UPPSALA UNIVERSITET

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

Nadezhda Koriakina[1], Nataša Sladoje[1], Ewert Bengtsson[1], Eva Darai Ramqvist[2], Jan-Michaél Hirsch[3], Christina Runow Stark[4], Joakim Lindblad[1]

[1]Centre for Image Analysis, Department of Information Technology, Uppsala University, Sweden; [2]Pathology and Cytology, Karolinska Institute, Stockholm, Sweden; [3]Surgical Sciences, Oral & Maxillofacial Surgery, Uppsala University, Uppsala, Sweden; [4]Public Dental Service, Södersjukhuset, Stockholm, Sweden

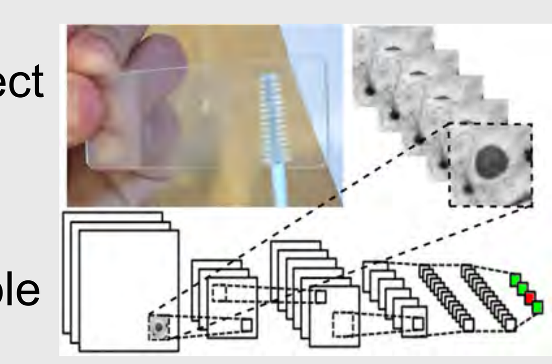
Context

- Cancer of the oral cavity is one of the most common malignancies in the world
- No routine screening tests for early detection yet
- Collection of samples with a brush would be a practical choice in this case
- Cytological examination supported by automated image classification

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



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?

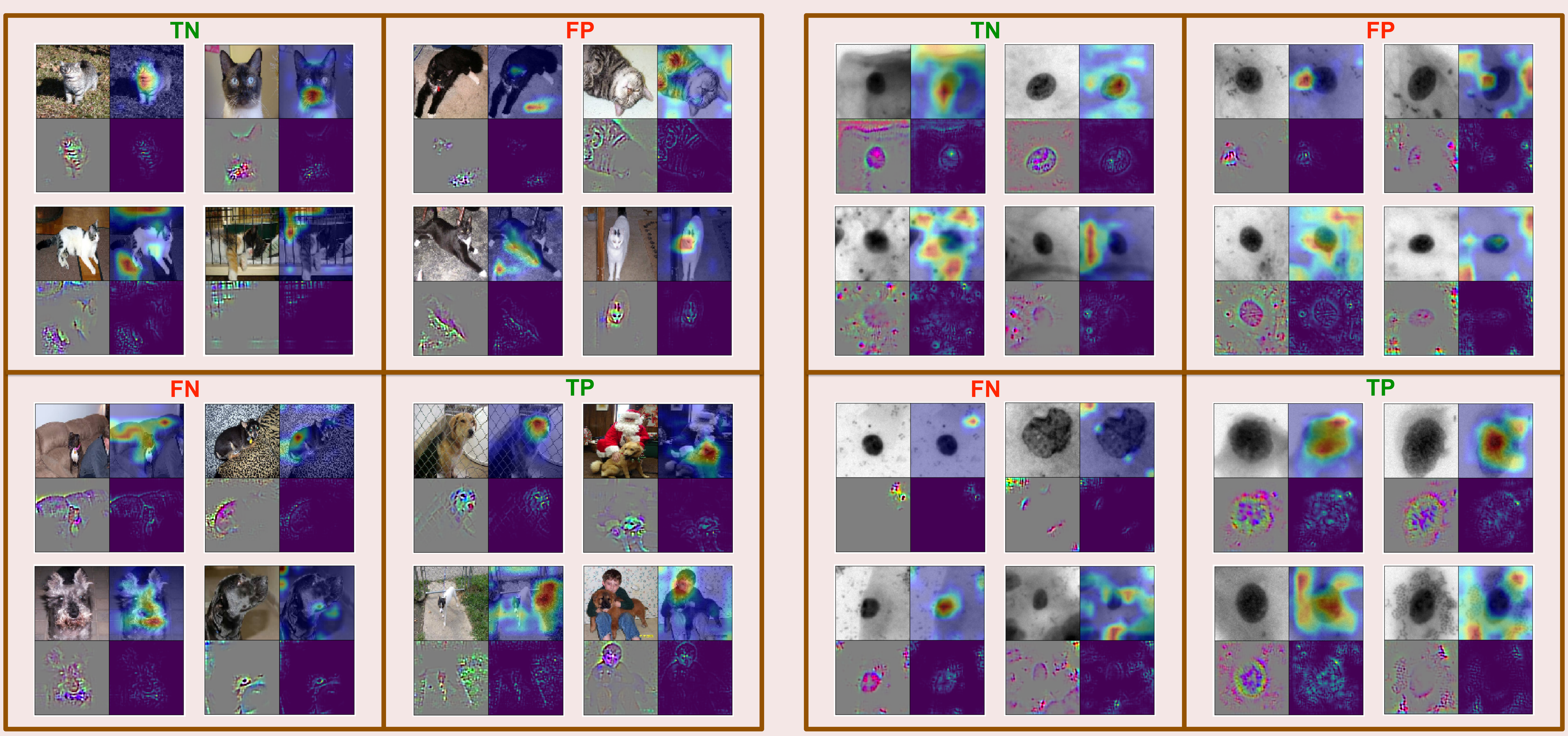
Explainable AI

- 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

Dogs (P) and cats (N) | Cells from malignant (P) and healthy (N) samples



Methods order in each image. Top left: original image, Top right: grad-CAM, Bottom left: guided grad-CAM, Bottom right: guided grad-CAM, positive saliency, sum along channels [3,5]

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

Centre for Image Analysis
Uppsala University

nadezhda.koriakina@it.uu.se

References:

[1] G. Forslid, H. Wieslander, E. Bengtsson, C. Wahlby, J. Hirsch, C. R. Stark, and S. K. Sadanandan. Deep convolutional neural networks for detecting cellular changes due to malignancy. In 2017 IEEE ICCVW, pages 82–89, Oct 2017.
[2] M. D. Zeiler, G. W. Taylor, and R. Fergus. Adaptive deconvolutional networks for mid and high level feature learning. In 2011 ICCV, pages 2018–2025, Nov 2011.
[3] Selvaraju, Ramprasaath R. et al. "Grad-CAM: Visual Explanations from Deep Networks via Gradient-Based Localization." *2017 IEEE International Conference on Computer Vision (ICCV)* (2017): 618-626.
[4] K. Simonyan, A. Vedaldi, and A. Zisserman. Deep inside convolutional networks: Visualising image classification models and saliency maps. CoRR, abs/1312.6034, 2013.
[5] J. T. Springenberg, A. Dosovitskiy, T. Brox, and M. A. Riedmiller. Striving for simplicity: The all convolutional net. CoRR, abs/1412.6806, 2014.
[6] M. Bojarski, A. Choromanska, K. Choromanski, B. Firner, L. D. Jackel, U. Muller, and K. Zieba. Visualbackprop: visualizing cnns for autonomous driving. CoRR, abs/1611.05418, 2016.
[7] S. Bach, A. Binder, G. Montavon, F. Klauschen, K.-R.Muller, and W. Samek. On pixel-wise explanations for non-linear classifier decisions by layer-wise relevance propagation. PLOS ONE, 10(7):1–46, 07 2015.
[8] B. Zhou, A. Khosla, A. Lapedriza, A. Oliva, and A. Torralba. Learning deep features for discriminative localization. CoRR, abs/1512.04150, 2015.