CNNs projects - #4

Whole slides attention maps software

Project outline

The project is focused on the digital pathology field. This sphere of research points at developing tools usable by pathologists. A pathologist is a doctor who performs diagnosis based on the visualization of histological samples (Figure 1 on the right). An histological sample is a piece of human tissue, obtained via a surgical operation. For example in Figure 1 on the left is shown a removal of a lesion in the colon, during a colonscopy. The pathologist traditionally performs a visual examination on such sample, looking for abnormalities, and decide for a diagnosis. A WSI (whole histological image) is a multi-resolution image (Figure 2) obtained from such sample, which may be elaborated numerically.

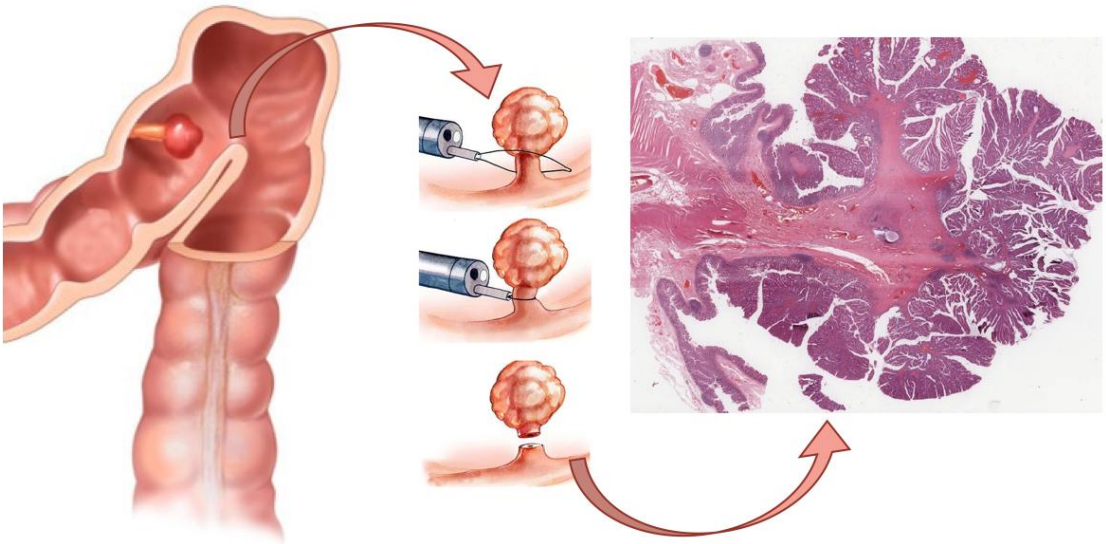


Figure 1

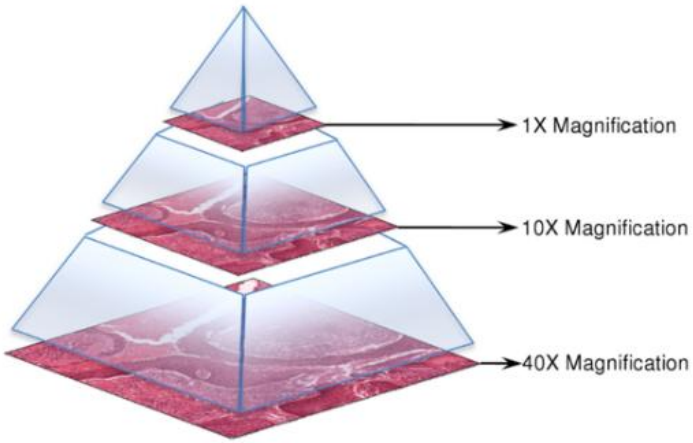


Figure 2

Project focus

The aim is to develop a software able to produce a so-called attention map for cancer detection. An example of such map is depicted in the below Figure 3. This visualization may serve to drive the attention of the pathologist during the examination of a slide. In Figure 3 the red part corresponds to regions which are likely cancer. The uncertainty of the map will be modeled with Bayesian CNNs (BNNs). Going into further details, instead of point estimate of weights, a BNN approximates the distribution of weights usually with a Gaussian distribution with two hyper-parameters (mean and standard deviation), based on prior information and data. The prediction uses the posterior distribution of weights, computed following the Bayes theorem. The back-propagation algorithm is used to update the hyper-parameters of the weights (i.e. mean and standard deviation: two numbers for each weight). This way, the model can provide uncertainty estimates of the weights and predictions, at the price of training double parameters with respect to normal ConvNets.

