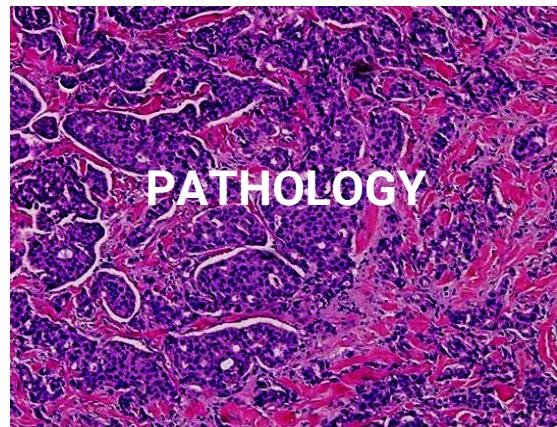
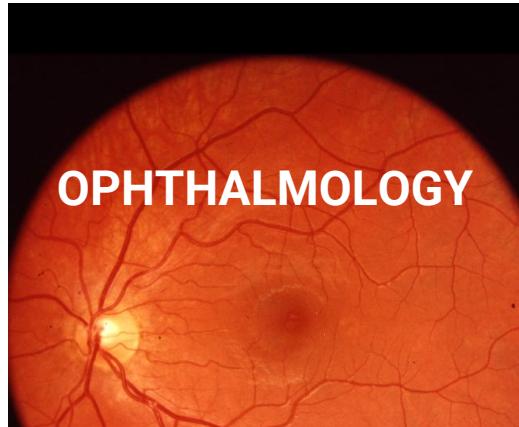
- bit.ly/easy-gpt

Quick discussion

Would you have shared the full model? Why or why not?

Today's agenda

- Medical imaging
- Project descriptions



<https://research.googleblog.com/2016/11/deep-learning-for-detection-of-diabetic.html>

<https://research.googleblog.com/2017/03/assisting-pathologists-in-detecting.html>

From bench -> bedside

- **Goal:** radically reduce the misdiagnosis rate for common diagnostic tasks (using mostly off the shelf software and well known algorithms).
- Challenges: **mostly not technical.**

Making this a reality requires a combination of technical skills, interpersonal skills, and domain knowledge - I hope you will be able to make a difference during your careers.

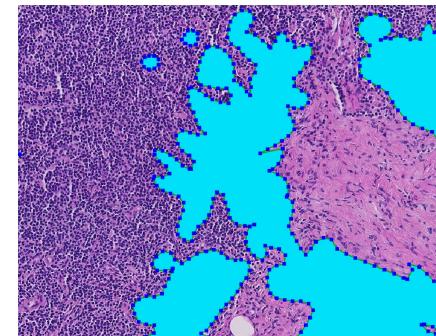
Projects

Choose either the class project or propose your own

Class project

- Detect cancer in gigapixel pathology images.

*Use techniques similar to the paper described in class today.
Starter code provided. This is feasible without a cluster,
using the single GPU on Colab.*



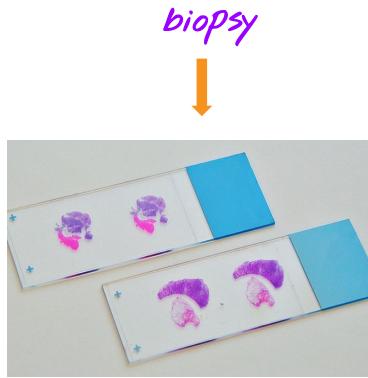
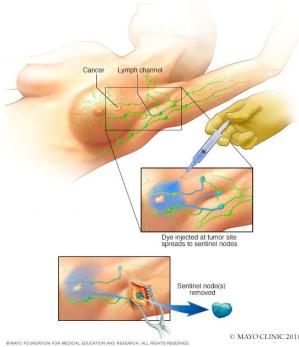
Custom project

- If you'd like to do a custom project, please submit a proposal (this can be short, just a paragraph and/or bullet points outlining your idea on CourseWorks by **3/7** -- this isn't graded, it's just so we can offer feedback / suggestions, see if it's feasible - though you may know more about your area of interest than I do!).

Deliverables

*Can be unlisted.
,*

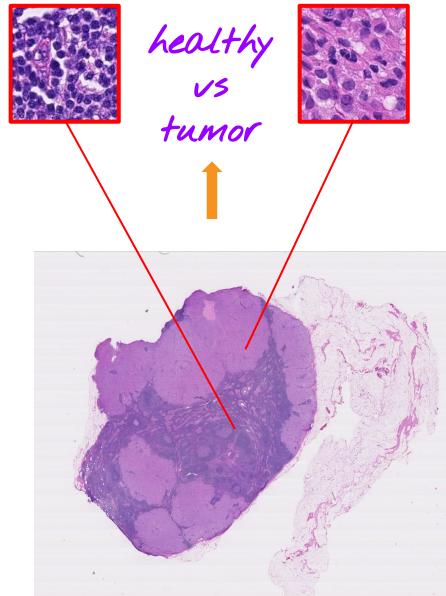
- **Video:** a 10 minute YouTube video explaining your project / the key pieces of code / experiments you ran / and results. Should contain slides + code walkthrough.
- **Complete code:** either a Jupyter notebook that runs end-to-end on Colab, or a zip including source, a README, and a shell script to run it. Please include a saved model as well.
- **Demo:** a demo on the last day of class, sign up TBD.
- **Writeup:** optional. If you have experiments or results that don't fit into one of the above formats, feel free to submit a PDF or HTML report along with your project.



Preparation



Diagnosis → Treatment plan



Visual inspection

<https://camelyon17.grand-challenge.org/Data/>

<https://www.mayoclinic.org/tests-procedures/sentinel-node-biopsy/about/pac-20385264>

If you have a medical background, let me know if there's a better way to describe any of these concepts!

Visual inspection

Quick discussion: what can we help with? What can go wrong?

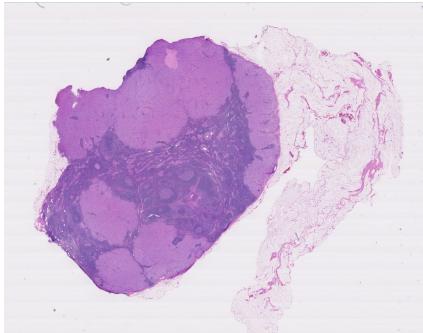
Visual inspection

- Can be tedious and error prone (despite best efforts of pathologists).
- Tissue samples are volumes: when sliced, many images / sample.
- Each image is high resolution (order of 100,000 x 100,000 pixels).
- Expertise is rare. Pathologists have many patients, and limited time to spend on each slide.
- Tumors can be small. Needle in a haystack.
- Fatigue. Mistakes happen.

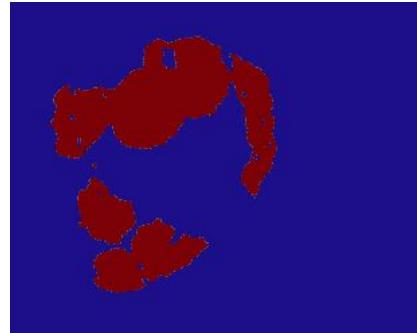
Project: Develop a tool to assist physicians

- Given a collection of training data, develop a model that outputs a heatmap showing regions of a biopsy image likely to contain cancer.
- Emphasis on **assist**. Not replace.

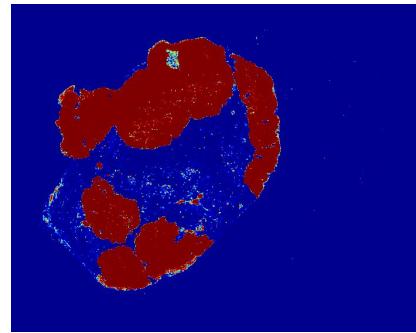
Of course, making this useful in a real-world clinical workflow is far more difficult than developing a model - this is just the first step.



Biopsy image



*Ground truth
(from pathologist)*



Model predictions

Assist or automate?

Quick discussion

Assist, not automate

- Best bet: insert into workflow as an **automatic second opinion** (if reads differ, flag and ask for second pathologist).
- Inserting into workflow before a pathologist reads the slide could lead to over reliance on the tool. Models / datasets aren't up for that yet in an open-world setting.
- You can imagine with enough training data, we can develop models with super-physician level accuracy, and **radically reduce misdiagnosis at little cost**.

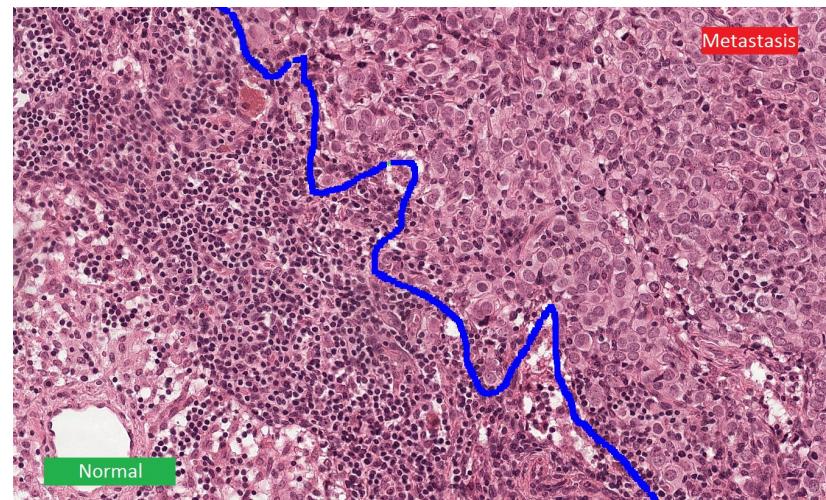
CAMELYON16 challenge

<https://camelyon16.grand-challenge.org/Data/>

400 WSI (whole slide images) collected independently from two medical centers in the Netherlands.

- Slide level annotations.
- Importantly, licensed under [CC0](#).
- About 600GB.

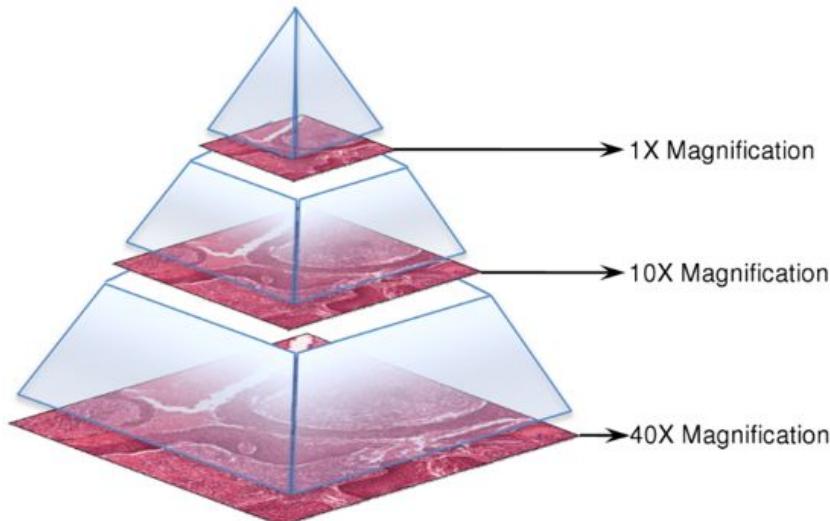
Important: see the README before using this data for notes on annotation quality in several slides.



