Visualization of convolutional neural network class activations in automated oral cancer detection

Nadezhda Korigakina¹, Natása Sladoje¹, Ewert Bengtsson¹, Eva Darai Ramqvist², Jan-Michael Hirsch³, Christina Runow Stark⁴, Joakim Lindblad¹
¹Centre for Image Analysis, Department of Information Technology, Uppsala University, Sweden; ²Pathology and Cytology, Karolinska Institute, Stockholm, Sweden; ³Surgical Sciences, Oral & Maxillofacial Surgery, Uppsala University, Uppsala, Sweden; ⁴Public Dental Service, Södersjukhuset, Stockholm, Sweden

Motivation
Why do CNNs show such an impressive performance in automated image classification tasks?
- On the one hand, deep neural networks have a complex multi-layer structure that allows them to fit the data in a nonlinear way
- On the other hand, it is still very hard to conclude what makes them arrive at a particular decision
- Interpreting a neural network outcome is especially important for medical tasks, such as early cancer detection, where automated methods should assist cytologists in decision making
- How can we improve understanding and gain trust in CNN-supported decision making?

Workflow
- Brush samples collection
- Staining according to standard PAP procedures
- Acquisition of digital images of cells’ nuclei and surrounding tissue
- Train CNN model to categorise cells into malignant vs healthy type
- Convolutional neural networks (CNNs) have previously shown the ability to detect the difference between healthy and malignant samples [1]
- Ground truth labels are defined only at the patients’ level, not at the cellular level
- Not looking for clearly malignant cells, but using randomly selected cells in a sample

Explainable AI
- Recently, a variety of methods have been introduced to improve understanding of neural networks in different ways
- We focus on methods that visualize what aspects of input data affect the network’s decision [2-8]

Visual analysis of decisions made by networks
The same architecture is applied to two different datasets: cats & dogs and cells

Conclusions
- We have selected a set of most promising approaches for visualisation of CNN class activations
- We demonstrate applicability of these methods to cytological image data, however, a number of challenges remain:
  - The evaluation of visualisation by human is subjective; an objective measure of a map quality is not straightforward to design
  - Explanations tightly depend on the data; any feature of an input image is used if it helps the network to reduce the loss during training
  - There is no guarantee that human perception of a class would coincide with the neural network perception

References: