CamelyonAI

Automatic Detection and Classification of Breast Cancer Metastases for the Camelyon16 Grand Challenge

Terence Conlon, Aaron Sadholz
Columbia University

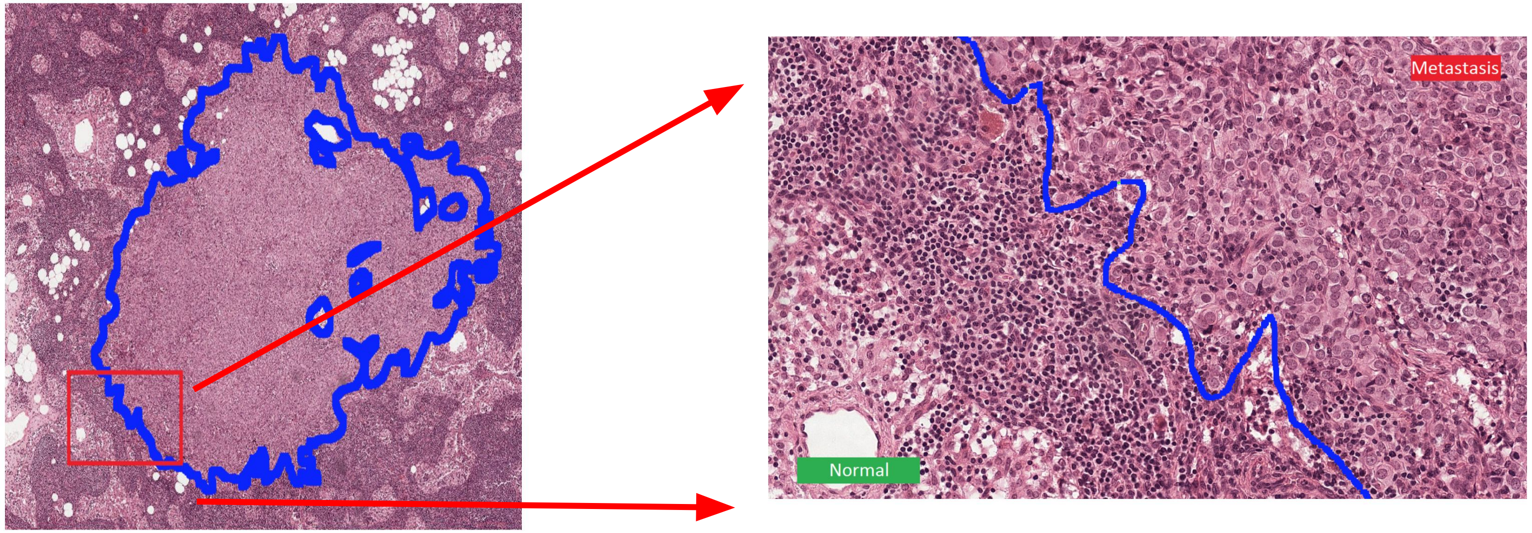
Overview and Methodology

Camelyon 16 Grand Challenge

Can a neural network help physicians detect breast cancer metastases?