Demo

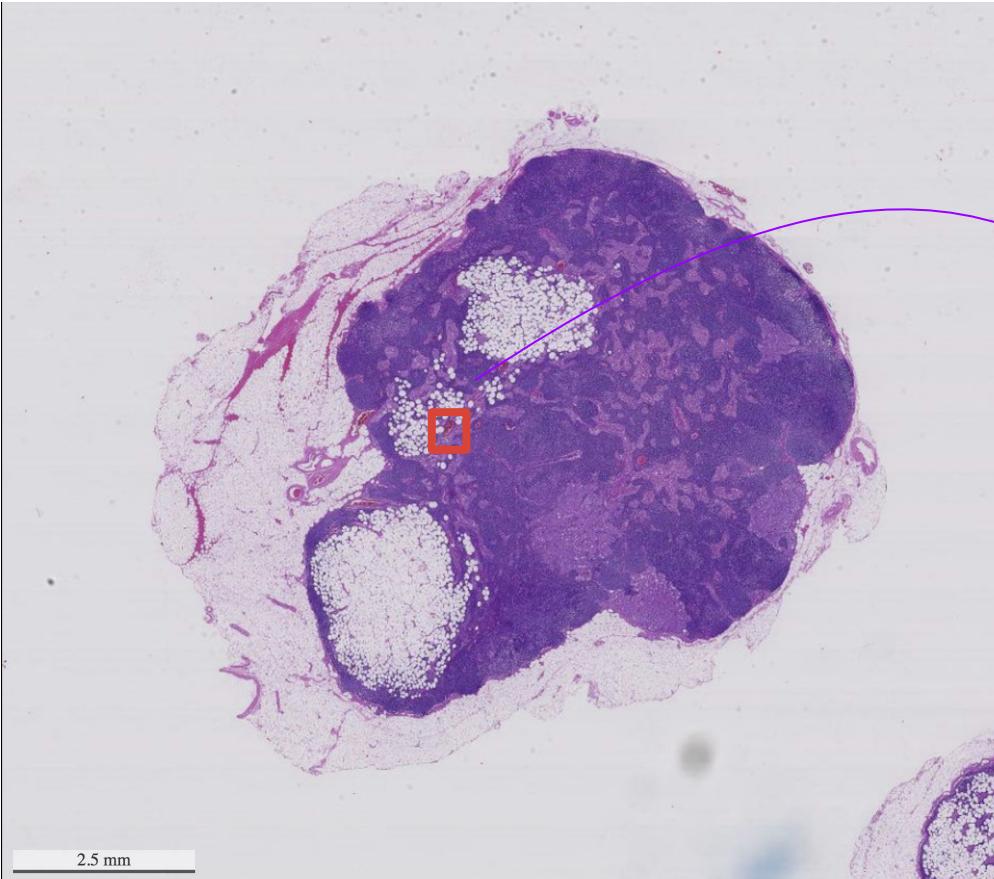
```
$ python deepzoom_server.py tumor_091.tif
```

Included with the open-slide python package (install the C library first).



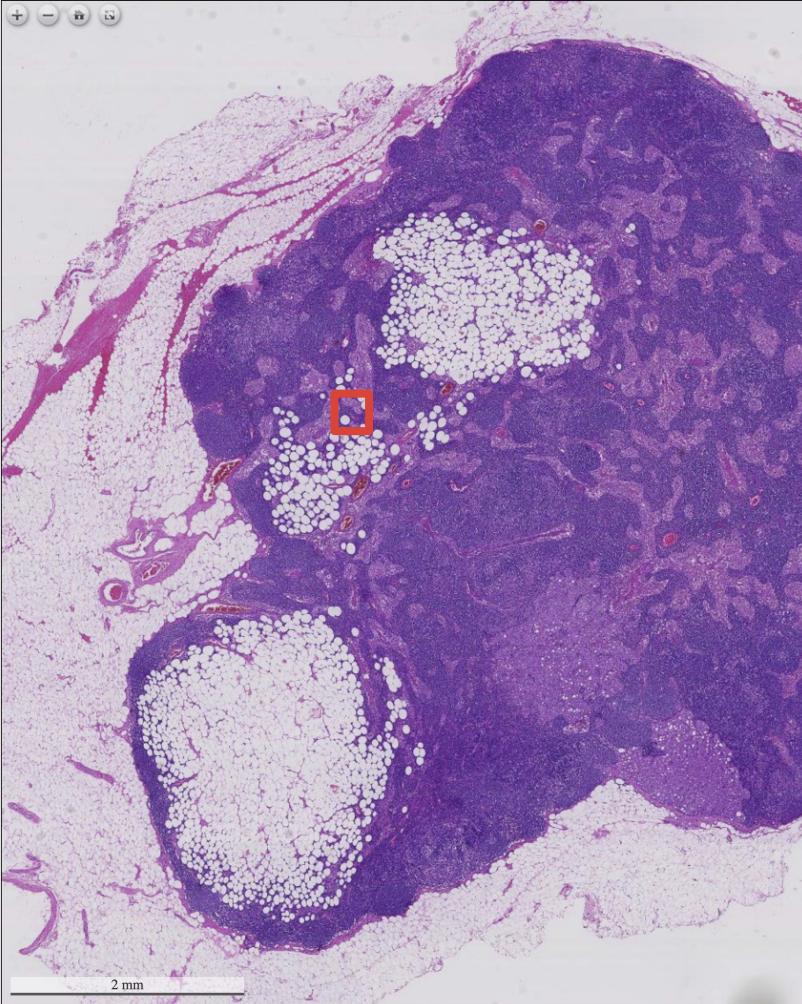
<https://openslide.org/>

This is the only non-Python library needed to work with this data, if you use the subset we provided where annotations have been pre-converted to masks. This may be easier than installing the ASAP tool provided by the competition organizers (tricky on mac, and possibly linux).

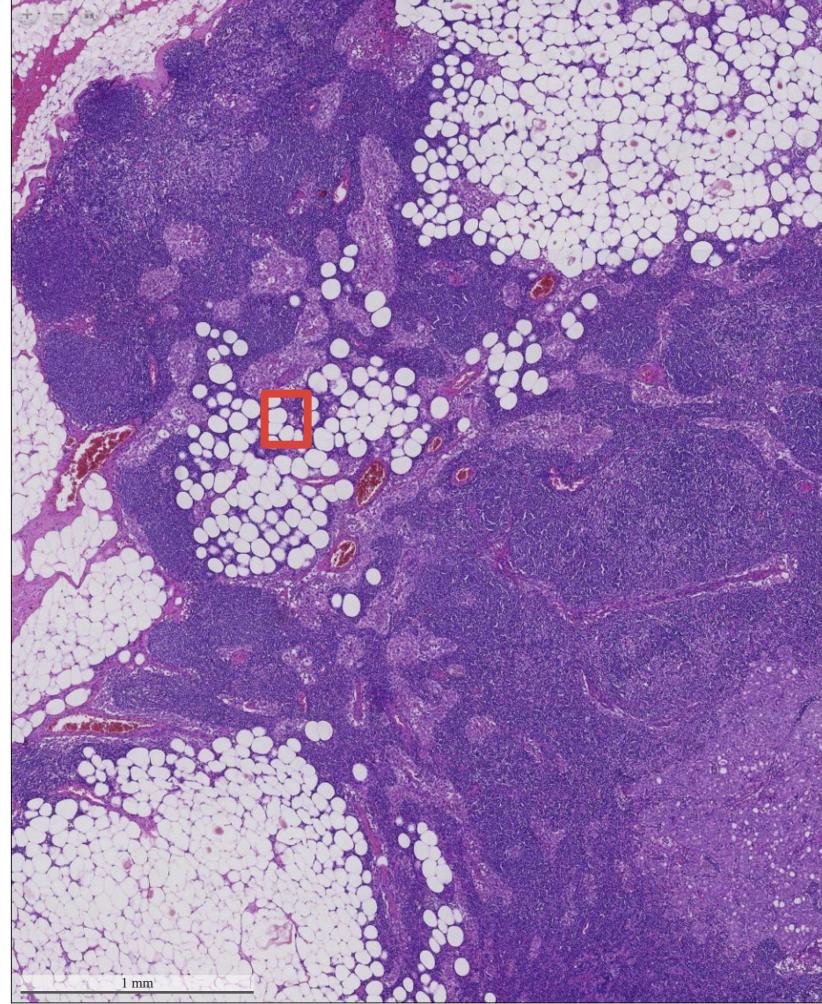


Notice the scale in the lower left.

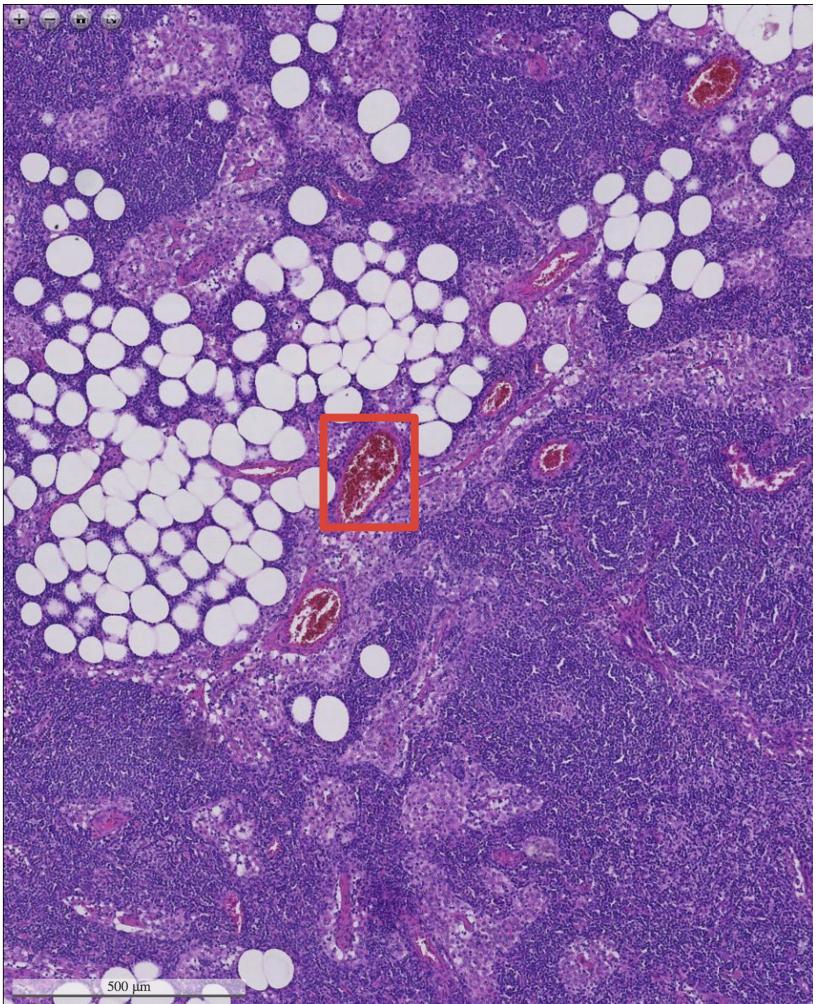
tumor091.tif from the 2016 training set



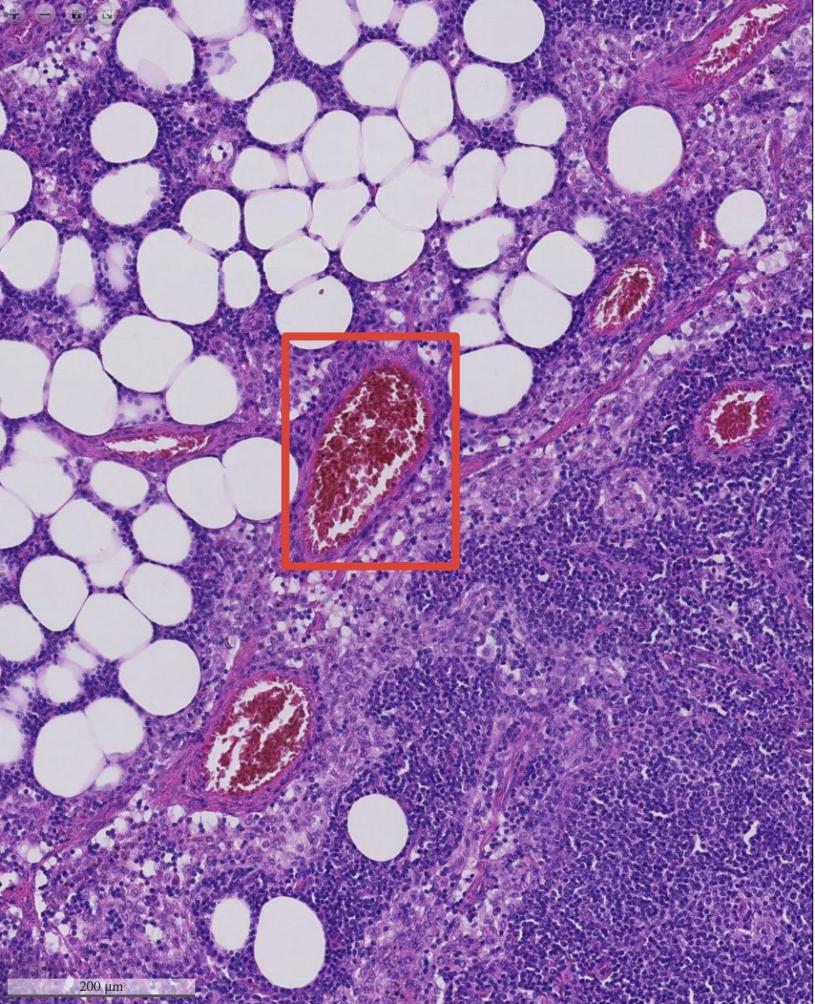
COMS 4995.07. Spring 19.



