CNNs projects - #3 Growing DeepTree

The aim of the current project is two-folds: speaking about application, it 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, 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 decides for a diagnosis. A WSI (whole histological image) is a multi-resolution image (Figure 2) obtained from such sample, which may be elaborated numerically.

As regarding computer science sphere of research, the project is focused on the so-called *incremental learning*. In last years, Convolutional Neural Networks (CNNs) have shown great performance in many computer vision tasks. The network learns to extract features and classify images. This trained model is then applied to other unlabeled images to classify them. In such training, all the training data is presented to the network during the same training process. However, in real world, we hardly have all the information at once. Instead, data is gathered incrementally over time. With current model, if you want to add one class to the training set, you must re-train the model with the loss, at least partial, of the previous knowledge. This phenomena is known as the catastrophic forgetting. Incremental learning research investigates solution in this context.

Disclaimer:

This project is centered in a quite unexplored field of research, hence consider two aspects:

i. Results may be not good. Don’t worry too much for this: even bad results are results; we evaluate the development of the project.

ii. Data are strictly personal, don’t upload them nowhere and don’t share them.

Figure 1

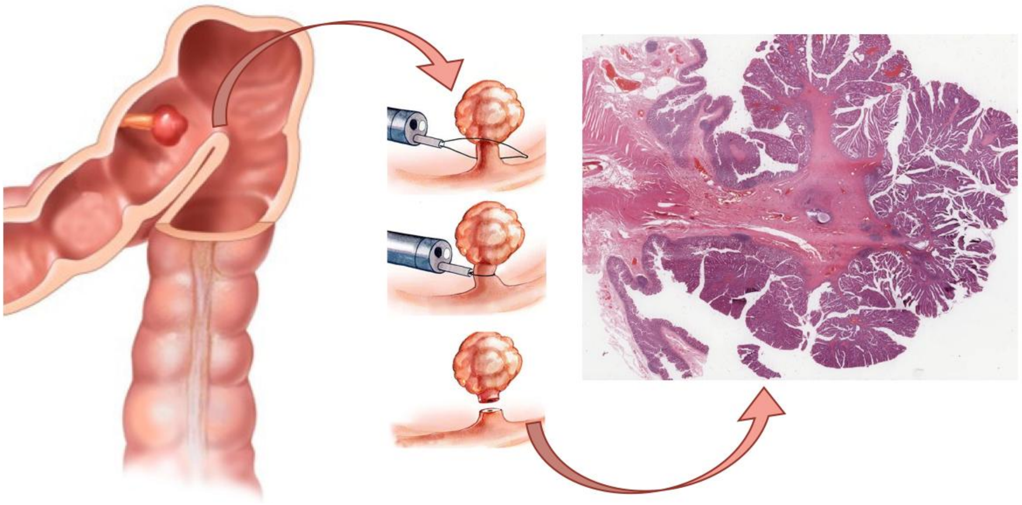
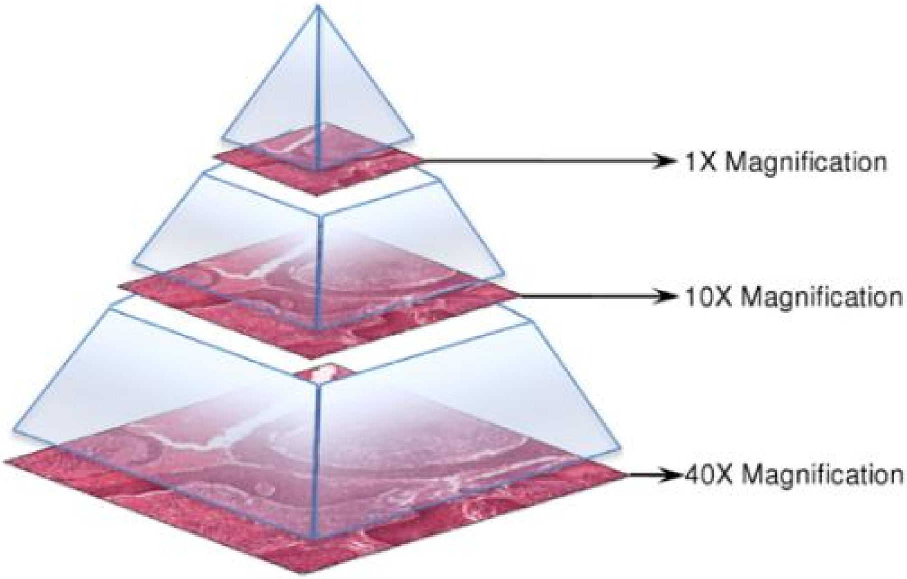


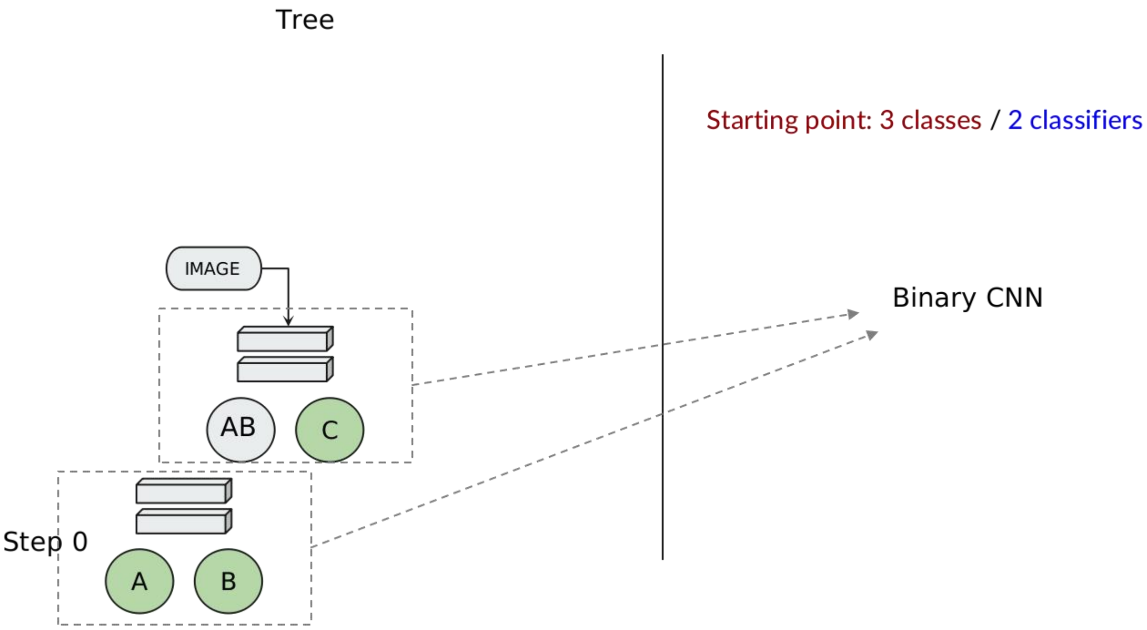
Figure 2

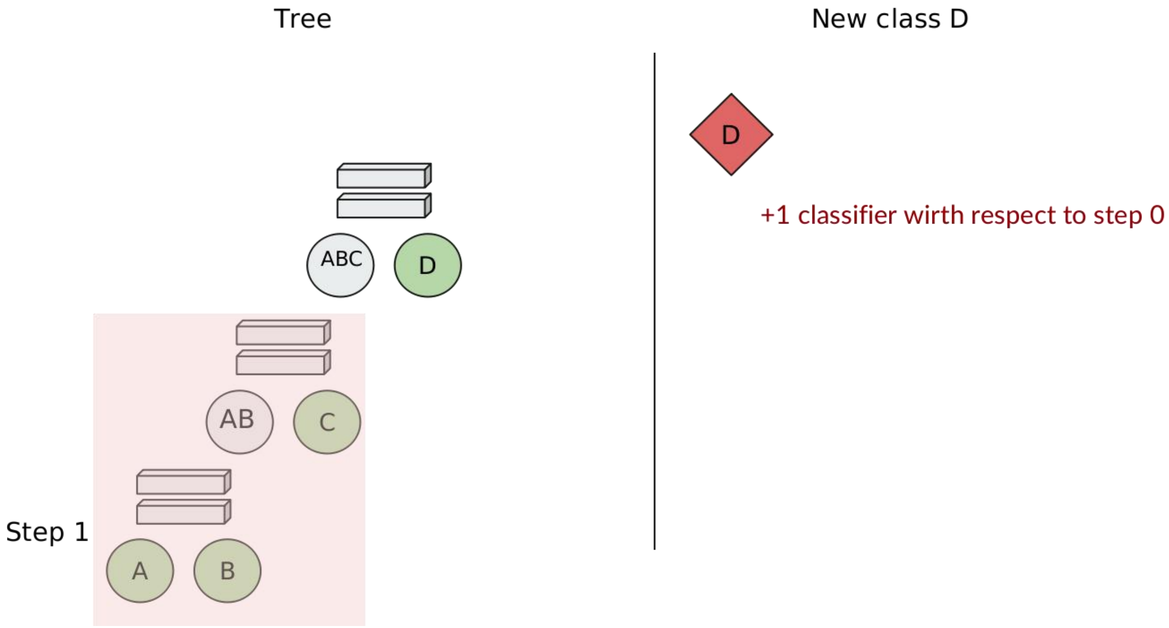


