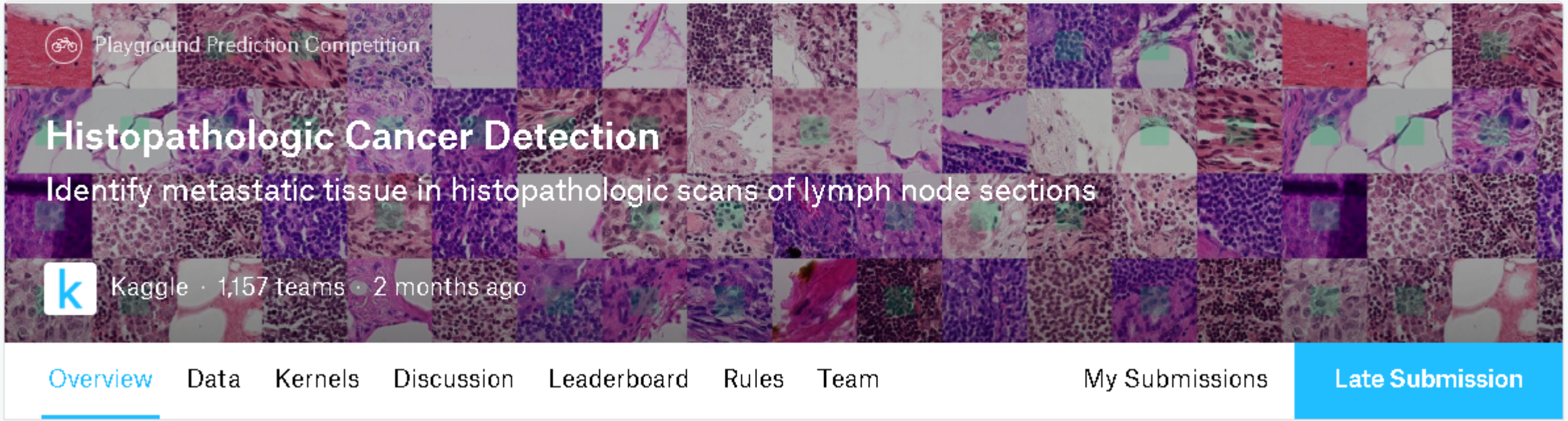
Histopathologic cancer detection using Convolutional Neural Networks

By:

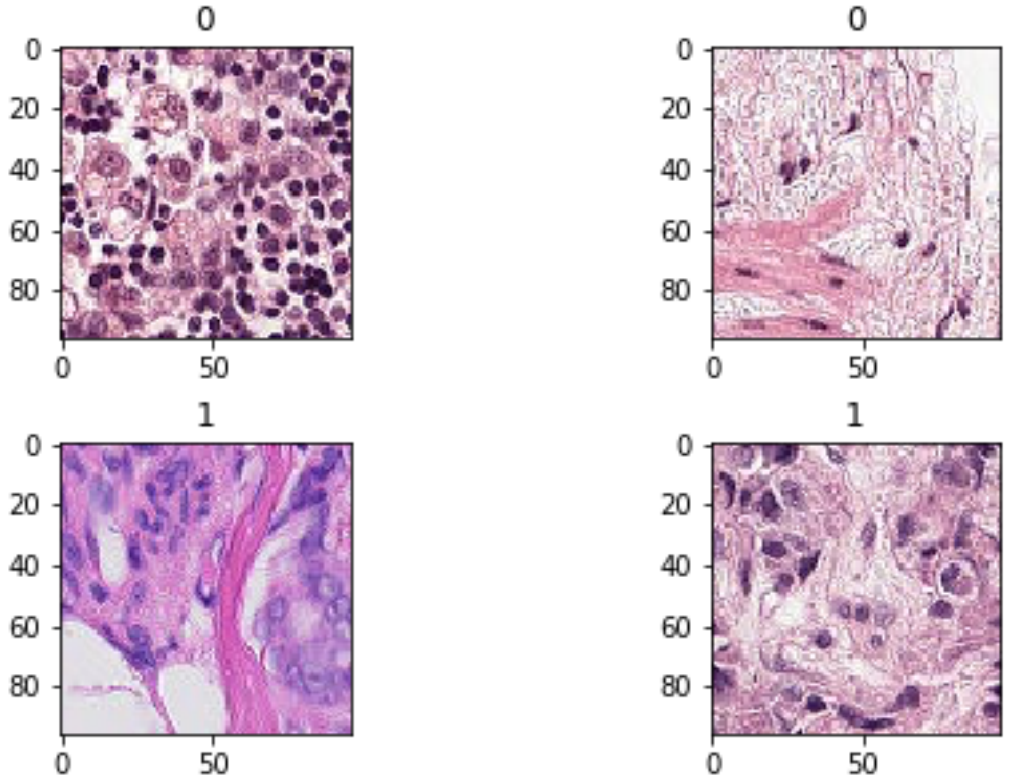
Hampus Rosvall



- Kaggle
  - Primarily classifications competitions on various types of data using Machine Learning
  - Around 1200 participants in my competition
- Histopathology – examination of changes human tissue

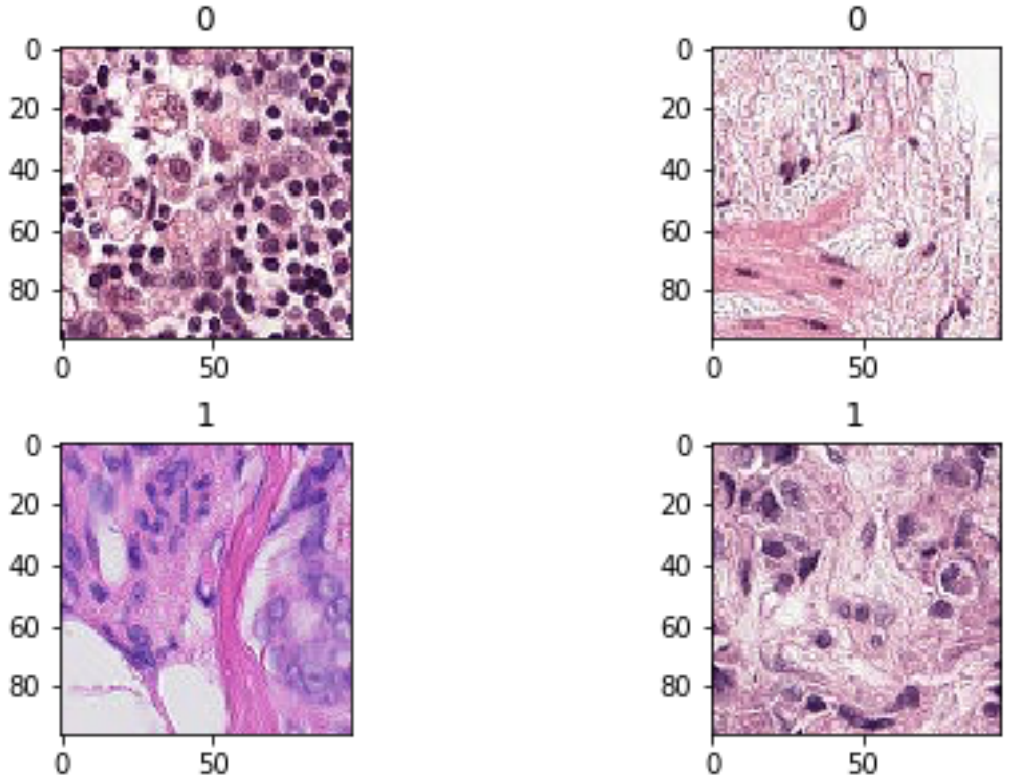
The data

- Classify wheter small microscopic images of human tissue have tumour present or not
- Input images of the size (96, 96) in RBG
- Hard to tell the difference between positive and negative class



The data

- 220025 training and 57458 test images.
- Distribution amonst training images
  - 130908 negative examples
  - 89117 positive examples