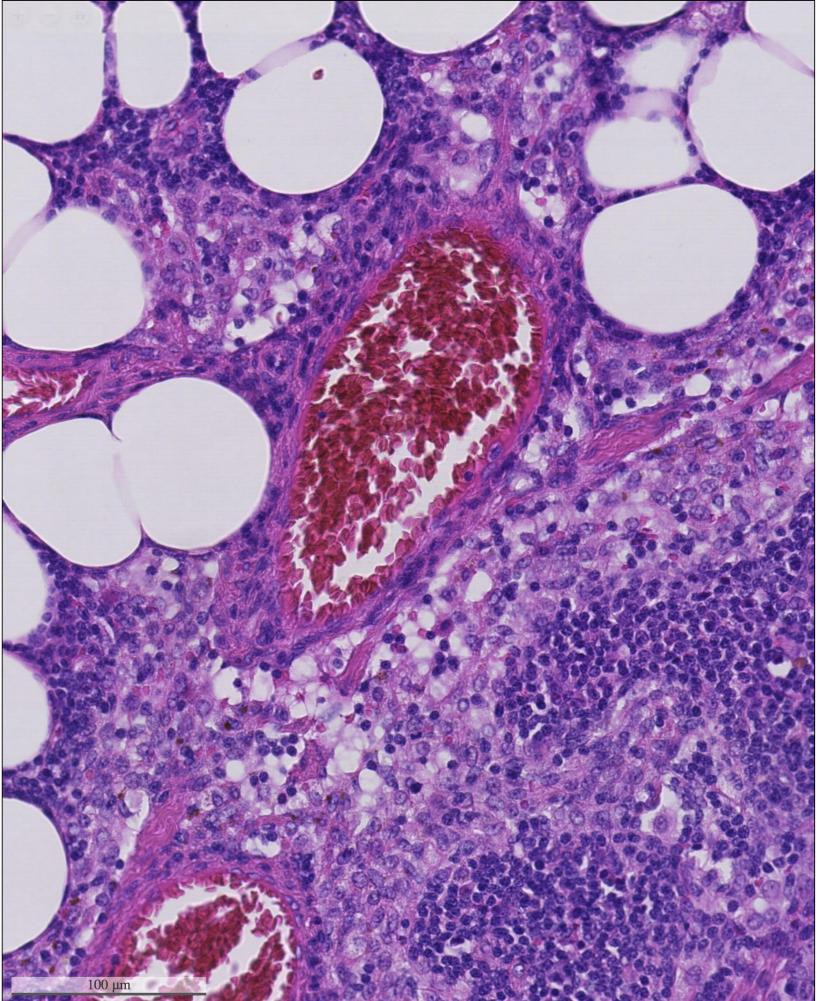
Lecture 5 - 21



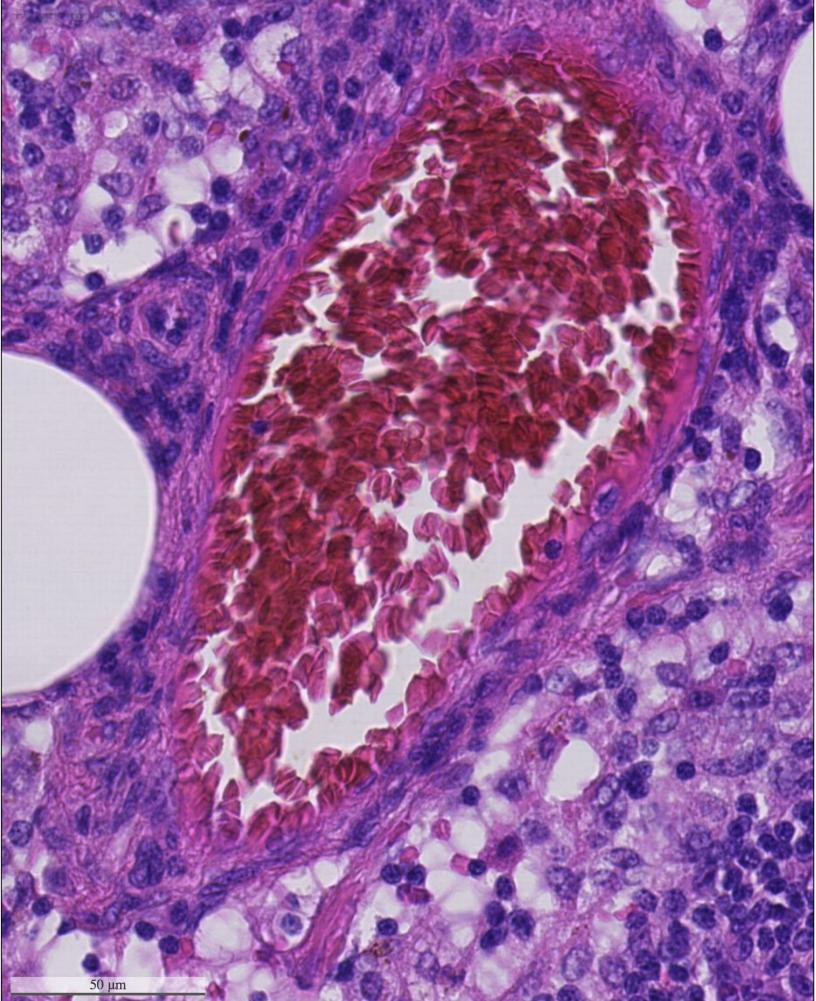
COMS 4995.07. Spring 19.



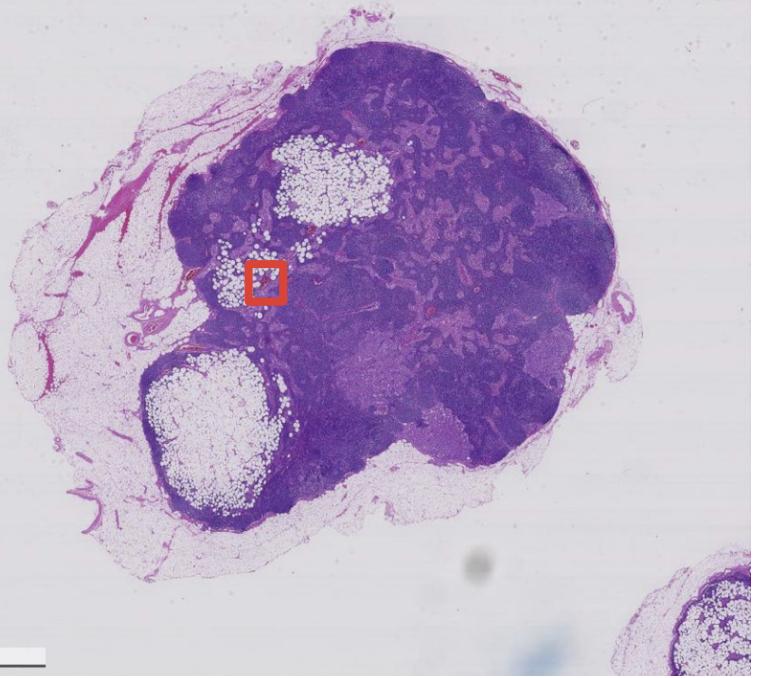
Lecture 5 - 22



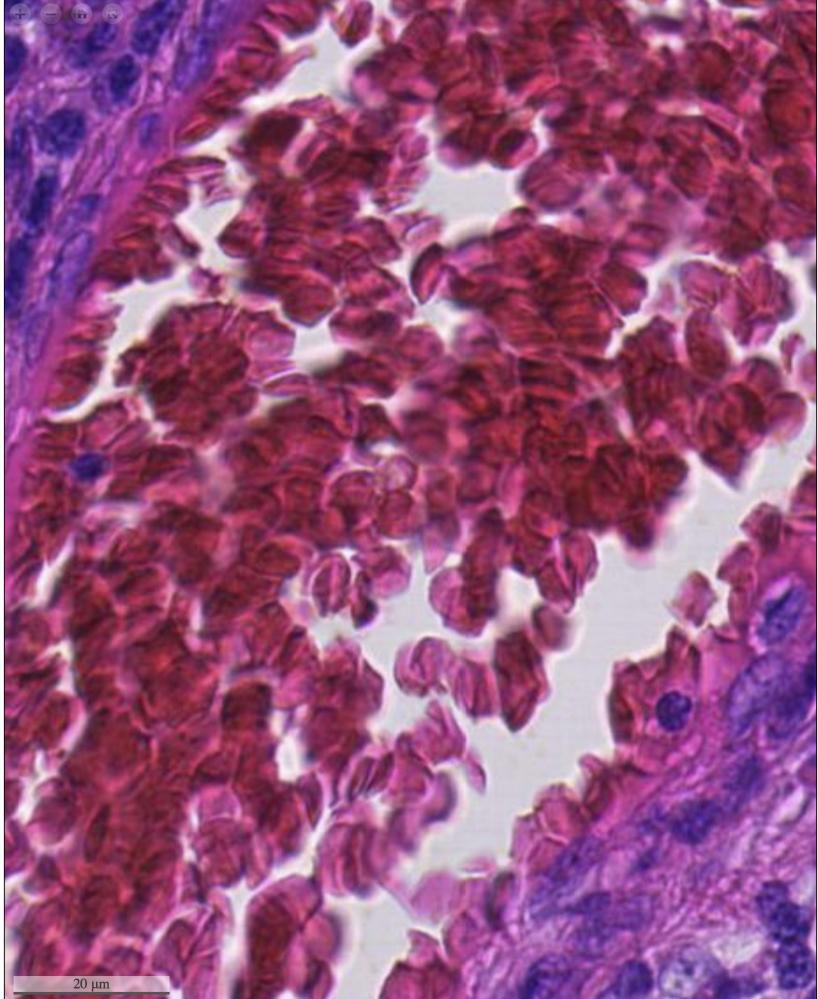
COMS 4995.07. Spring 19.



Lecture 5 - 23



7 magnification levels available per slide, up to 128x.



CAMELYON17 challenge

<https://camelyon17.grand-challenge.org/>

- Similar idea, except now with **patient level labels** (multiple slides from the same patient with a diagnosis), and of course, more data (around 3TB).

Recommend working with the CAMELYON16 dataset for the course project. Each slide is independent, it's simpler for building an end-to-end proof of concept that works reasonably well, without training on a huge dataset, or requiring a ton of GPU hours on Colab or elsewhere.

Your project should include

Code

- Train a model using a subset of the training data from CAMELYON16.
- Include a script to run it on a testing image and generate a heatmap (note: you do not need to create a 100,000 x 100,000 heatmap -- use a much lower resolution).

Results

- Design a **thoughtful evaluation method**, and include the results on at least three images from the testing set.
- Include a saved, trained version of your model with your submission.

Feasible without a fast machine or cluster

Tips from the paper

We provided starter code for this trick for you.

- “To reduce computation, we removed background patches (gray value > 0.8 and verified visually that lymph node tissue was not discarded.”
- “We surprisingly found that slimmed-down Inception architectures with **only 3%** of the parameters achieved similar performance to the full version”.
- Authors found that using a pretrained model on ImageNet ultimately hurt accuracy (Why? Pathology images do not match the natural images from ImageNet) – but – it **improved convergence speed**. This is a good place to start. **Try transfer learning and/or fine-tuning.**

Other tips

- Start with a low zoom level (much smaller images), a simple model, and a small amount of data.
- Implement an end-to-end prototype (training data in → heatmap on a test image out), then slowly scale up.
- Begin with transfer learning from a model trained on ImageNet, then fine-tuning, then consider training a model from scratch. Only after this is working end-to-end, consider data augmentation, using higher zooms, etc.
- Is Inceptionv3 still the best architecture to use?

Starter code + demo

On courseworks as project-starter-code.ipynb

- You will need to install OpenSlide first (both the C and Python libraries) before running.
- Let's look at the approach from [this paper](#).

Recommended approach to get started

Mechanics (software)

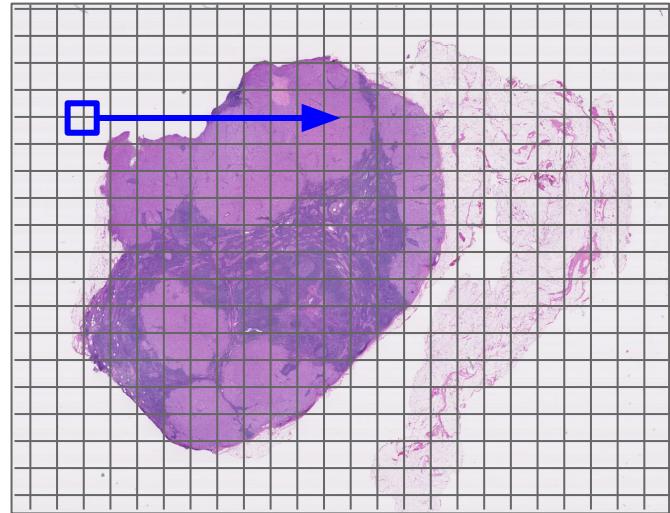
- Use the starter code! This will install [OpenSlide](#) (it's all you will need to read the training data we've provided). You do **not** need to use [ASAP](#) unless you'd like to create more training data.

ASAP may difficult to compile from source on a Mac, and may be difficult to install on Linux - when in doubt (don't laugh), download [VirtualBox](#), and a free [Windows VM](#) - use the prebuilt Windows installer for ASAP. You can map a directory on your local machine to the VM so you can open the images and masks linked from the starter code that you've already downloaded.

Recommended approach to get started

Mechanics (image processing)

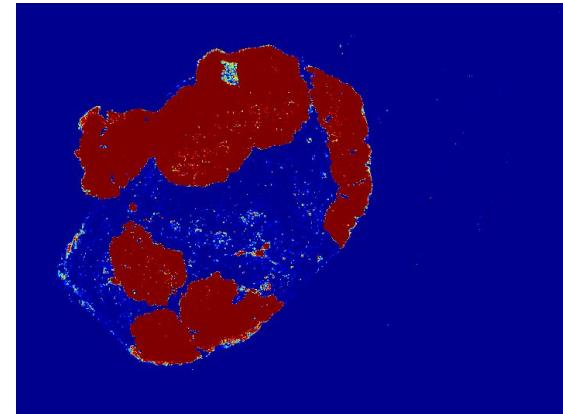
- Read this [paper](#) to understand their approach (discussed in a moment).
- Start small (use a single slide, at the lowest available zoom level - the effective resolution should be on the order of 1,000 by 1,000 pixels).
- Write code to slide a window across the slide. Extract patches and labels (using the tissue mask).



Recommended approach to get started

Modeling (proof of concept)

- Although transfer learning may be less effective than training a model from scratch , it is the **best place** to start.
- Choose a model previously trained on Imagenet. Use the techniques in Chapter 5 of Francois's book to try transfer learning (add a single Dense layer on top of that model), and train it on your own data.
- Write a script that takes your trained model and a testing image, and outputs a heat map showing the cancerous regions.
- Design an evaluation metric, write a script to report your results.



Model predictions

Let's see how this was accomplished in 2017.

Detecting Cancer Metastases on Gigapixel Pathology Images

1 out of 12 breast cancer biopsies is misdiagnosed

		Participating Pathologists' Interpretation				Total
Consensus Reference Diagnosis	Benign without atypia	Benign without atypia	Atypia	DCIS	Invasive carcinoma	
	Benign without atypia	1803	200	46	21	2070
	Atypia	719	990	353	8	2070
	DCIS	133	146	1764	54	2097
	Invasive carcinoma	3	0	23	637	663
Total		2658	1336	2186	720	6900

Underdiagnosis

Overdiagnosis

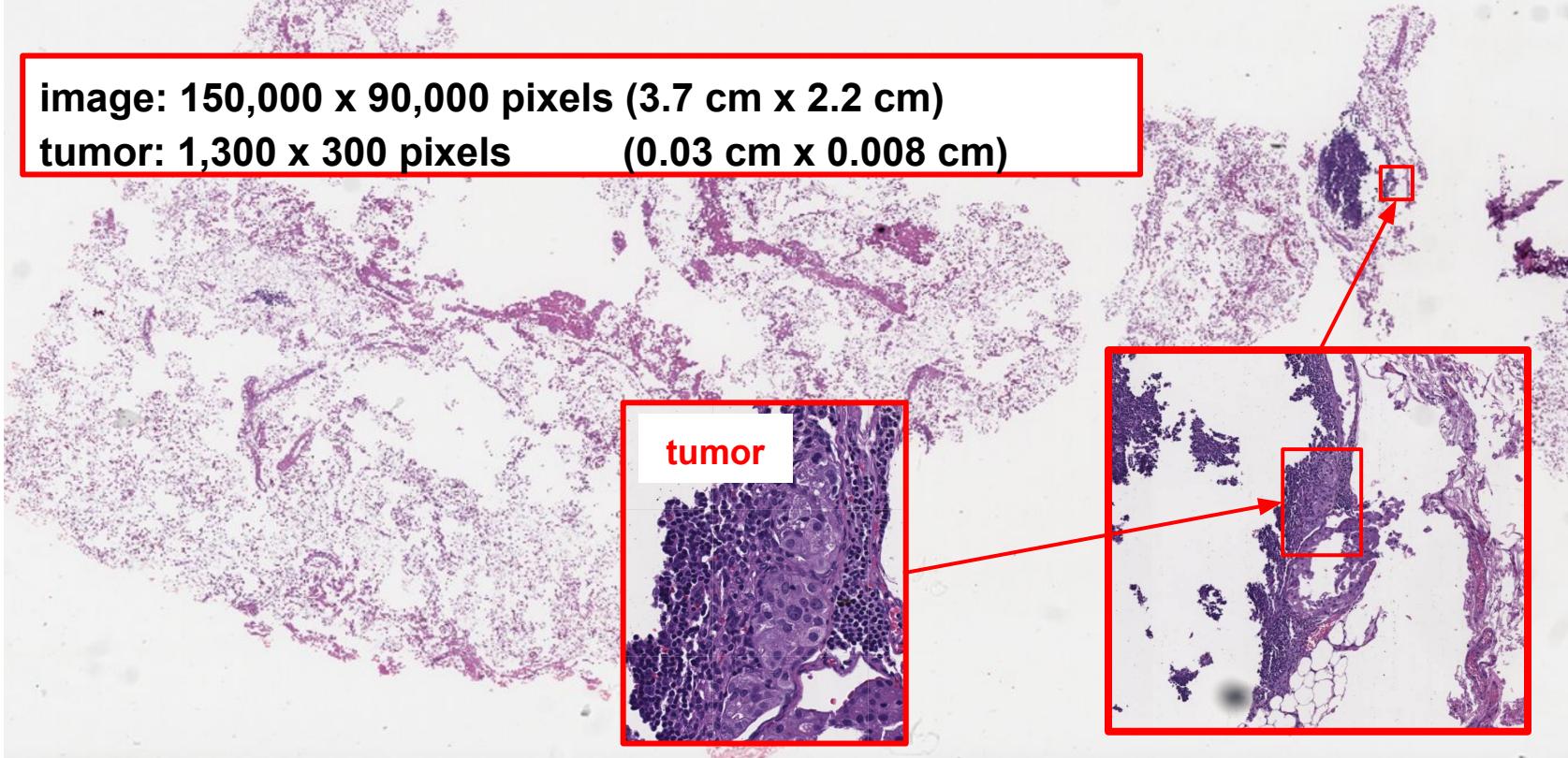
Correct diagnosis
87%
48%
84%
96%
75%

Diagnosis by three experienced pathologists, internationally recognized for research and continuing medical education on diagnostic breast pathology, before and after reaching a consensus agreement.

Diagnostic Concordance Among Pathologists Interpreting Breast Biopsy Specimens

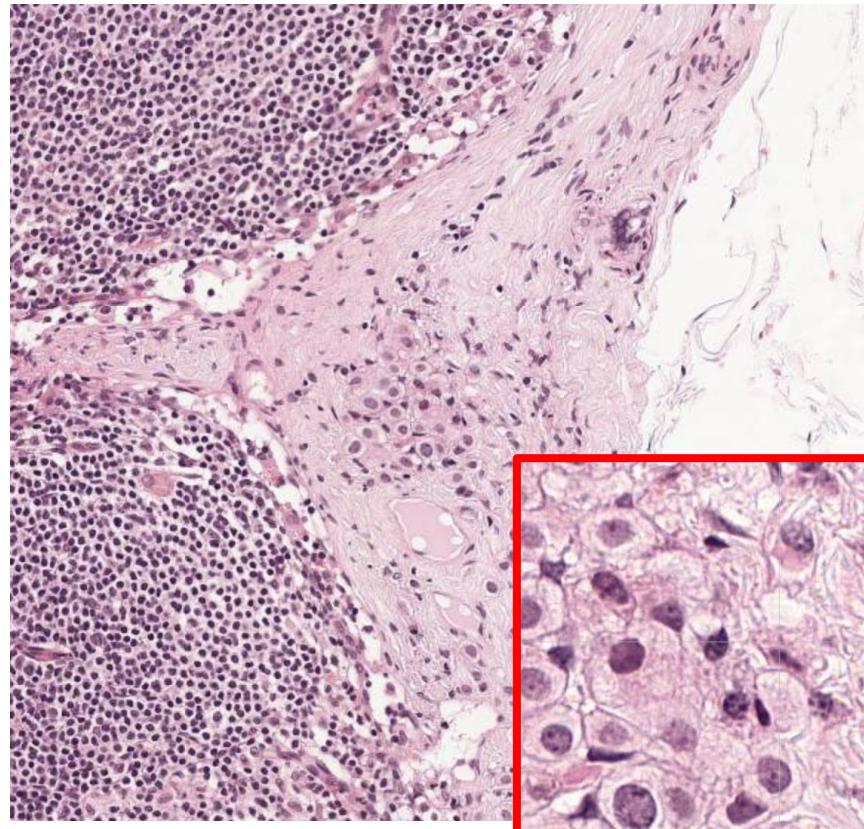
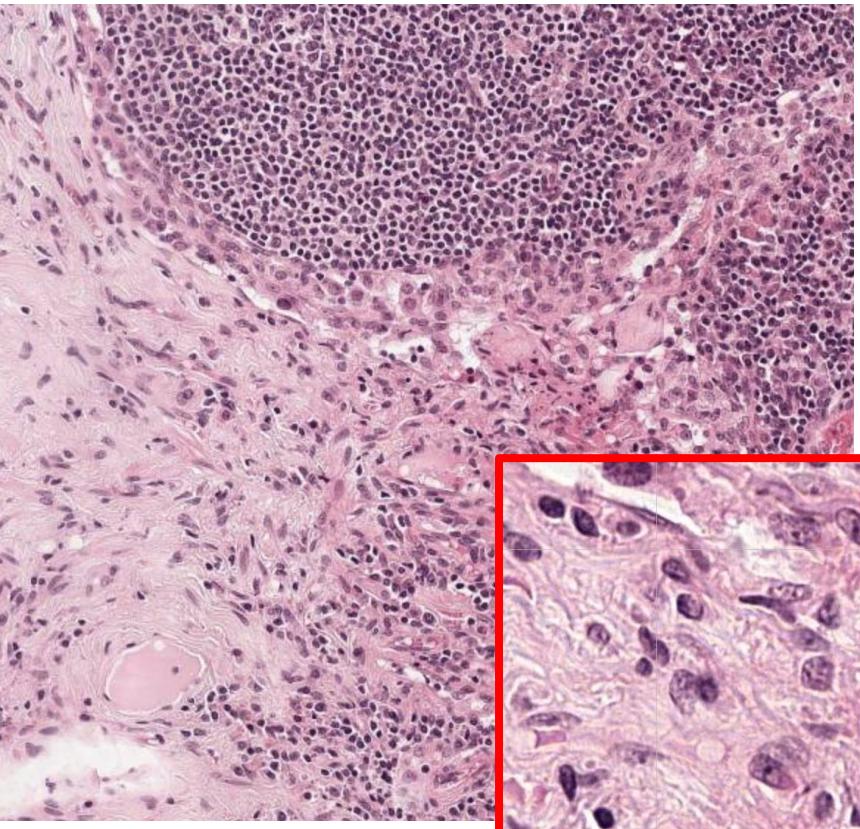
Why is it hard?