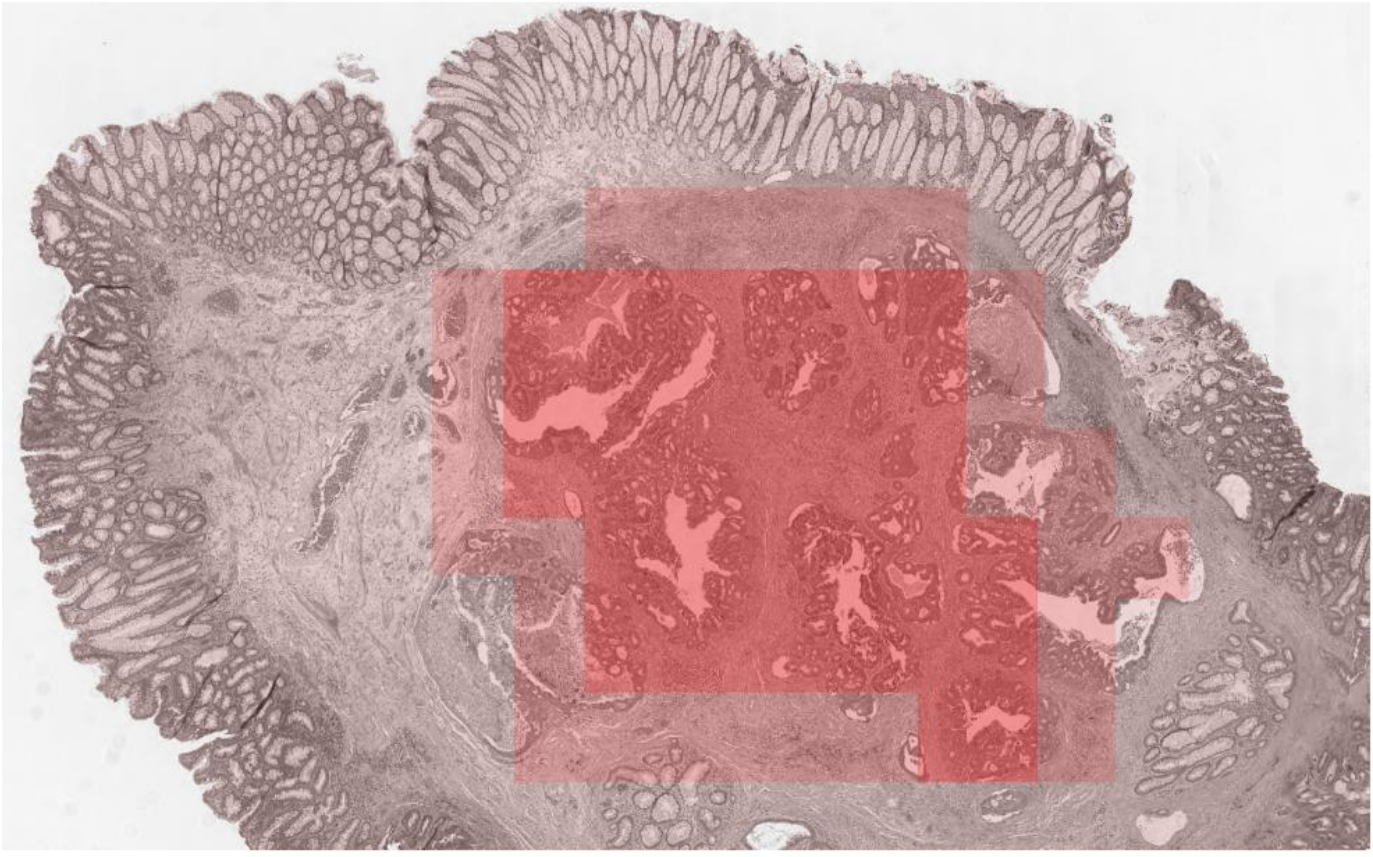


Figure 3

Assignments:

1) Data management

In the Google Drive folder attached in the mail, you can find a certain number of WSIs belonging to different patients and classes of interest. Such file may be viewed by the software available here: https://www.leicabiosystems.com/digital-pathology/manage/aperio-imagescope/

To move files on the Politecnico cluster, please follow the guide available at http://hpc.polito.it/

In order to feed a CNN with this kind of images, is necessary to crop them in a very high number of patches, as shown in Figure 4.