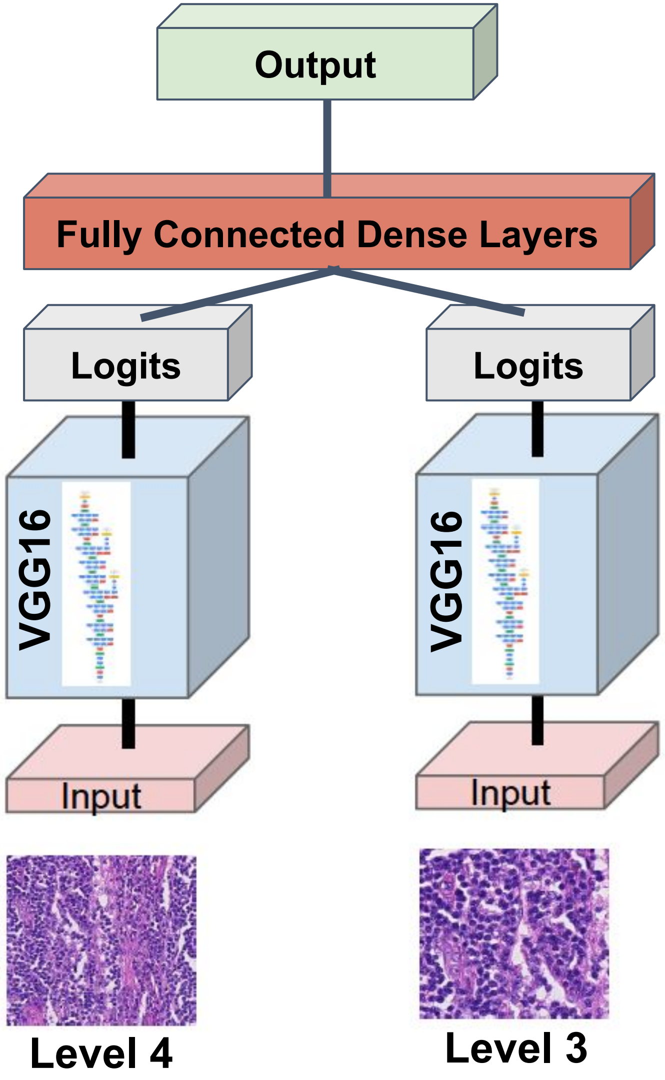
The Camelyon16 challenge supplied 400 whole slide images with tumors annotated by expert oncologists.

These images and annotations are the neural network inputs, i.e. the groundtruth necessary for training.



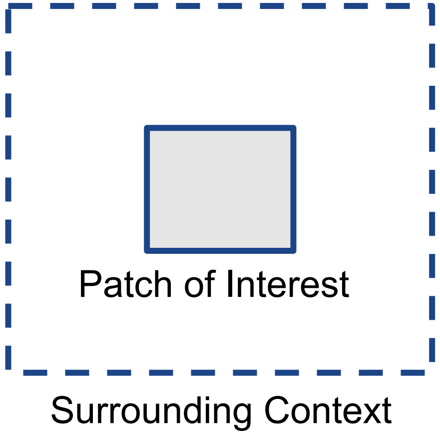
https://camelyon16.grand-challenge.org/

Parallel Network Architecture



Why a parallel architecture?

- Image context is important for tumor classification.
- By training on inputs at multiple zoom levels, CNNs can account for both local image characteristics and relevant surrounding features.



How is this implemented?

- Model subclassing with tf.keras
  - Two convolutional bases
  - Concatenation of convolutional base output, followed by fully connected dense layers.

https://arxiv.org/abs/1703.02442

Why not use Google Colab?

- Given the amount of training data and in order to train parallel models, a virtual machine with more than 12 GB RAM is necessary for training.

Virtual Machine Specifications:

- 8 vCPUs, 52 GB RAM
- 2 NVIDIA Tesla K80s, 24 GB RAM
- 125 GB Disk
- Openslide, TensorFlow-GPU, CUDA, etc.