image: 150,000 x 90,000 pixels (3.7 cm x 2.2 cm)
tumor: 1,300 x 300 pixels (0.03 cm x 0.008 cm)

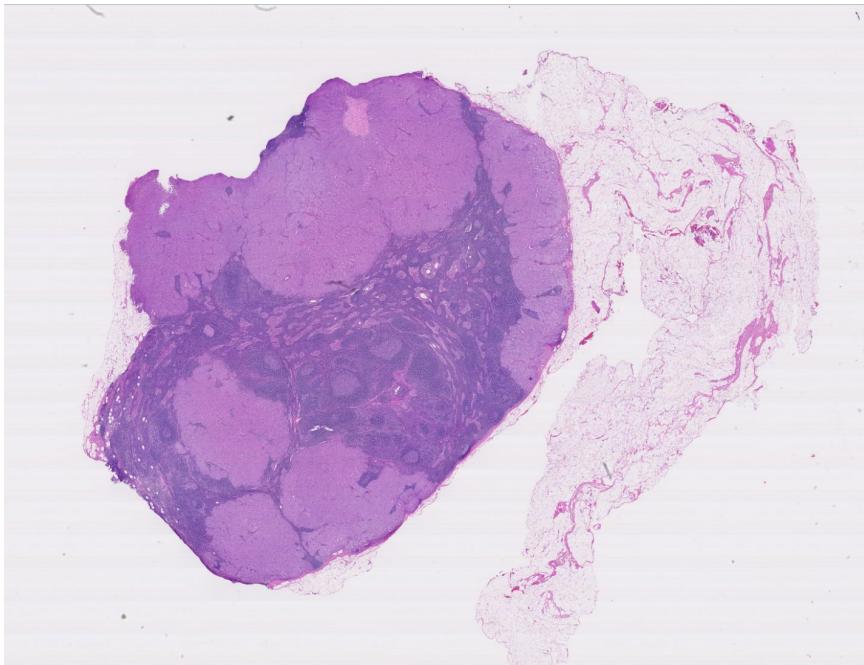


[Detecting Cancer Metastases on Gigapixel Pathology Images](#), 2017

Tumor mimics



Large images (> 100k x 100k pixels)

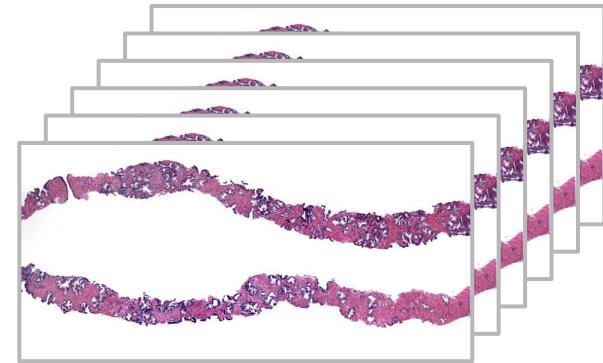
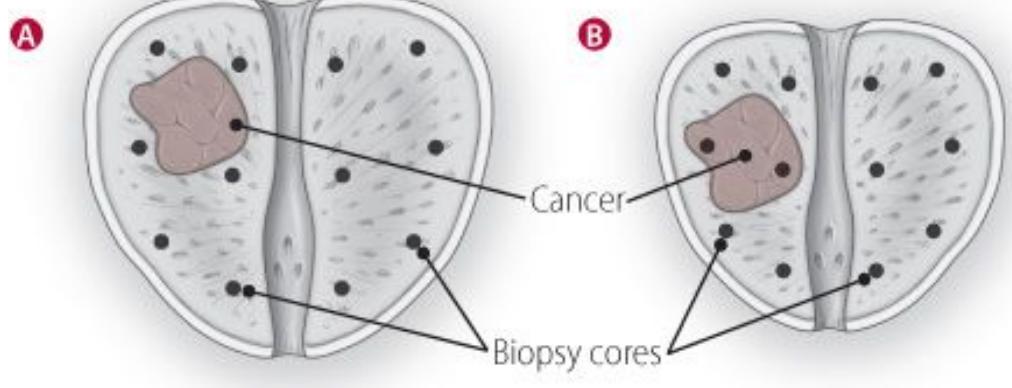


150k pixels (15 Gigapixel image)

1 biopsy image:

- 10,000,000,000 pixels
- ~1000 DSLR photos

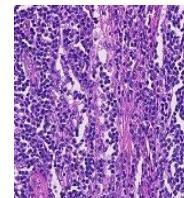
Multiple images per case



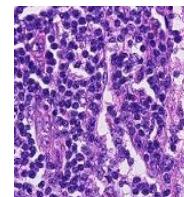
Work is being done now to process 3d volumes directly - this can be done with the convolutional layers you've already seen.

- 1) An insightful piece of work from observing the pathologists workflow - when examining an image, they often zoomed in and out with the microscope - looking at details, as well as context.
- 2) Using something like YOLO (which must consider the full image at once) it's feasible with this amount of data.

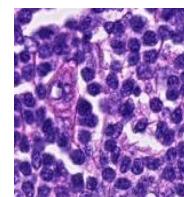
Quick discussion: what can we do?



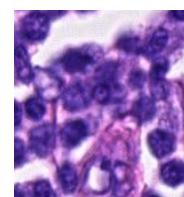
5x



10x



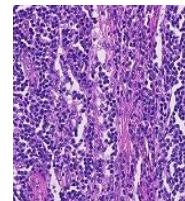
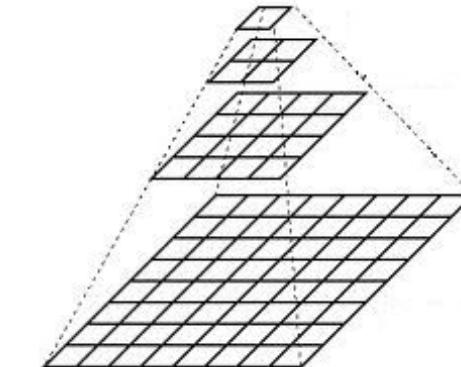
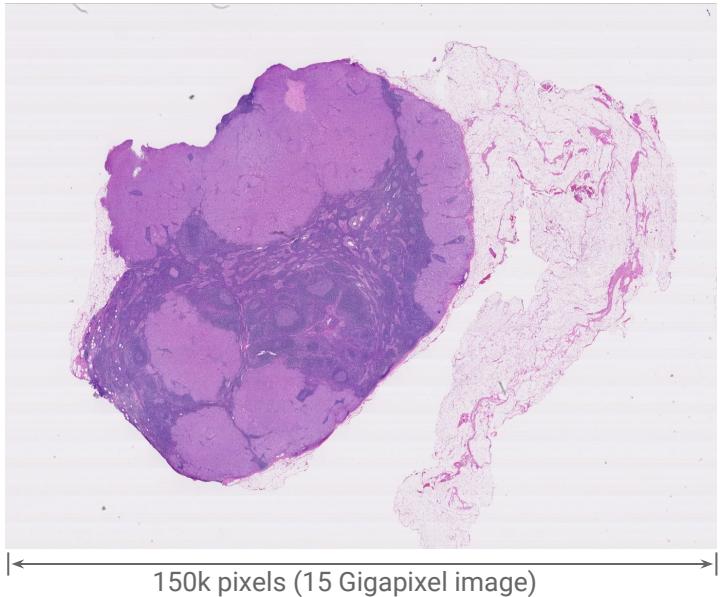
20x



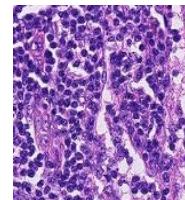
40x

[Detecting Cancer Metastases on Gigapixel Pathology Images](#), 2017

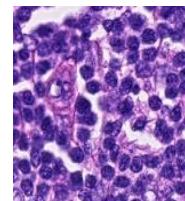
Multiscale problem - need detail as well as context



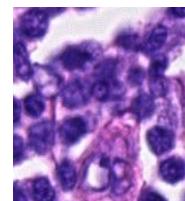
5x



10x



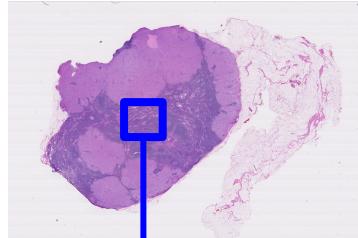
20x



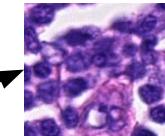
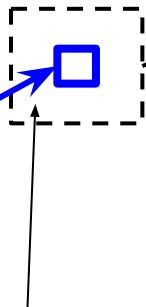
40x

[Detecting Cancer Metastases on Gigapixel Pathology Images](#), 2017

Patch based approach

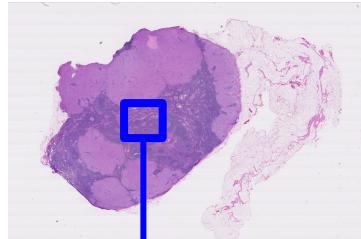


128 x 128 patch
of interest
(not to scale)

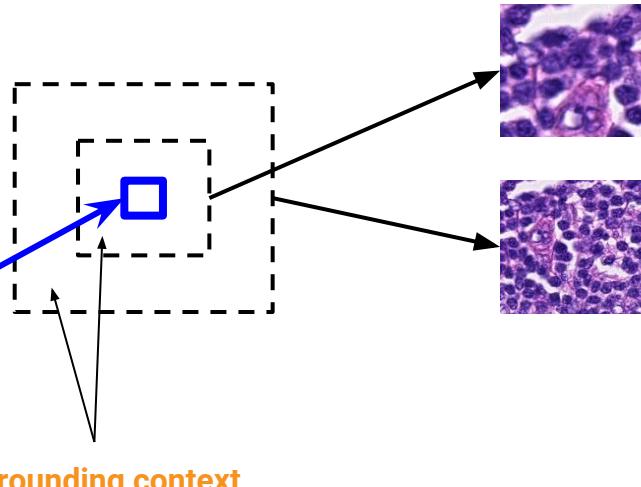


40X patch
(299 x 299)