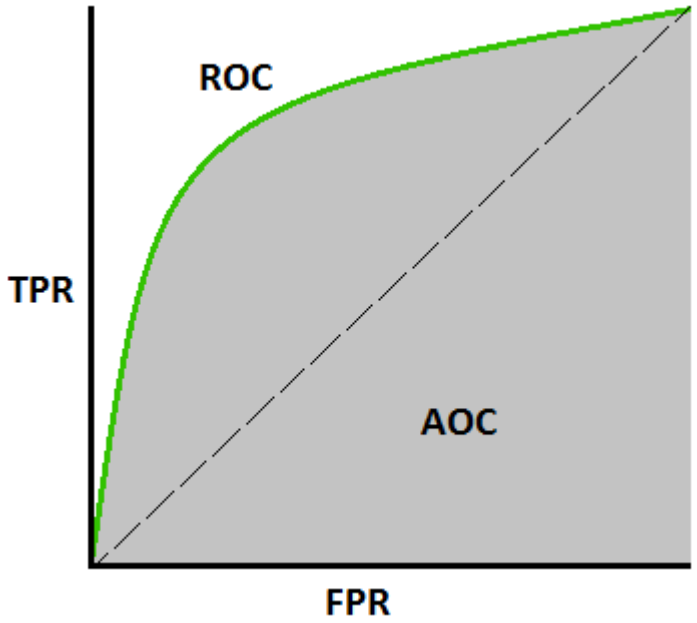


Data preprocessing

- Normalize data before feeding it to the neural net
  - Subract mean and divide by the standard deviation for each of the training example
- Image augmentation as one experiment – more on that later

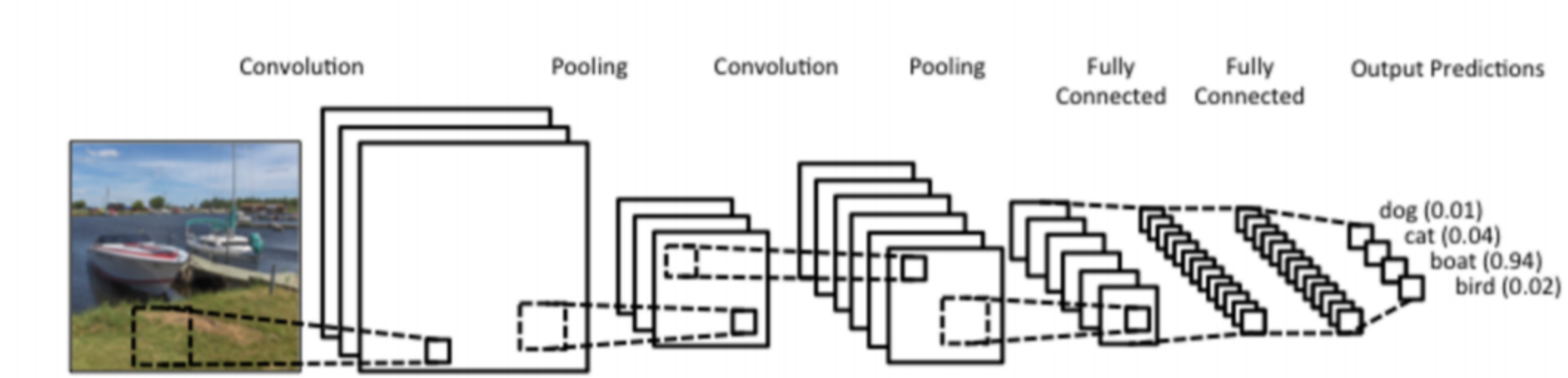
Evaluation metric

- Predetermined evaluation metric on Kaggle
- ROC AUC score
  - Plots TPR (true positive rate) versus FPR (false positive rate)
  - Score is based on the area under the curve (0.5 for a purely random classifier and 1 for a perfect one)



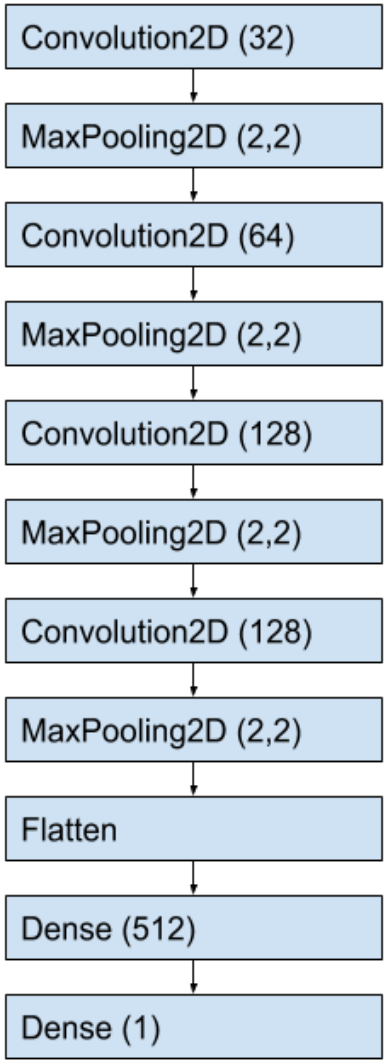
Implementation and CNNs

- Trained on a Google Cloud virtual machine
  - Nvidia TESLA K80 GPU
- Implemented using Python and high-level library Keras and the use of Neural Networks
  - Convolutional layers
  - Fully connected layers



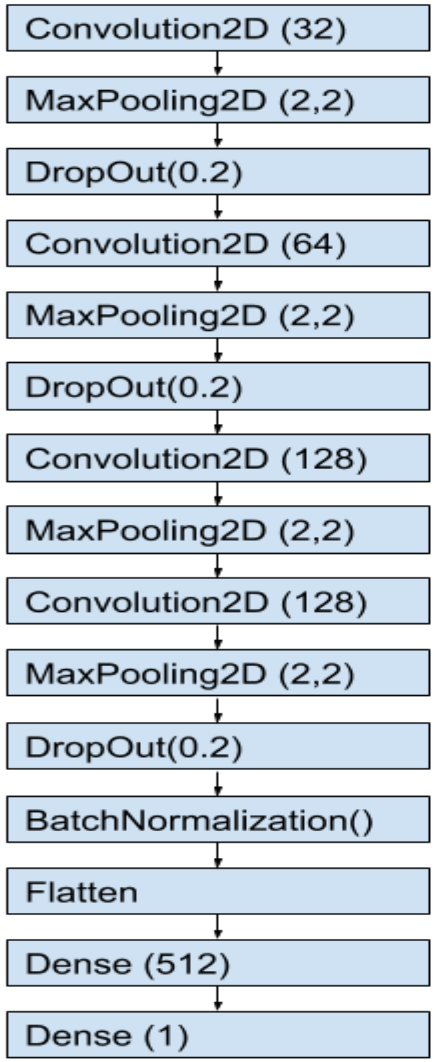
Architectures – user defined

Baseline



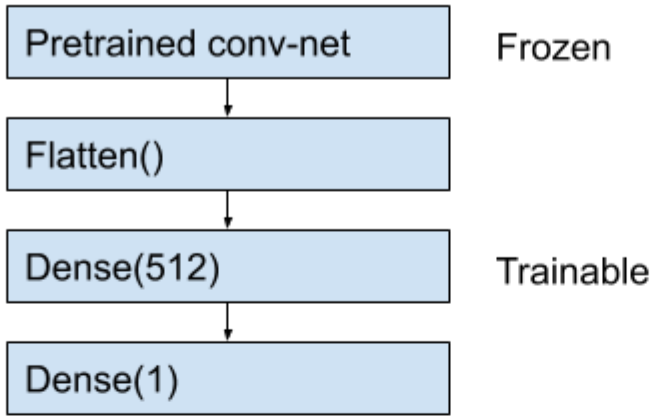
DropOut and BatchNormalization

Final architecture



Architectures – pre trained

- Tried two pretrained convolutional neural networks
  - VGG16 – 23 layers
  - InceptionResNetV2 – 572 layers
- Both pre trained on ImageNet data.



Results

Model	Score
Baseline	0.9358
VGG16	0.9347
Image augmentation	0.9279
InceptionResNetV2	0.9485
Final	**0.9530**

Suggestions for improvement

- Ensemble learning
- More extensive image augmentation
- Build a classifier that classes more false negatives than false positives.