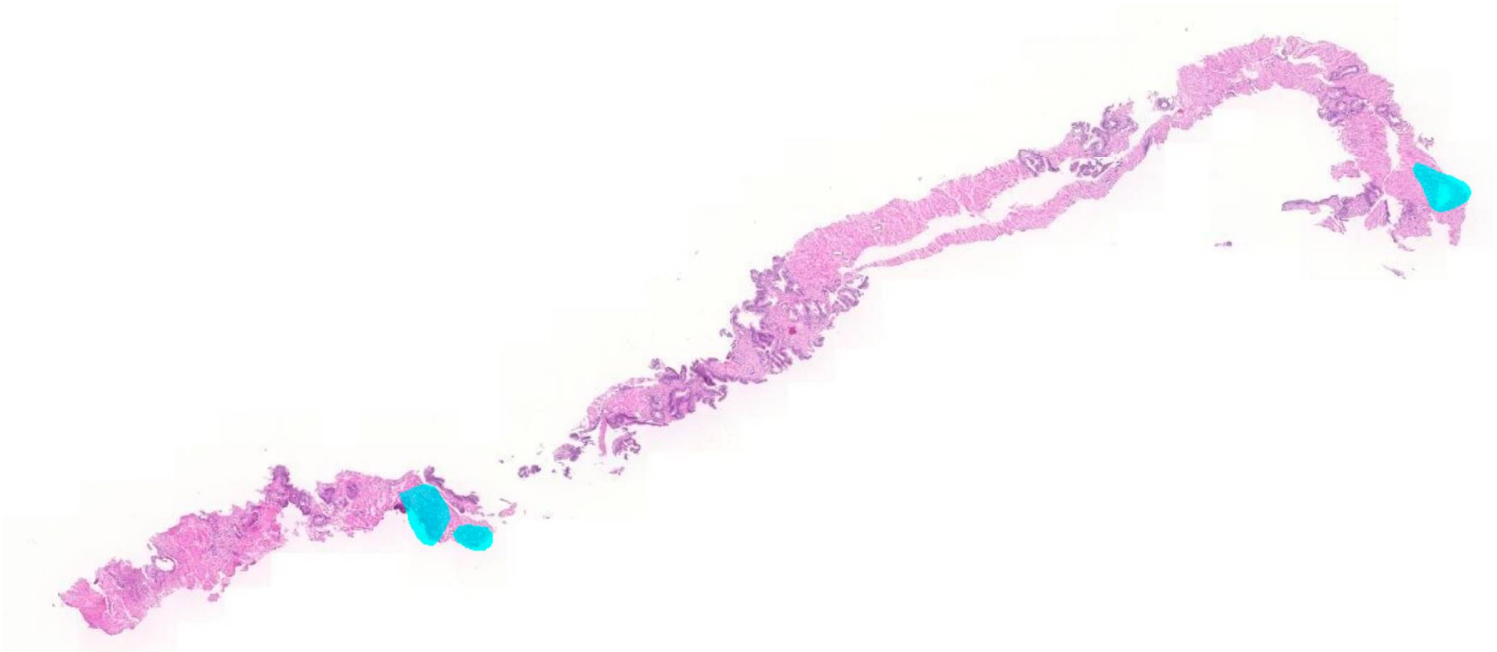
Machine Learning & Understanding

- Working with manually annotated whole slide images (WSIs)
  - annotations of cancer vs. non-cancer regions, exclude regions
  - using non-cancer regions from healthy patients
  - rate limiting steps are manual annotations