[Detecting Cancer Metastases on Gigapixel Pathology Images](#), 2017



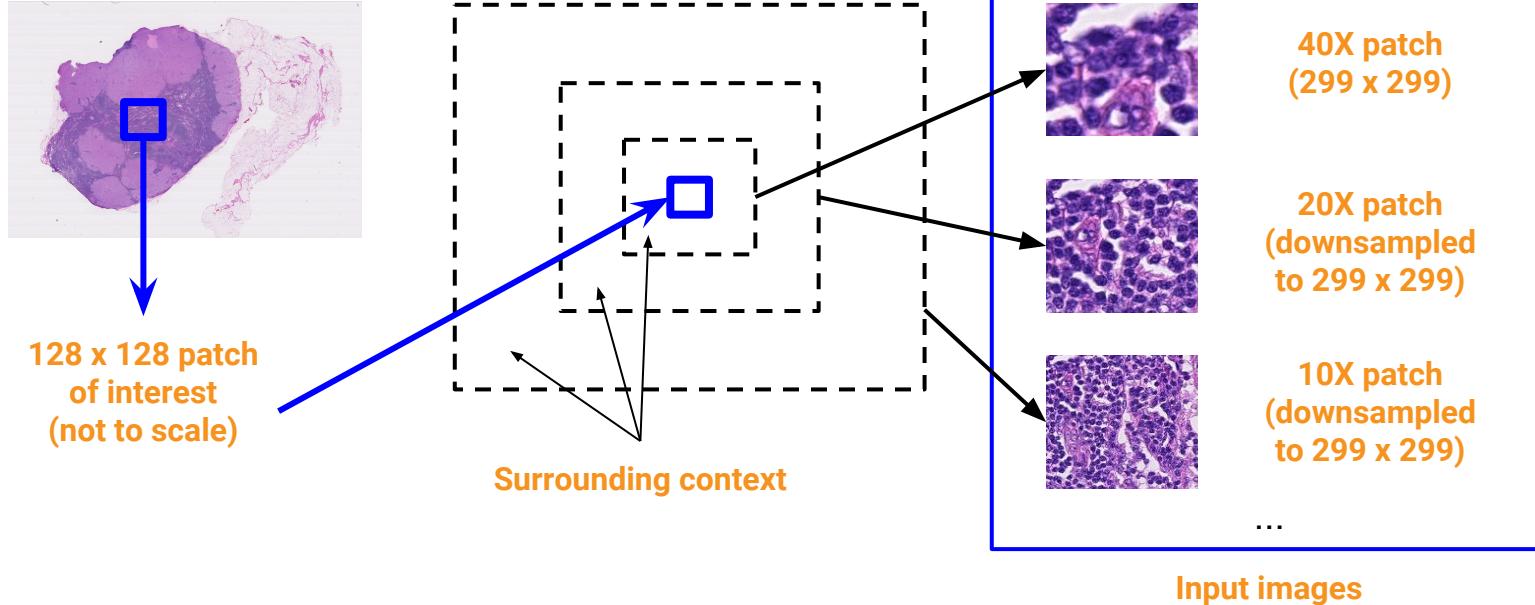
128 x 128 patch
of interest
(not to scale)



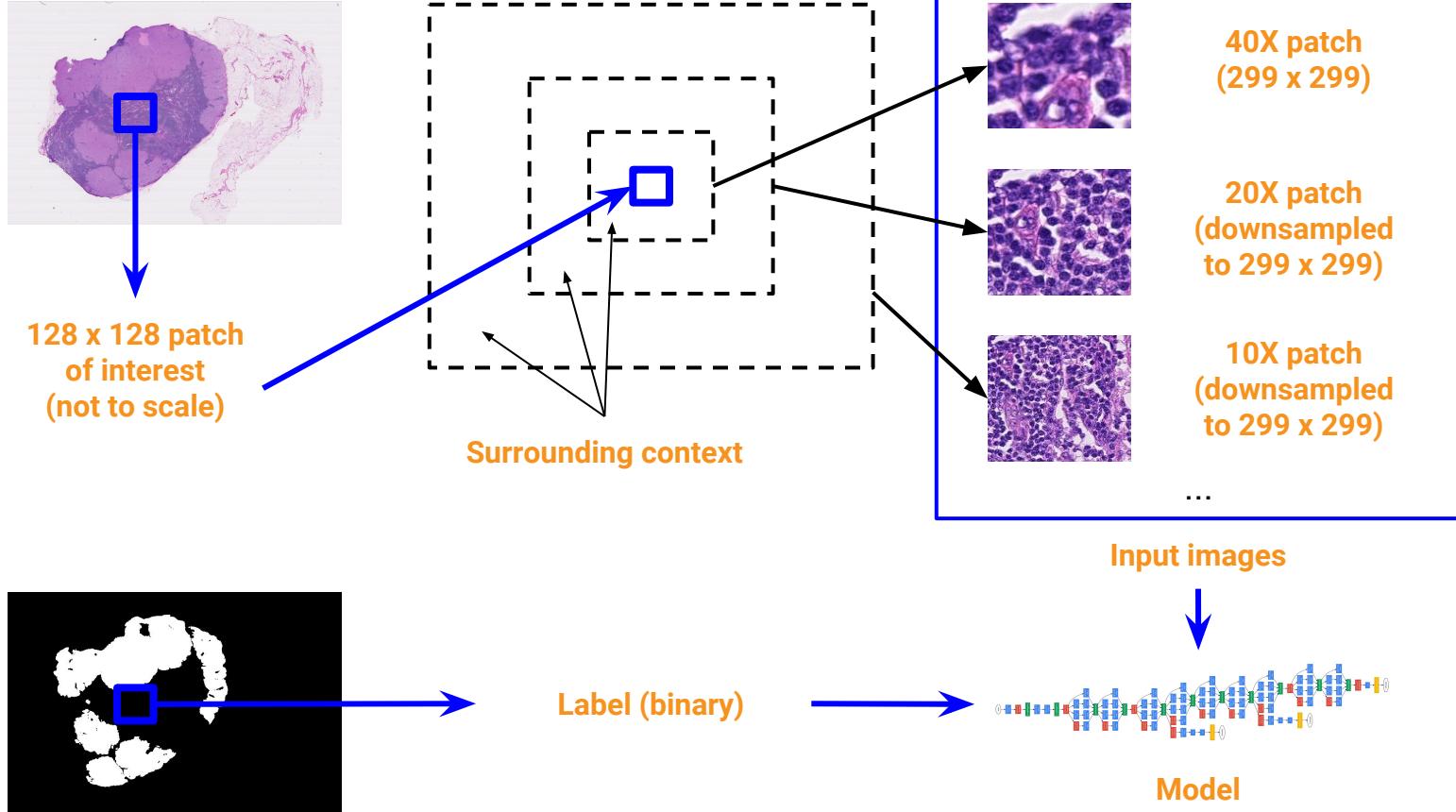
40X patch
(299 x 299)

20X patch
(downsampled
to 299 x 299)

[Detecting Cancer Metastases on Gigapixel Pathology Images](#), 2017



[Detecting Cancer Metastases on Gigapixel Pathology Images](#), 2017



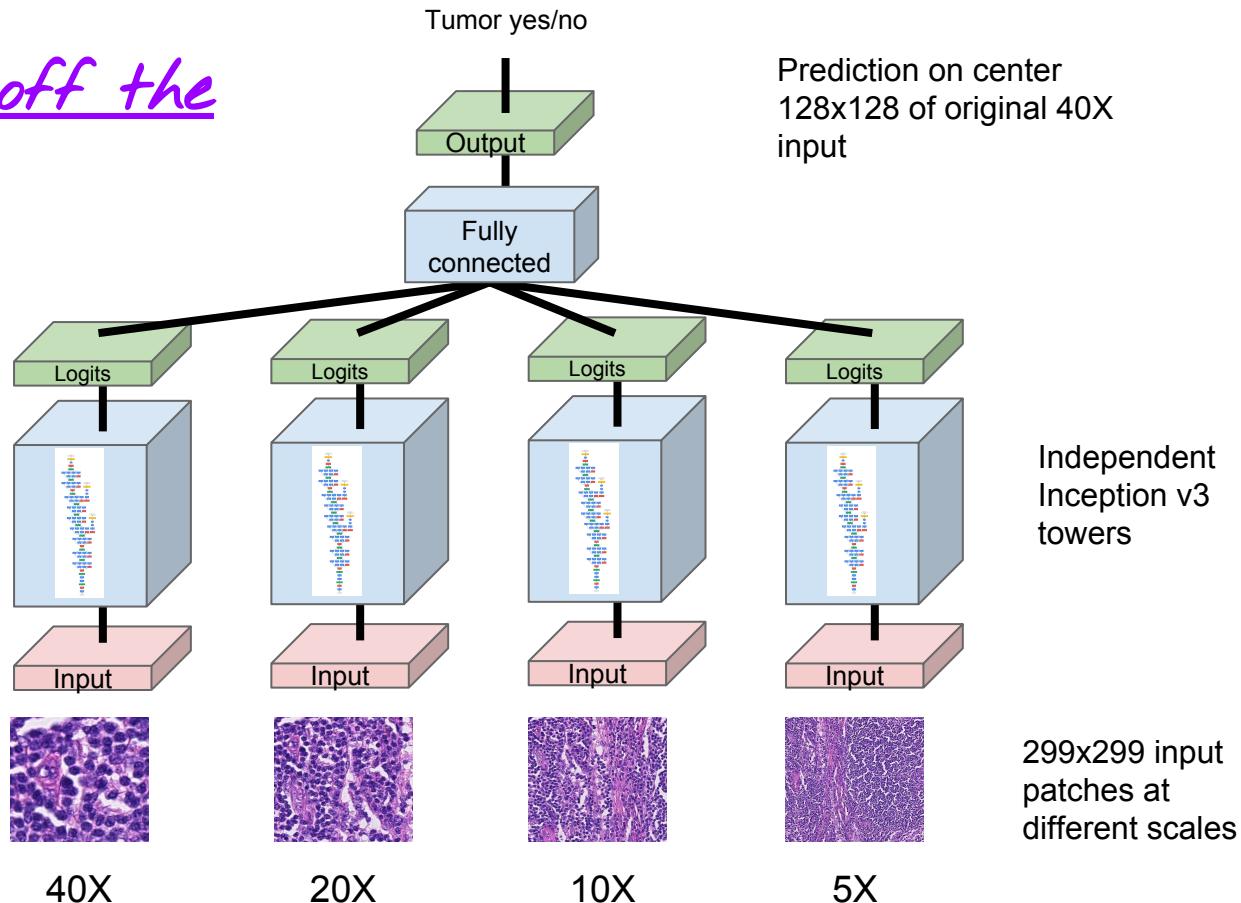
[Detecting Cancer Metastases on Gigapixel Pathology Images](#), 2017

Model is mostly off the shelf!

Multi scale model

detail \longleftrightarrow context

resembles microscope magnifications



[Detecting Cancer Metastases on Gigapixel Pathology Images](#), 2017

Data augmentation

- Brightness, saturation, contrast, hue
- Rotate / flip: 8X augmentation

This can result in an 8x to 32x increase in training data, depending on how you do it - don't try this when you get started (build end-to-end first, it becomes computationally difficult fast).

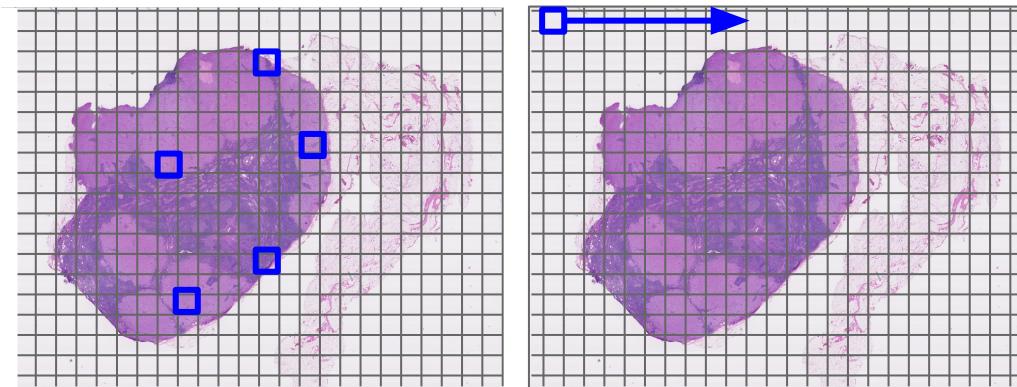
Train vs inference

Makes inference 8x more expensive.

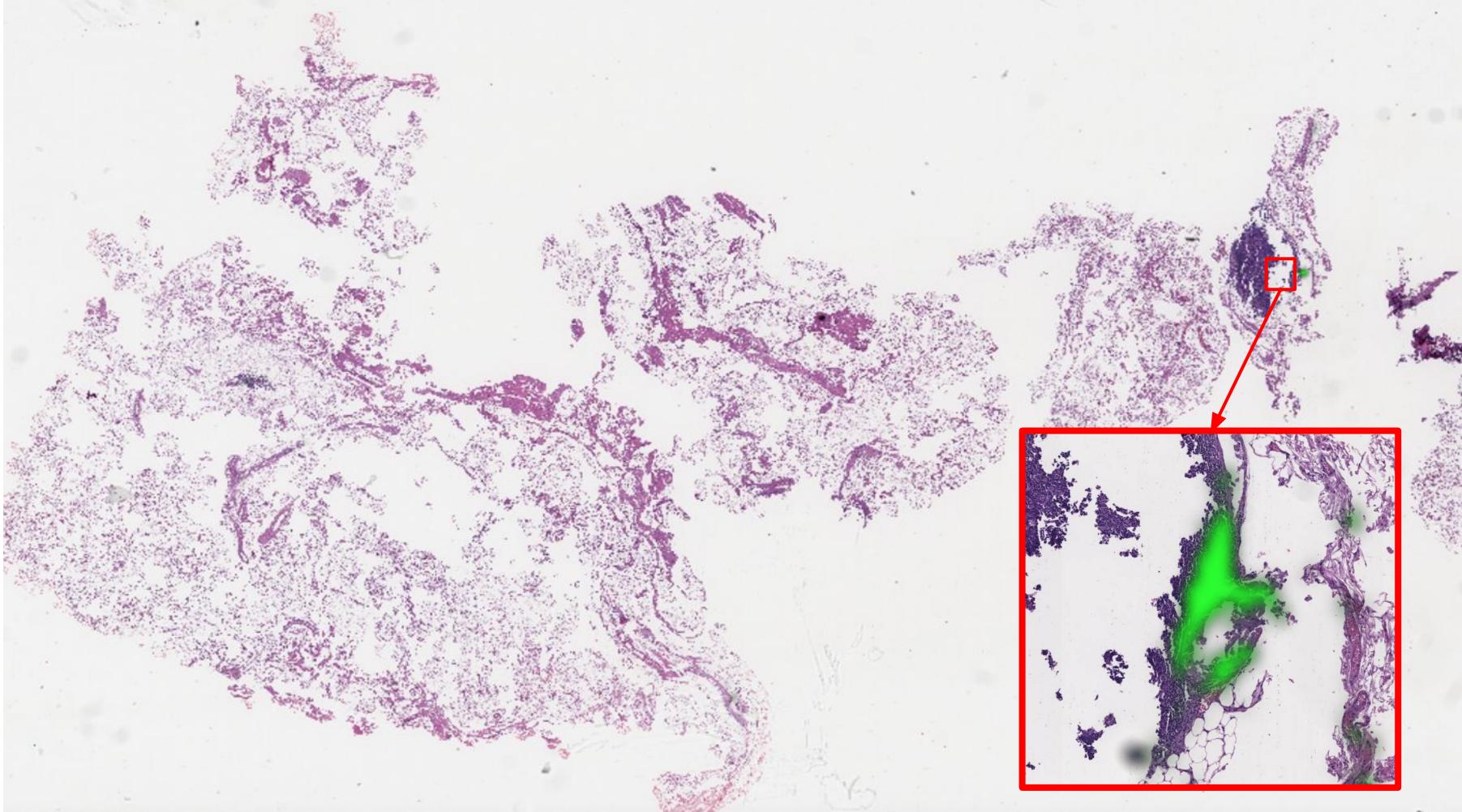
	Train	Inference
Patch extraction	Random	Sliding window
Color perturbations	Yes	No
Rotate / flip augmentation	Yes	Yes (average across)

These augmentations will be too expensive in the smaller scope version of this paper that you implement - start small and slowly scale up.

Quick discussion: How will you handle imbalanced data?



[Detecting Cancer Metastases on Gigapixel Pathology Images](#), 2017



Lessons learned

- **Generate actionable insights**
- Iterate quickly (get to **end-to-end as soon as possible**).
- Using pre-trained networks and transfer learning is important for fast coverage, especially when experimenting.
- Smaller models worked well.

What's next?

- Can this be done in real-time?

[Detecting Cancer Metastases on Gigapixel Pathology Images](#), 2017

An augmented reality microscope



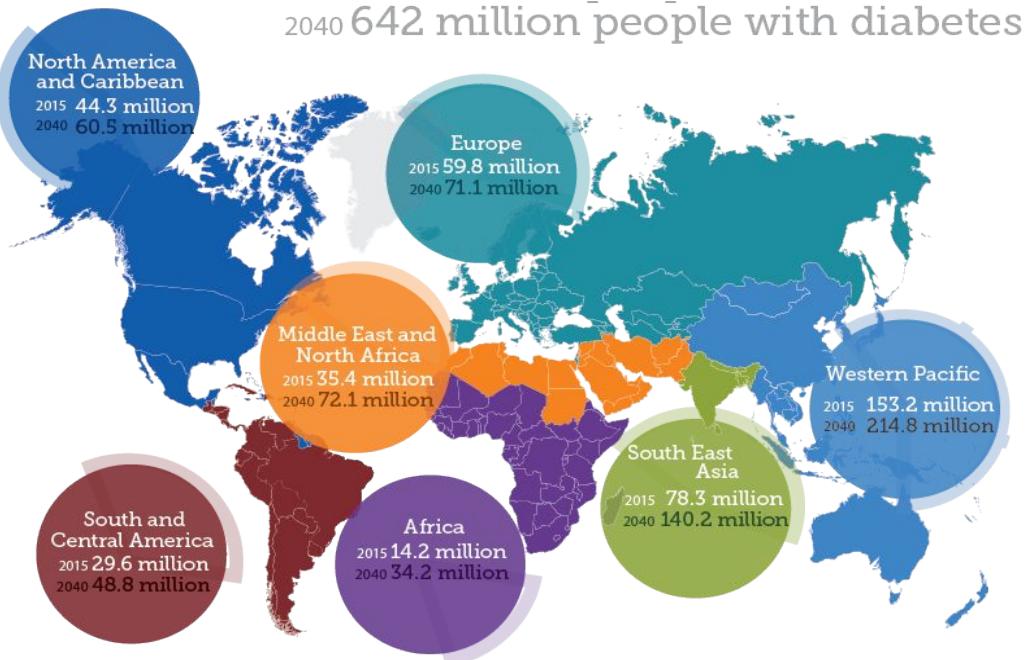
Personal opinion, this is cool (understatement!) - but I think a more "boring" static, offline analysis performed as an automatic second opinion is more practical in the medium future.

<https://ai.googleblog.com/2018/04/an-augmented-reality-microscope.html>

Questions?

Diabetic retinopathy.
Earlier, outstanding work. Similar ideas as
before, even simpler.

415M people with diabetes



Diabetic Retinopathy

- An eye disease that affects 1 out of every 3 people with diabetes.
- Can result in blindness if left untreated.
- Treatable once detected / early detection is important.

[Development and Validation of a Deep Learning Algorithm for Detection of Diabetic Retinopathy in Retinal Fundus Photographs](#)

Screened by analyzing a retinal fundus image

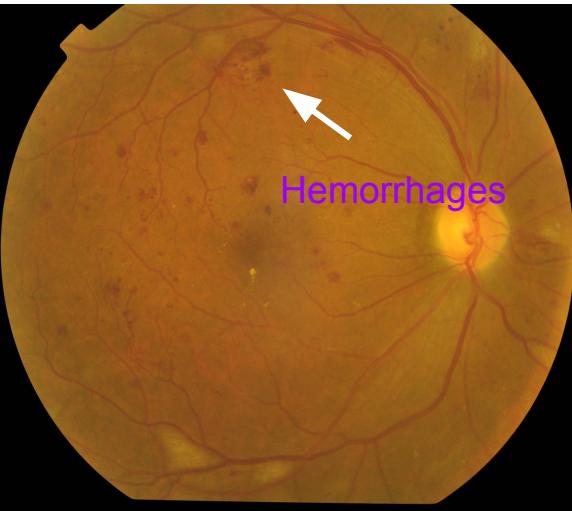


[Development and Validation of a Deep Learning Algorithm for Detection of Diabetic Retinopathy in Retinal Fundus Photographs](#)

Signs of Diabetic Retinopathy



A healthy eye



A diseased eye

In the early stages of the disease, microaneurysms develop in the blood vessels.

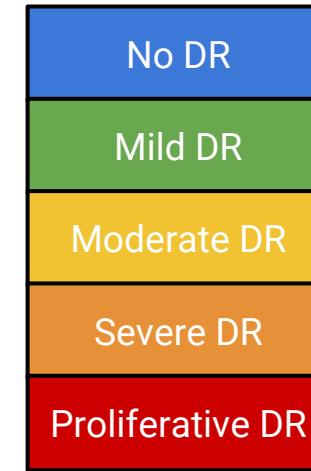
- Blocks oxygen to the eye
- Eye grows new blood vessels to compensate.
- Excess vessels results in vision loss.

Development and Validation of a Deep Learning Algorithm for Detection of Diabetic Retinopathy in Retinal Fundus Photographs

Data collection



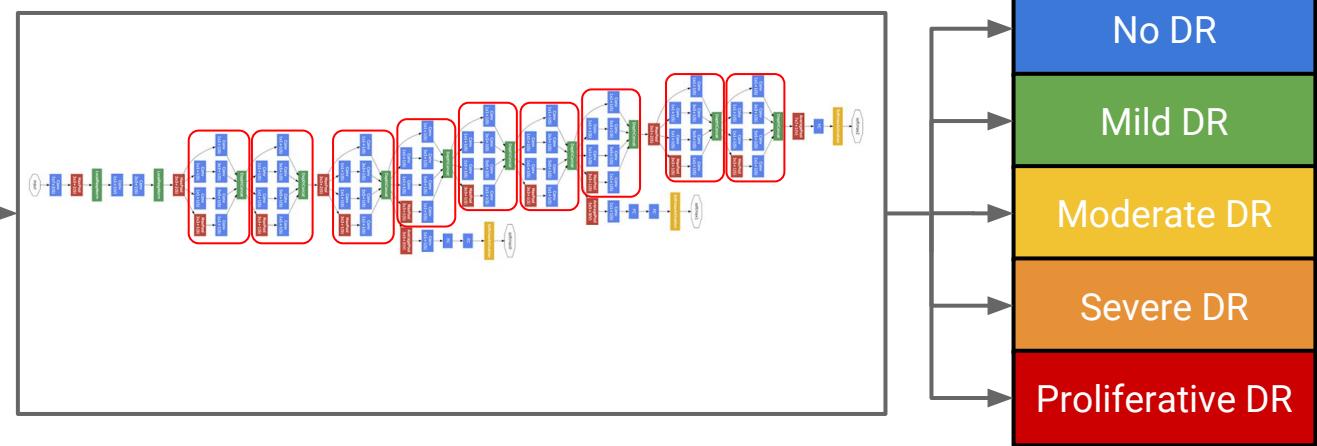
130k images graded by
54 ophthalmologists



[Development and Validation of a Deep Learning Algorithm for Detection of Diabetic Retinopathy in Retinal Fundus Photographs](#)

Similar idea as the previous paper

Unlike the Pathology work, here we're classify the entire image at once (downsized to a manageable resolution). In this case 299x299 (the default for this model).



Actually, this work came first - so I should the other paper had a similar idea to this one.