To automatize such process the following library is suggested https://openslide.org/api/python/, which offers a python wrapper.

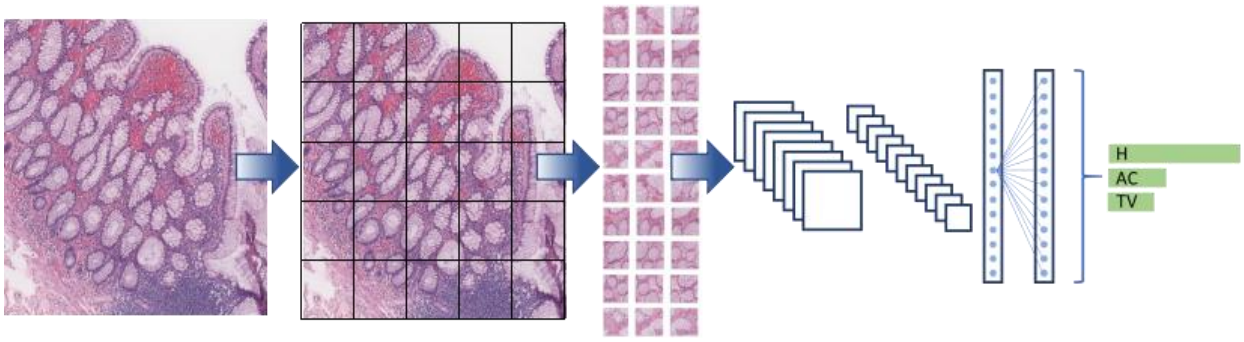


Figure 4

The dimension of crops and their overlap are parameters of your pipeline which must be analyzed, optimized and discussed. During dataset preparation, pay attention in avoiding class imbalance and eventually consider data augmentation or down-sampling of some class.

2) Network training

Once your dataset is ready, you can train your network. You will train a BNN using **tensorflow-probability**. Pay attention in keeping different patients between train and test dataset.

3) Attention map (Healthy versus all-the-rest)

You will build the attention maps using the whole images in MAP sub-folder. After processing each image crop, the whole slide must be reconstructed. The colormap assigned to the image will be proportional to the probability of the classification in cancer class. The whole process is summarized in Figure 5. Keep in consideration the information related to uncertainty.

