Visual interpretability for patch-based classification of breast cancer histopathology images

Mara Graziani, Vincent Andrearczyk, Henning Müller
University of Applied Sciences Western Switzerland (HES-SO), Sierre, Switzerland
University of Geneva (UNIGE), Geneva, Switzerland
mara.graziani@hevs.ch

Abstract

Decision support for digital histopathology has increased in an important way thanks to very good results using deep learning techniques in the past few years but neural networks have largely remained black boxes. Visualization methods address the interpretability of neural networks. If applied to histopathology images, they can improve the trust of pathologists in automated support tools for cancer diagnosis. We perform model interpretation and decision explanation of our binary patch-based breast cancer classifier. Experimental results show that morphological features of the nuclei influence network decisions in an important way.

1 Introduction

Visualizing deep Convolutional Neural Network (CNN) decisions for the determination of cancer stages can enhance the trust of physicians in automated support tools for cancer diagnosis. The visualization of input regions meaningful for classification can be obtained through backpropagation of the network response [2]. For instance, Gradient-weighted Class Activation Mapping (gradCAM) [3] produces a localization map using the gradients of any target class with respect to the last convolutional layer. Activation Maximization (AM), first proposed in [1], visualizes patterns sought by neurons by finding the input that maximizes the neural response. In medical imaging, MDNet was proposed in [5] to obtain explanations on the diagnosis process of the network by a language model with visual attention. In this paper, we propose preliminary visualizations of our patch-based tumor classifier for breast cancer histopathology. Particularly, we interpret deep layers with AM and we explain model decisions on testing data with gradCAM.

2 Experiments and Results

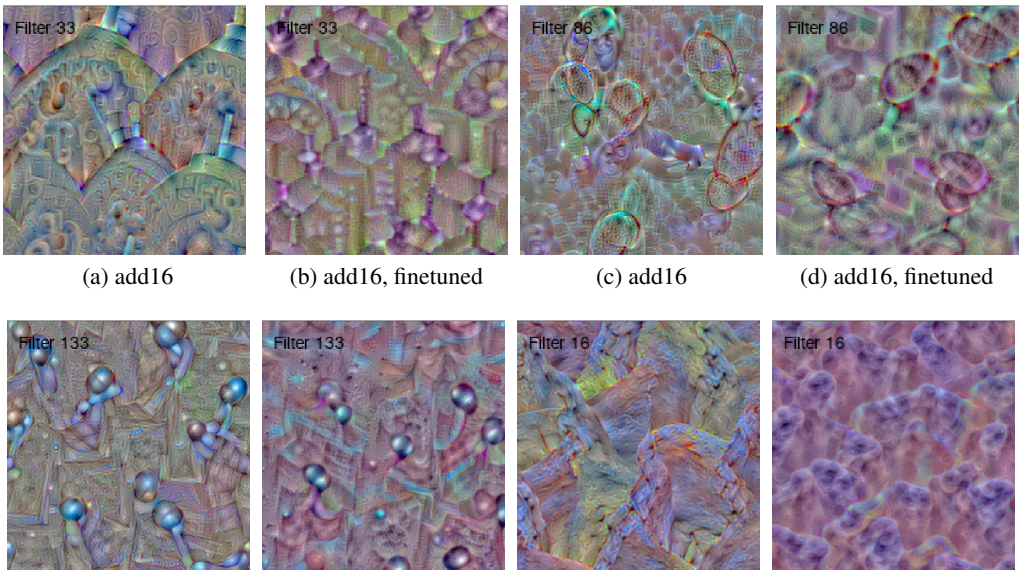
The grand-challenge dataset Camelyon17[1] (for classification of breast cancer metastases in lymph node sections) was used for the experiments. The decision layer of ResNet50 [4] pretrained on ImageNet was replaced with a single decision node for binary classification of tumor and non-tumor patches. The network was finetuned for 10 epochs with stochastic gradient descent and binary cross-entropy loss. For this preliminary investigation, we focused on achieving a model that performs sufficiently well to obtain meaningful visualizations and we did not target optimal performance. We sampled 10,390 patches of size 224×224 with uniform probability from non-tumor and tumor annotated tissues of the Whole Slide Images (WSIs) at resolution level 1. The need for data augmentation was reduced by using WSIs from only three of the five acquisition centers. Model performance was validated on 6,483 validation patches, obtaining an AUC of 0.965. In the following section, we show some of the visualization results on which we based our observations.

[1] https://camelyon17.grand-challenge.org

1st Conference on Medical Imaging with Deep Learning (MIDL 2018), Amsterdam, The Netherland.