A prostate sample WSI

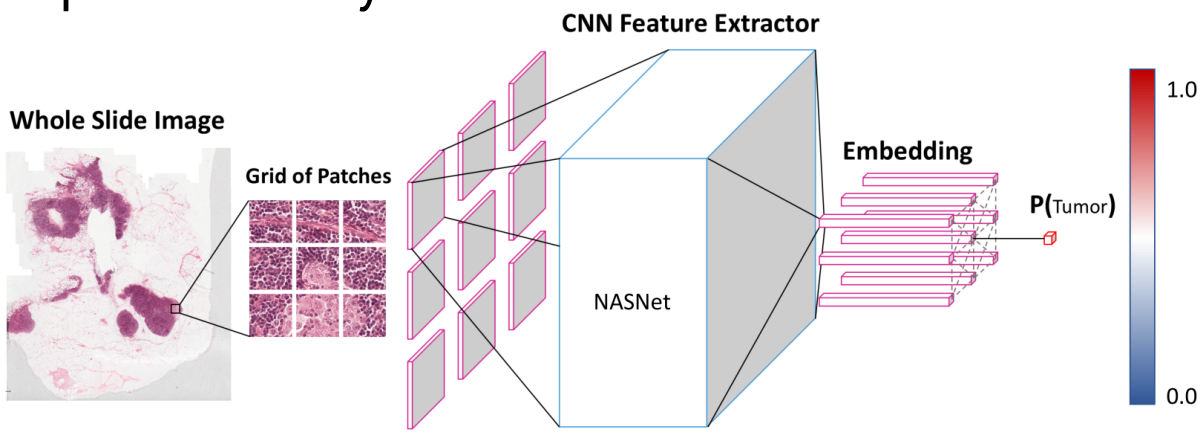


Machine Learning & Understanding

- Developing models that can process gigapixel images (e.g., 20 Gpix) efficiently
  - tile-based input into deep convolutional networks
  - experiments with different architectures

Pipeline

- Developed complete modular pipelines:
  - from reading/tiling big WSIs dynamically and reproducibly based on annotations to generate positive/negative tile sets
  - processing tiles – augmentations
  - training
  - evaluation
- Storing compressed history of each run
- Issues with reproducibility



https://github.com/baidu-research/NCRF

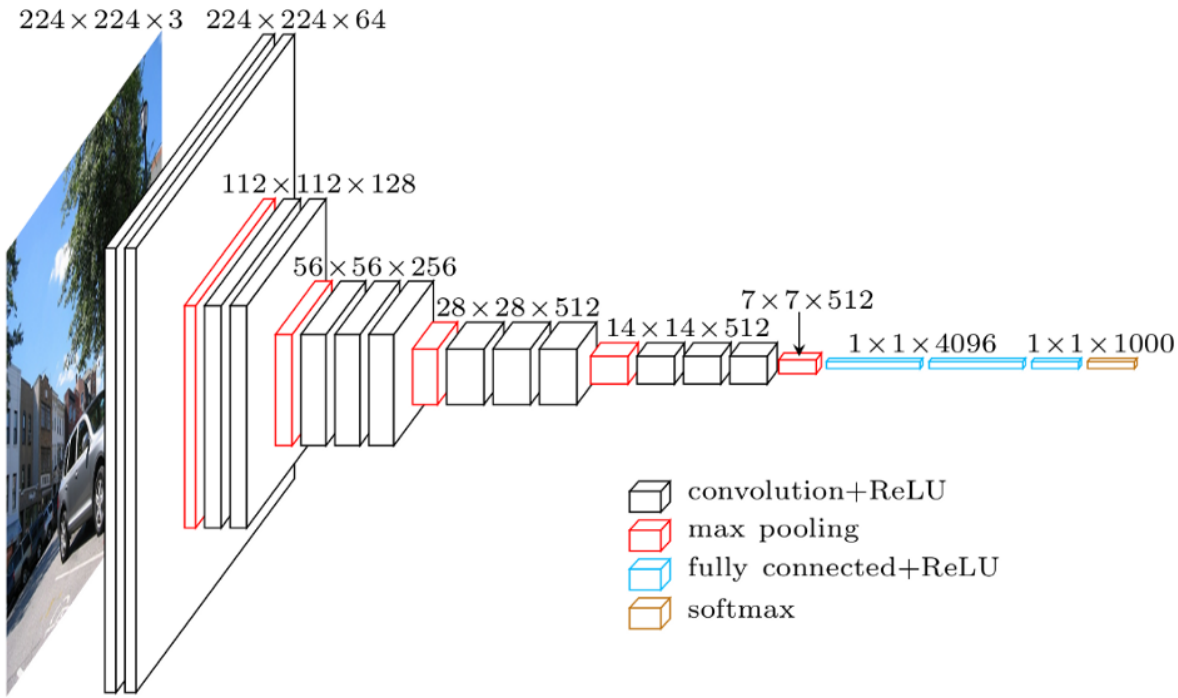
Training Data

- 698 slides; 156 patients
- The training set: 6,513,435 negative patches and 1,365,240 positive patches.
- augmentation techniques:
  - random vertical and horizontal flips with 50
  - random brightness perturbations in range [−64,64],
  - random hue perturbations in range [−10,10],
  - random saturation perturbations in range [−64,64] and
  - random contrast perturbations in range [0.5,2.0].
- All patches were scaled to [−1,1] range.
- Training data three step sampling:
  1. a label is selected uniformly at random
  2. single slide is picked uniformly at random from all the slides containing at least one patch with label
  3. a patch with label is selected uniformly at random from the slide

  A performance issue: A new slide (i.e. something large) is opened every time a new patch is sampled.

Model

- Convolutional Neural Networks (CNNs)
  - VGG16
  - combined with ImageNet initial weights
  - experimenting with different ways how to include bigger context – aggregating 3×3 vs. big tiles with evaluating centers only



Training & Testing

- RMSprop optimizer with the following parameters: momentum $= 0.9$, $\rho = 0.9$, initial learning rate of $\eta = 5 * 10^{-5}$
- Learning rate was halved after every 5 consecutive epochs without improvement on a validation data.
- The training would stop if no improvement on a validation data is made for 10 consecutive epochs.

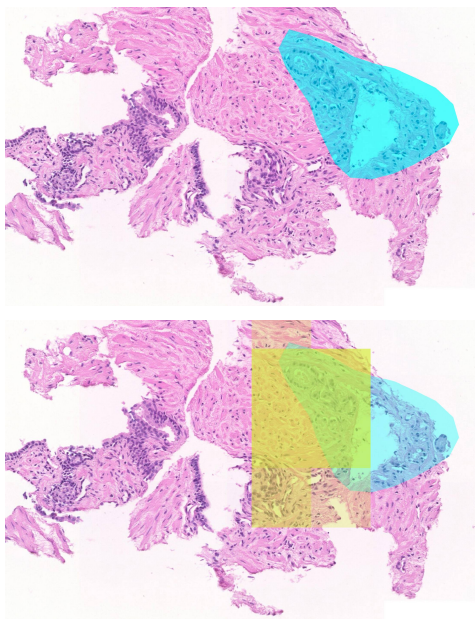
Testing:

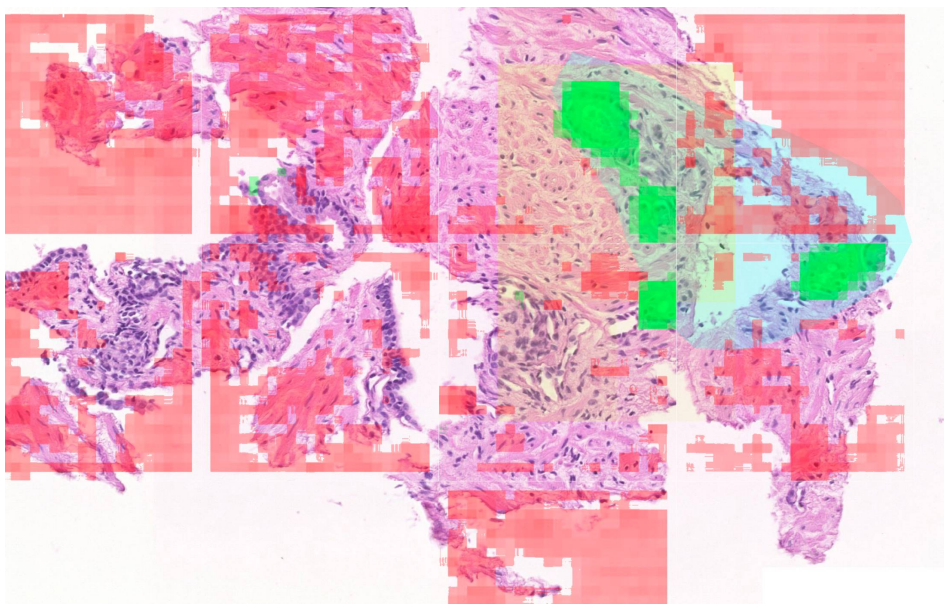
- Test set: 87 slides; 10 patients; 193,235 tiles

Machine Learning & Understanding

- Understanding how the deep neural networks work – explainability
  - to gain trust in the diagnostic process
  - to assess limitations of what the network can assess

Here gradient-based saliency maps





Interactive Visualizations

- Interaction with pathologists
  - visualizing results vs. original annoations
  - visualizing explainability information
  - annotating all these results
- Developed a visualization pipeline: from automated harvesting of trained models