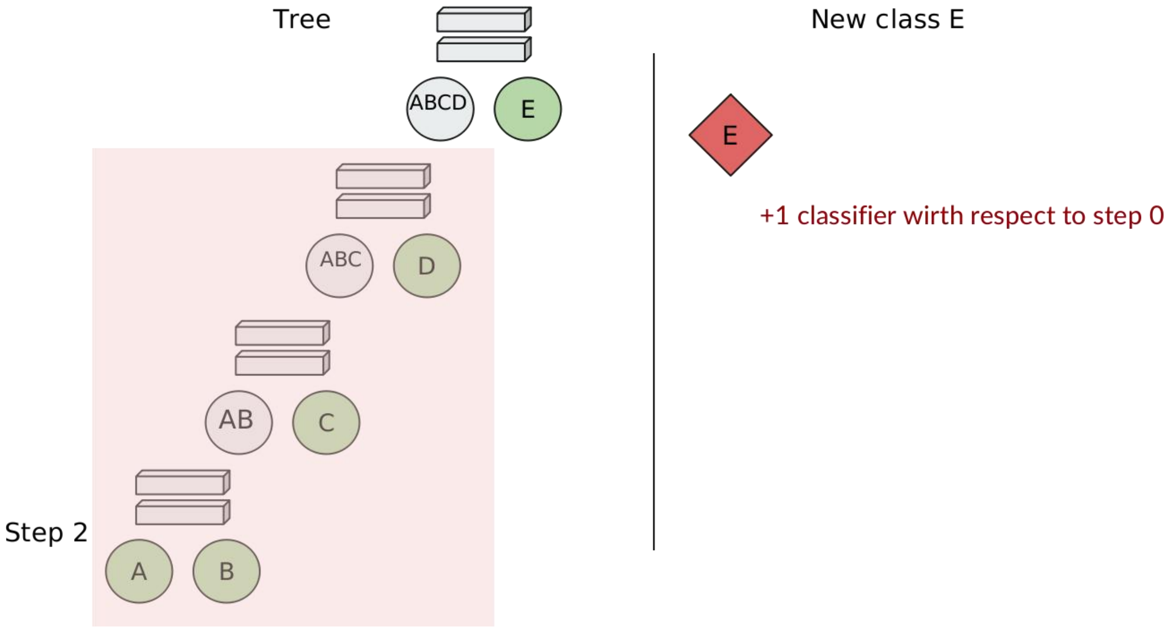
Main steps of the project:

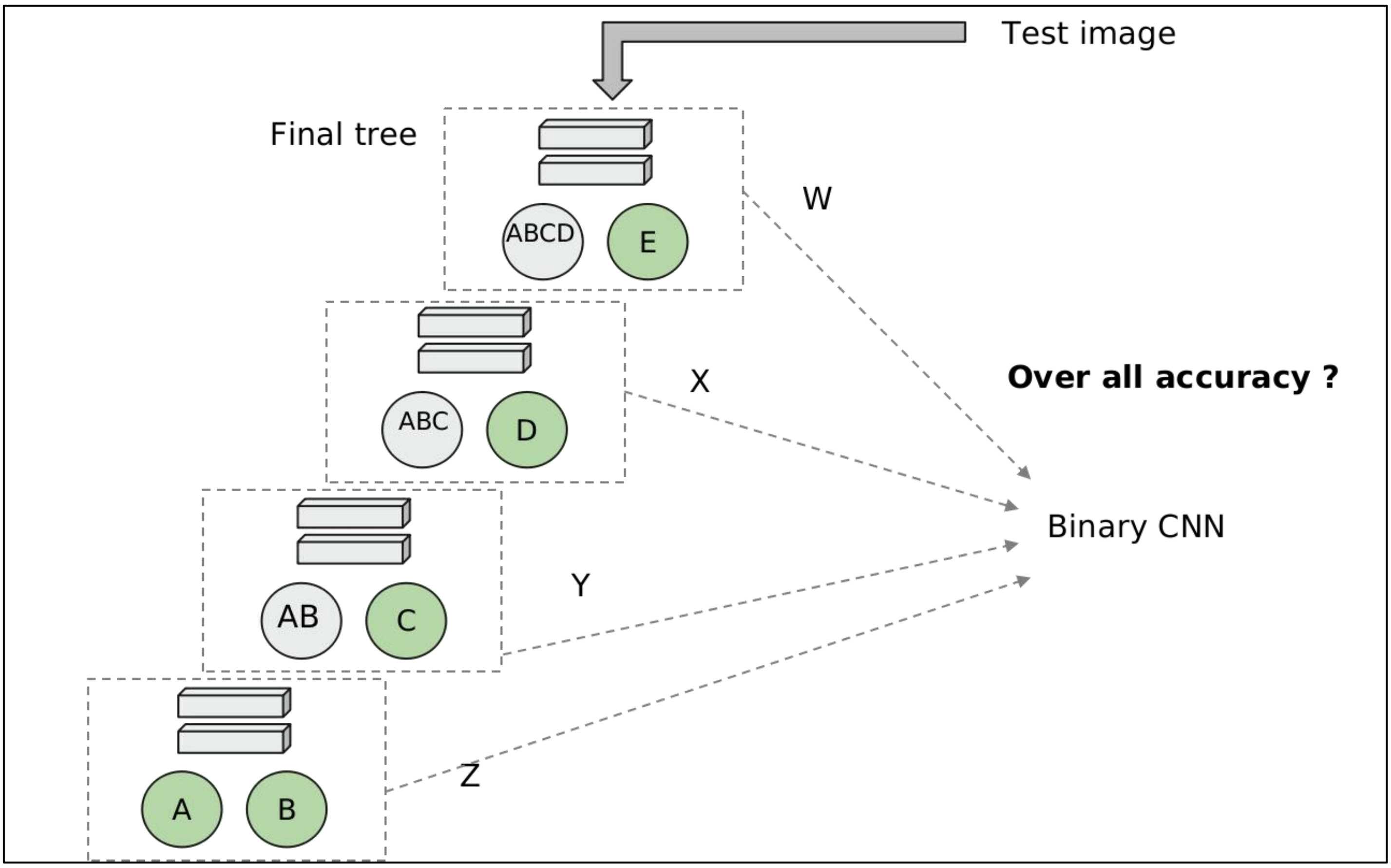
The assignment of the project is the design and development a **growing DeepTree, which is able to classify images in a context of incremental learning**. Here further details.

1. In the below Figure there is a visual explanation of the growing process when a new class arrives. A, B, C, D, E are *dummies* classes, read the following for real data details.






Tree
ABCD
E
New class E
E
+1 classifier wirth respect to step 0
ABC
D
AB
C
Step 2
A
B


Test image
Final tree
ABCD
E
W
ABC
D
X
Over all accuracy ?
Binary CNN
AB
C
Y
A
B
Z

2. As it can be gathered from above figure, there are, **at last step**, **4 binary CNN classifiers** in the tree (namely Z, Y, X, W). Z classifies A versus B. Y classifies A+B versus C, X classifies A+B+C versus D and so on. A+B is a super-class, union of classes A and B.

3. Pay attention to **keep the dataset balanced while training**. For example, consider classifier Y. Here, in A+B super-class, A must have the same number of training examples as B, and A+B must have the same number of training example as C.

4. In **test phase**, the testing image, which is unknown, will enter from the top, i.e. from W classifier, and will face n steps of classification **until it will reach a *leaf***, which means a class. Analyze the accuracy during each step of the growing and for the last final configuration.

5. You will face **two classification contexts**:

   a) The well-known Cifar10 dataset(look on the web for details). Select 5 classes among the 10 of the dataset.

   b) The biological dataset described in the following.

Biological Dataset

The biological dataset comes from Colon tissue of different patients and presents 5 classes of interest:

i. Adenocarcinoma, tumour (AC).

ii. Healthy tissue, normal colonic glands (H). See (a) in below figure.

iii. Serrated Adenoma, precursive lesion of cancer that may turn into it (Serr).

iv. Tubular adenoma, precursive lesion of cancer that may turn into it (T).