2.1 Model interpretation

We visualize AM inputs that maximize the last and penultimate convolutional feature maps of ResNet50 before and after finetuning in Figure 1. Complex high-level structures in the pretrained model such as buildings or mountain patterns as in Figure 1a are finetuned into textures of repeated nuclei-resembling shapes. A shift of color towards WSI color ranges is noticeable. Round shapes like the ones shown in Figure 1e are kept, whereas pointed shapes and straight edges tend to be smoothed. To understand the model discrimination between tumor and non-tumor, we substituted the single

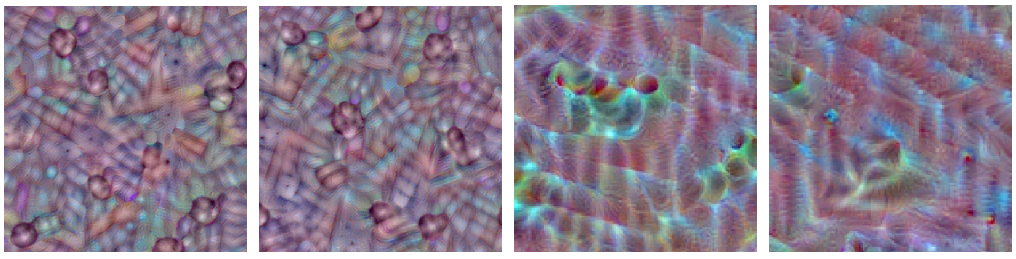


(a) add16 (b) add16, finetuned (c) add16 (d) add16, finetuned

(e) add16 (f) add16, finetuned (g) add15 (h) add15, finetuned

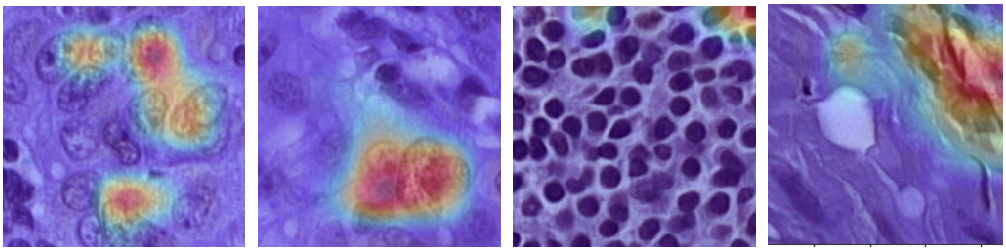
Figure 1: AM of the last (add16) and penultimate (add15) layers of the pretrained Resnet50 and our tumor classifier after finetuning.

decision node in the prediction layer with two nodes. Input patches that maximally activate either the tumor class node or the non-tumor class node can be obtained through AM. Figure 2 shows the AM of the two prediction nodes for different random seeds. Both the tumor and the non-tumor AM contain nuclei-resembling shapes, although a consistent difference in both classes can be noticed. The tumor AM seems to contain numerous round shapes which are markedly defined and spread in the patch. However, more research is needed to clearly interpret the discriminant factors between the classes.



(a) tumor AM (b) tumor AM (c) non-tumor AM (d) non-tumor AM

Figure 2: AM for tumor and non-tumor tissue. Input patches maximizing the tumor class node activation are shown in (a) and (b). AM for the non-tumor node is shown in (c) and (d).



(a) tumor, $p = 0.994$ (b) tumor, $p = 0.999$ (c) non-tumor,$p = 4.7e-4$(d) non-tumor, $p = 0.841$

Figure 3: GradCAM for testing patches. We report the ground truth class and the probability for a tumor as p.

2.2 Decisions explanation

GradCAM, sensitivity methods and saliency [2] are powerful tools to identify visual features in the input that explain the classification. Figure 3 shows gradCAM on testing patches. Nuclei pleomorphism seems to be the main factor affecting the output tumor prediction. Heatmaps of tumor patches mostly focus on nuclei with marked variations in size and irregular shapes. Similarly, larger cells with visible nucleoli or stippled nuclear chromatin explain the tumor prediction in Figure 3a. Non-tumor nuclei do not have such characteristics, hence non-tumor patches present low activations. Figure 3d shows the gradCAM for a false positive prediction. The deformation of the nuclei increases the probability of the classifier for tumor, although the patch belongs to non-tumor tissue.

3 Discussion and Future Work

This paper presents preliminary results for improving the interpretability of patch-based tumor classifiers. The experiments show that nuclei pleomorphism, especially size and color intensity, are the main factors affecting the probability of the class tumor. Irregularities of the non-tumorous tissue, which may be given to artifacts in the acquisition procedure, and the consequent variation of nuclei shapes, may be one of the causes of false positives. Future work is needed to investigate misclassification errors and build more robust classifiers with better generalization. Moreover, the analysis should be extended to different resolution levels to identify differences in the discriminating features and to visualize classifier decisions, also as a means to remove the black box image of deep learning in histopathology.

References

[1] Dumitru Erhan, Yoshua Bengio, Aaron Courville, and Pascal Vincent. Visualizing higher-layer features of a deep network. *Bernoulli*, (1341):1–13, 2009.

[2] Grégoire Montavon, Wojciech Samek, and Klaus Robert Müller. Methods for interpreting and understanding deep neural networks. *Digital Signal Processing: A Review Journal*, 73:1–15, 2018.

[3] Ramprasaath R. Selvaraju, Michael Cogswell, Abhishek Das, Ramakrishna Vedantam, Devi Parikh, and Dhruv Batra. Grad-CAM: Visual Explanations from Deep Networks via Gradient-Based Localization. *Proceedings of the IEEE International Conference on Computer Vision*, 2017-Octob:618–626, 2017.

[4] Ren Wu, Shengen Yan, Yi Shan, Qingqing Dang, and Gang Sun. Deep Image: Scaling up Image Recognition. 2015.

[5] Zizhao Zhang, Yuanpu Xie, Fuyong Xing, Mason McGough, and Lin Yang. MDNet: A Semantically and Visually Interpretable Medical Image Diagnosis Network. 2017.