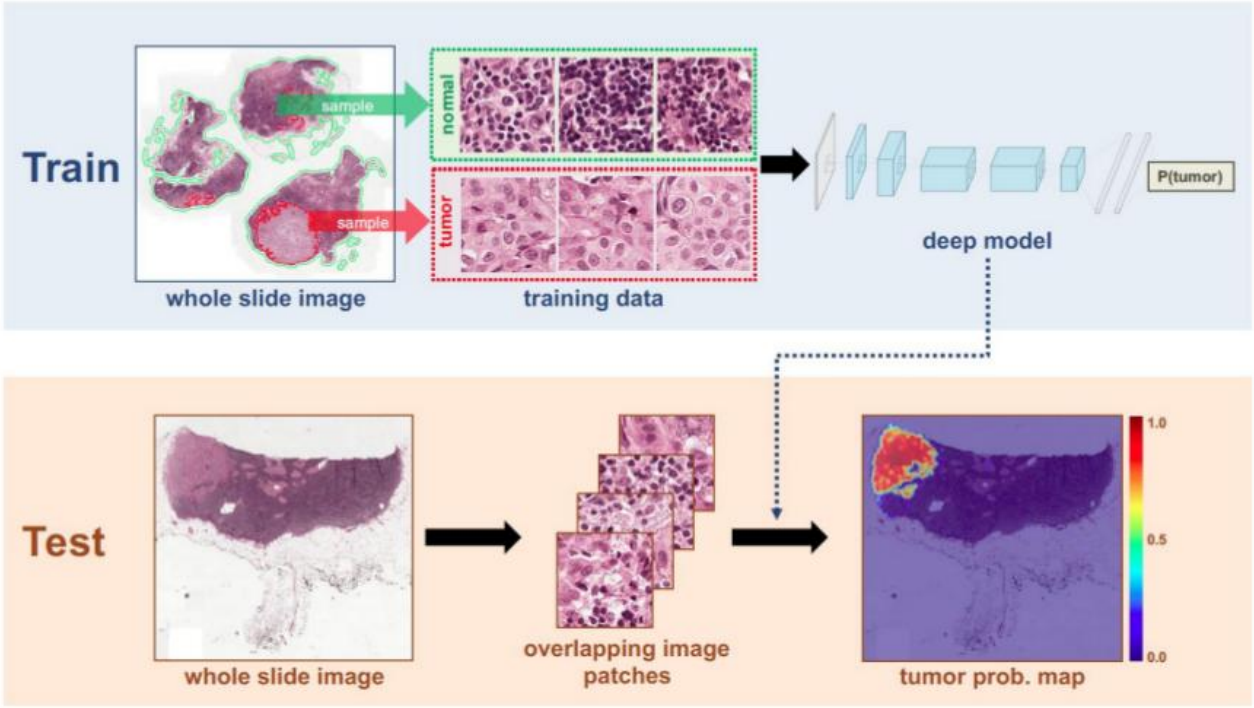


Figure 5

4) Software development

The design and development of a visualization software implementing the case Healthy versus all-the-rest is required. *Such software must be able to manage multi-reolution WSIs* overlapped by a mask proportional to cancer probability.

Classes of interest

Below (Figure 6) you can find some examples of cropped regions for the three classes of interest. The H class is characterized by the glands pattern (Figure 6 (a)) , which is lost in the other two classes. If some image crops depict a region which is not of interest (for instance a total white crop), they must be removed or not considered in training.

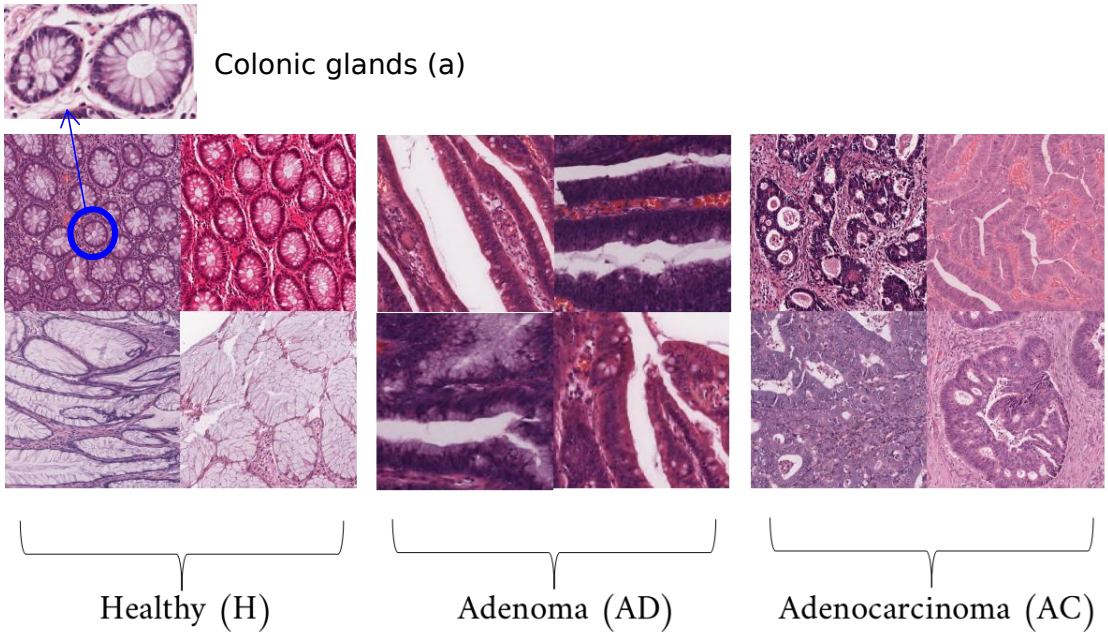


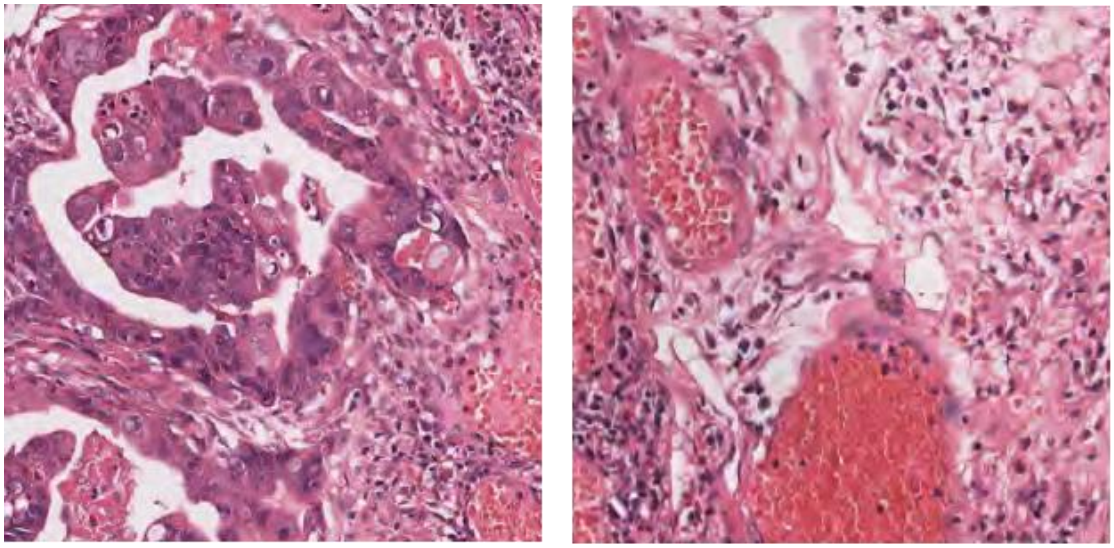
Figure 6

Dataset notation

{Patient-ID}_{Class-ID}_{Roi-ID}.svs

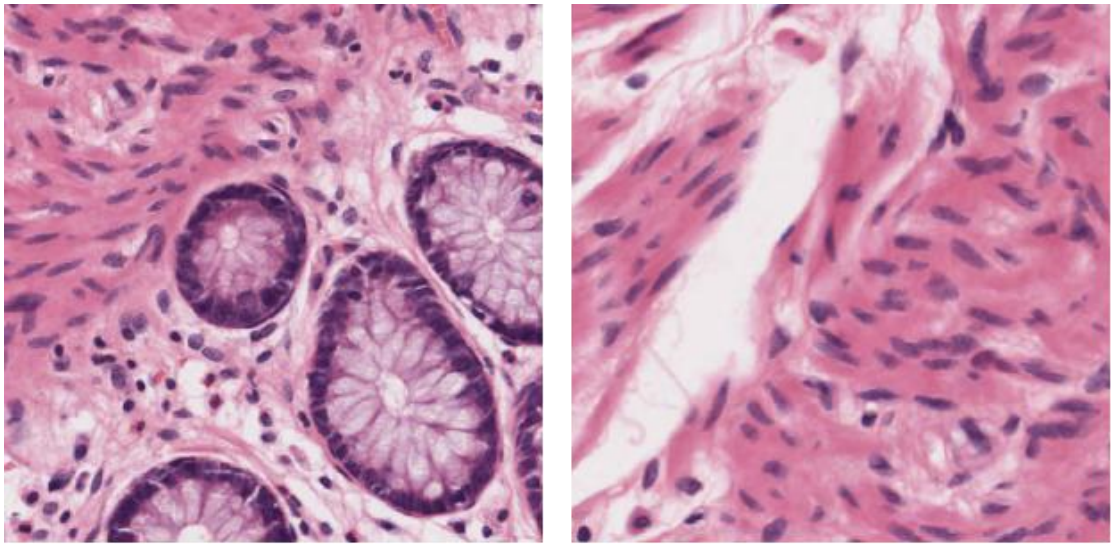
Patient-ID	patient number. Use different patients for train and test.
Class-ID	class of interest (AC, AD, H)
Roi-ID	each patients present some Region Of Interest (ROI), which are WSIs focused on a certain type of tissue. After cropping, you **can** remove those patches not belonging to the specified class

In Figure 7 (a) is shown a correct patch for AC class. You can see a gland structure (adulterated by cancer) among an other kind of tissue: the stroma (i.e connective tissue). On the other hand, in Figure 7 (b), is reported a sample containing only stroma. This is not a correct instance for class AC. Another example in Figure 8: (a) is a correct sample for H (glands immersed into stroma); (b) is a misleading sample depicting only stroma. Consider that in machine learning, the quality of your output will never be higher than the quality of your input. On the other hand, the removal of such not correct patches may be boring and time-consuming. Such removal is NOT MANDATORY, but optional.



(a) (b)

Figure 7



(a) (b)

Figure 8

References

CNNs theory

http://cs231n.github.io/, https://medium.com/@laumannfelix (BNNs)

Useful libraries

openslide (WSIs management), opencv (image elaboration and colormaps), tensorflow-probability, tensorflow.