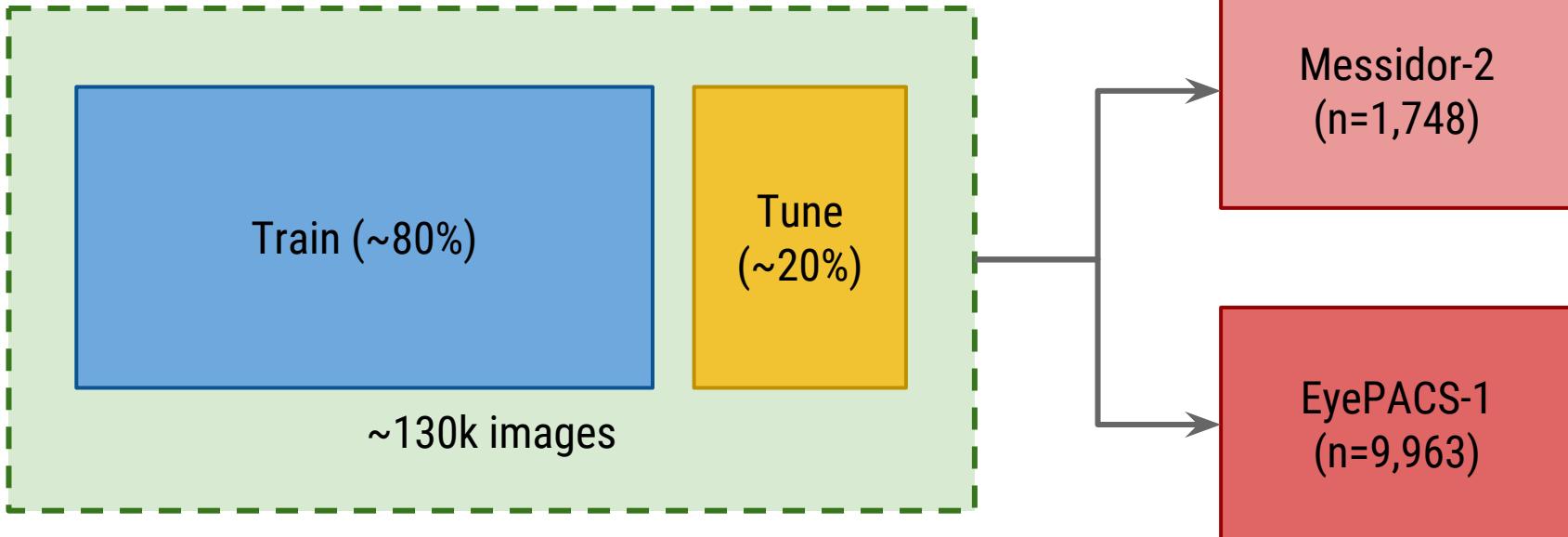
[Development and Validation of a Deep Learning Algorithm for Detection of Diabetic Retinopathy in Retinal Fundus Photographs](#)

This is important work **using a nearly off the shelf model and tools** with a thoughtful approach and interdisciplinary team. Many challenges in data science are not technical!

```
from tensorflow.keras.applications.inception_v3 import InceptionV3  
  
# create the base pre-trained model  
base_model = InceptionV3(weights='imagenet', include_top=False)
```

CLINICAL VALIDATION

DEVELOPMENT

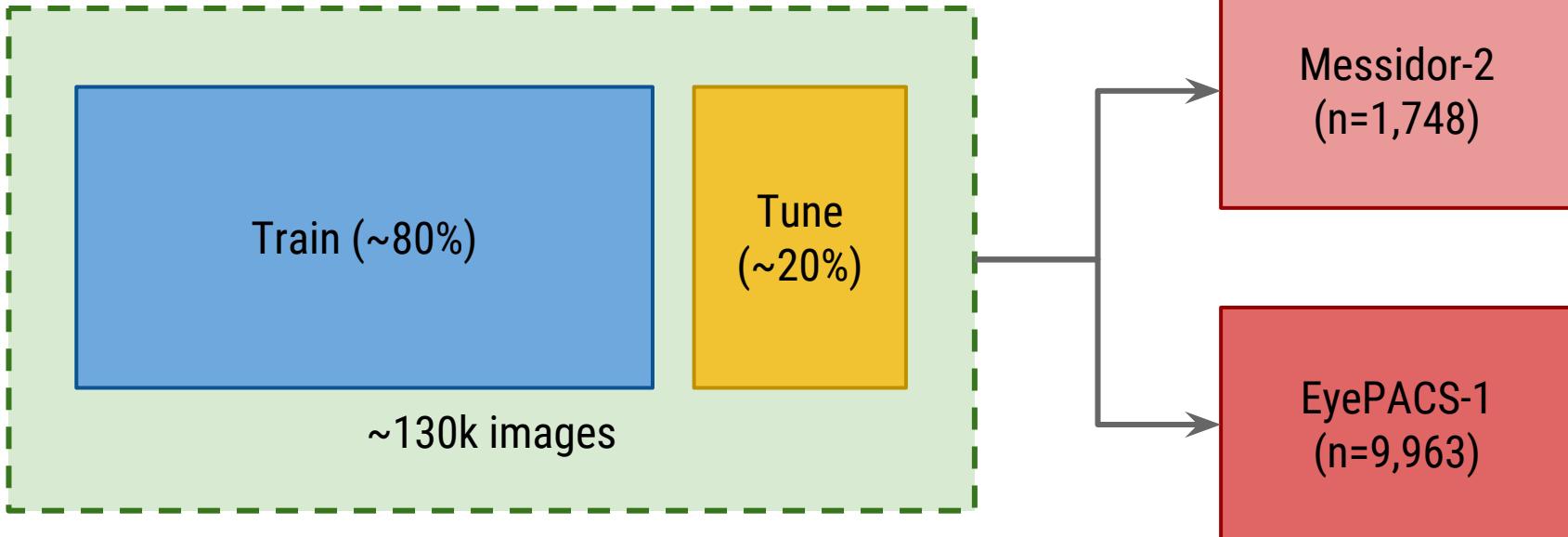


Quick discussion: if your training data is collected from many patients, and includes multiple images / patient, how should you divide it up for train / val? *Also:* why include test data from two institutions?

[Development and Validation of a Deep Learning Algorithm for Detection of Diabetic Retinopathy in Retinal Fundus Photographs](#)

CLINICAL VALIDATION

DEVELOPMENT



Note: make sure images from a single patient do not exist in both the train and tuning set. Multiple institutions = more diverse imaging equipment.

Development and Validation of a Deep Learning Algorithm for Detection of Diabetic Retinopathy in Retinal Fundus Photographs

To make this a **reality** across many specialities

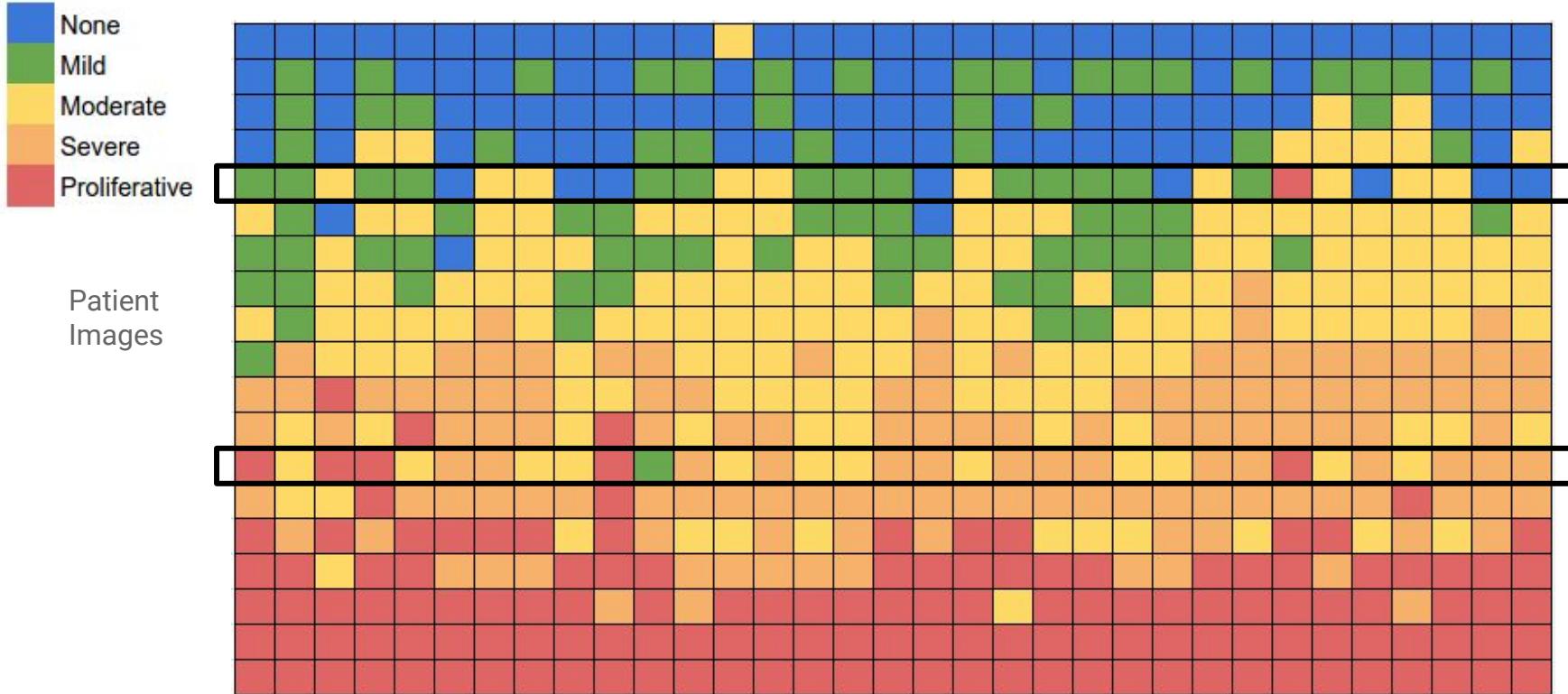
We need:

- High quality, publicly available and permissibly licensed datasets.
- Committed contributors with energy to actively maintain and expand them (in my experience, this is often the limiting factor).

The technical side, of course, can be further developed (3d convs, techniques to avoid this silly data augmentation we use currently).

- Already, super-physician accuracy is obtainable.
- Assistive tools like these **must** become common place.

Most of us are lucky to be able to even see a specialist (many worldwide are not -- we can use these tools to scale diagnostic capability as well), but even when seeing a specialist...



Consistency: intragrader ~65%, intergrader ~60%

Development and Validation of a Deep Learning Algorithm for Detection of Diabetic Retinopathy in Retinal Fundus Photographs

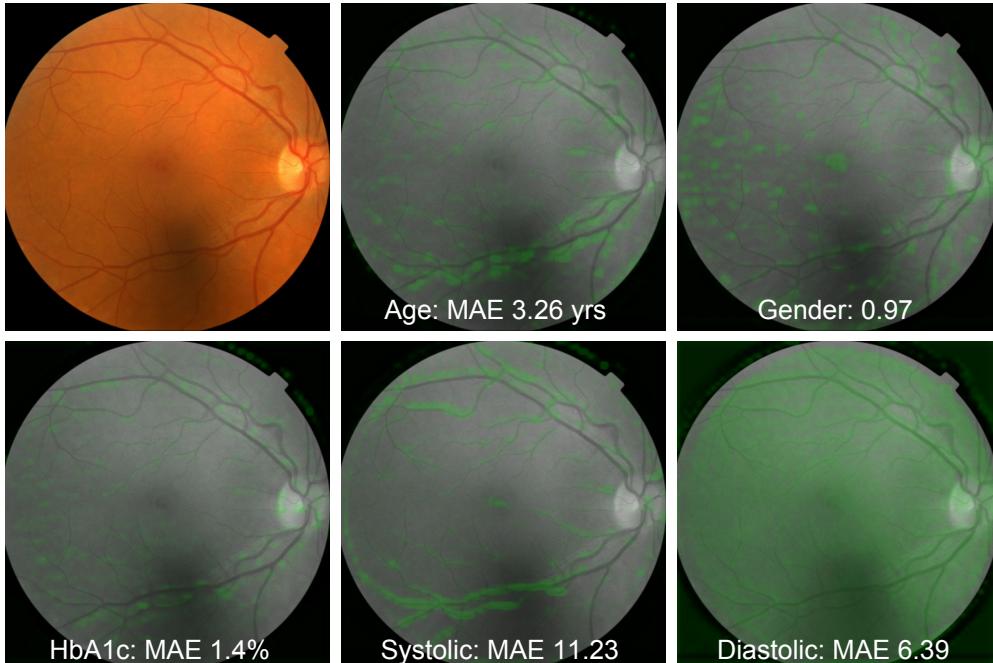
Determining ground truth with noisy labels

Several sensible strategies

- Majority vote
- Consensus labeling (labelers discuss and resolve discrepancies)
- Discard noisy labels/images
- Etc

An open area - no one right answer - think of what's sensible for your domain.

The future: supporting basic science



Can we predict cardiovascular risk? If so, this is a potentially non-invasive way of doing so.

As important: can we identify features used by the model to predict each task? (Could these potentially assist researchers doing basic science?)

Techniques from last lecture offer a simple, but reasonable start.

R. Poplin, A. Varadarajan et al. Predicting Cardiovascular Risk Factors from Retinal Fundus Photographs using Deep Learning. Nature Biomedical Engineering, 2018.

For next time

For next time

Reading

- Detecting Cancer Metastases on Gigapixel Pathology Images
- Development and Validation of a Deep Learning Algorithm for Detection of Diabetic Retinopathy in Retinal Fundus Photographs

Project descriptions

- **Will be posted to CourseWorks next week**