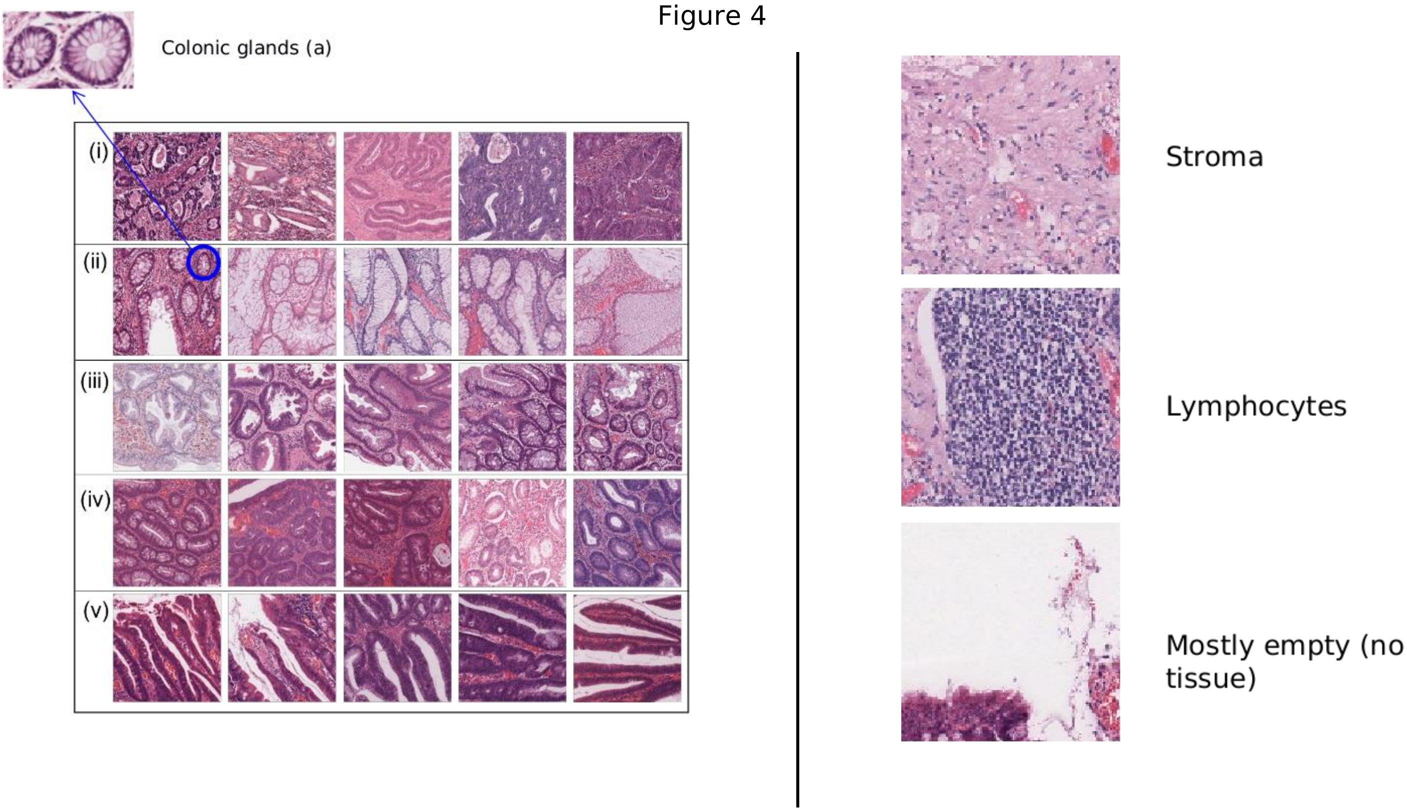
v. Villous Adenoma, precursive lesion of cancer that may turn into it (V).

The dataset presents some noise (below figure on the right), which can maybe be associated with high-uncertainty predictions.

Notation:

{Patient_ID}_{Class_ID}_{Image_ID}.png

Figure 4



Material

1) Background for CNNs and machine learning http://cs231n.github.io/

2) BCNNs https://medium.com/@laumannfelix and many papers on the argument (search online with PoliTo connection to download them)

3) Tensorflow https://jacobbuckman.com/post/tensorflow-the-confusing-parts-1/

4) Incremental learning https://github.com/xialeiliu/Awesome-Incremental-Learning