Training, Validation, and Testing

Model Development

Evaluation Metric

- F1 Score on level 3 pixel basis

Data Split

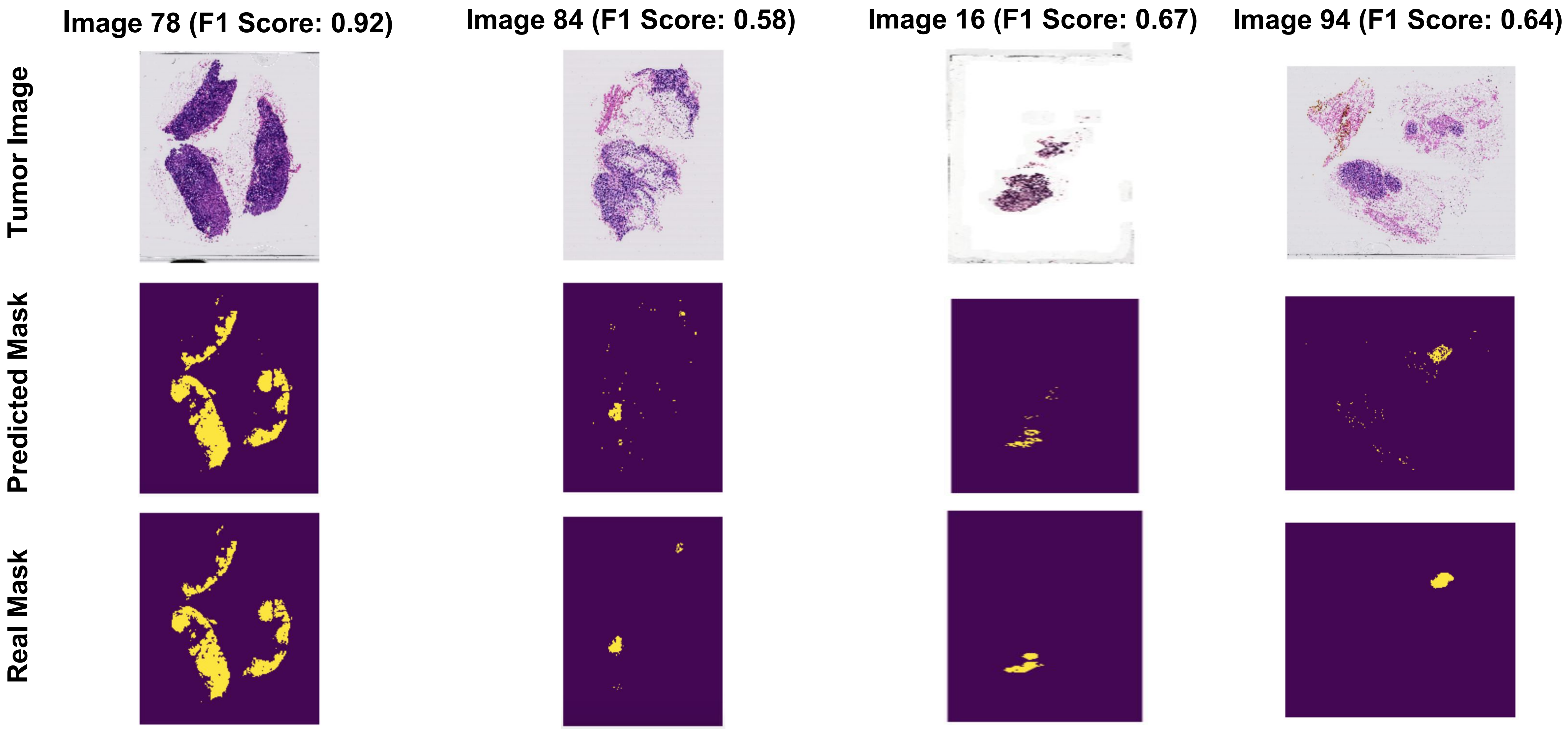
- Training set: 12 random tumors
- Validation set: 4 random tumors
- Test set: 4 random tumors

Model Tuning Process

1. Train model on training set
2. F1 score on validation set
3. Repeat steps 1 & 2 on potential models
4. Select best model (M4) - train on training+validation set
5. Evaluate on test set

	Model				
Features	**Baseline**	**M1**	**M2**	**M3**	**M4**
Image Levels	3	3,4	3,4	3,5	3,4
Conf. Threshold	70%	70%	70%	70%	85%
Conv. Base	VGG16	VGG16	Xception	VGG16	VGG16
Mean F1 Score	0.56	0.64	0.61	0.57	0.65

Optimized Model Performance -- Model M4



Conclusions

Primary Takeaways

Conclusions:

- Parallel networks allow for more accurate tumor detection.
- VGG performs better than Xception as a convolutional base in this scenario.
- Training on slides levels 3 and 4 results in the highest accuracy.
- Overprediction can be mitigated by adjusting the confidence threshold.

Future Work:

- Lower level images (1 & 2)
- Additional parallel models
- Larger training and validation sets
- Higher granularity model tuning
  - Epoch count
  - Alternative convolutional base models
  - Image augmentation
  - Number/size of dense layers after transfer learning, before output

Github Repository

Thank you!

Questions?