

Camelyon 2017: A two-step CNN classification workflow

VISILAB Group¹

<http://visilab.etsii.uclm.es/>

¹ University of Castilla La Mancha, Ciudad Real, Spain,

Author for correspondence: grupo.visilab@uclm.es

ABSTRACT

In this paper, the approach followed by our team to face the Camelyon 2017 pathology challenge is explained. The proposed method relies on a two-step process. The first one employs a dataset obtained from the XML information about the provided WSIs. This dataset is used to train a first Convolutional Neural Network (CNN) model that performs a high-level classification between benign and malignant regions (macro-metastases). The next one makes use of a handcrafted dataset that represents the variability of this problem more precisely, training a second CNN model that is available to classify lower-level cases (micro-metastases and isolated tumour cells (ITC)). Using these results, a patient level diagnosis is obtained, according to the challenge rules.

Index Terms— Breast cancer, WSI, Digital pathology, Convolutional neural networks

1. INTRODUCTION

Whole-slide images (WSI) processing is slowly becoming more relevant for clinical trials, to ensure proper classification of cases and lower diagnosis discrepancy rates. The challenges in digital pathology have led to improvement in image analysis techniques resulting in better opportunities offering to the pathologist for treatment of benign tissues. There are multitude of histopathological processes which are still made by hand and a digital workflow is needed.

One of the most demanding diagnosis is the cancer detection, especially, breast cancer detection. The goal of this challenge 1 was to evaluate new and existing algorithms for automated detection of metastases in hematoxylin and eosin (H&E) stained whole-slide images of lymph node sections. To this end the CNN approach has been explored in this work.

2. MATERIALS AND METHOD

This section explains the datasets that have been developed, the processing techniques that have been applied, the developed classification workflow and how the final diagnosis is stated.

2.1. Dataset 1

The first step that has been taken is to develop a dataset that contains significant patches from benign and malignant regions, using the information about the WSIs (Whole Slide Images) provided by the challenge organizers in XML format. Thus, this dataset has two classes: Benign and Malignant.

The benign samples are obtained from negative patient node slides. These regions are cropped into 91 x 91 pixel patches.

The malignant samples are obtained from bounding boxes of the regions provided by the XML. These are later divided into 91 x 91 pixel patches, since this size is usually enough to isolate single tumour cells. To increase the size of this dataset, a data augmentation process has been applied, performing rotations of 90, 180 and 270 degrees.

In these patches, the differences between hospitals are clearly appreciable, as it is observed in Figure 1. Because of that, a colour standardization is applied, in which a reference of every hospital is selected, applying it to the rest of them.

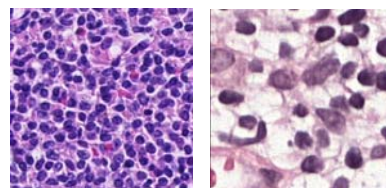


Figure 1.- Tissue samples from different hospitals

As a result, this dataset has 623,000 samples.

2.2. CNN model 1

Using the previous dataset, a Convolutional Neural Network (CNN) is trained. The selected network is the state-of-art AlexNet [1], using a fine-tuning approach. This way, a pretrained version of the model is used to slightly modify the weights in order to adapt them to this problem. The architecture of this network is shown in Figure 2.

The hyperparameters of the training process are the following: the learning rate is started out with an initial of 0.0001. For back propagation, the Stochastic Gradient Descent was used, establishing 0.0004 for L2-Regularization. With the previous parameters, the training reaches the best results with just 5 epochs (after that, the loss value and accuracy do not improve). It takes around 5 hours to perform the training process with this configuration, using the GPU NVIDIA Quadro M4000 with 8 GB of VRAM each.

2.3. Dataset 2

The first model was available to discern generally between benign and malignant samples, but the complexity of this problem makes it necessary to represent more extensively the variability of structures that compose these classes, reducing false positives, and increasing the accuracy. For this reason, a second dataset is built. The idea is to split the previous classes into more specific and representative cases, pointing out the different structures that belong to benign and malignant samples more precisely.

This dataset has been handcrafted using expert knowledge about the problem, obtaining 5 different classes. Three of them belong to benign samples and the other two to malignant ones. They represent different cellular density. Figure 3 show samples of each class.

The dataset is built using 181 x 181 pixel patches.

As a result, this dataset has 233,000 samples.

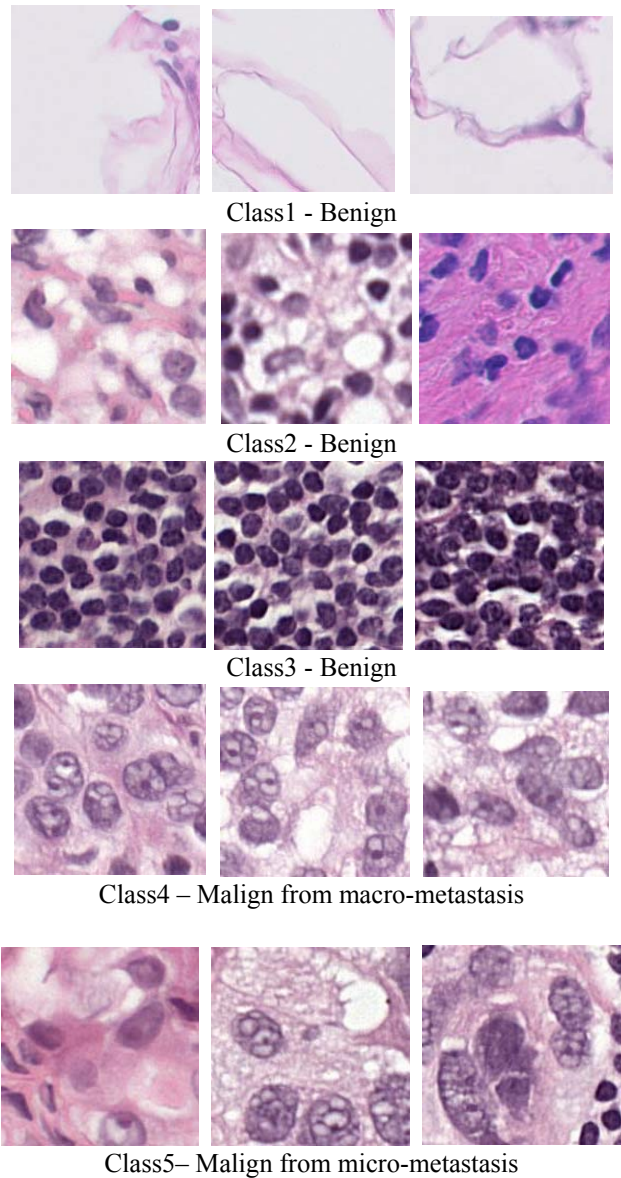


Figure 3.- Patch samples from the different selected classes

CNN model 2

Using the dataset 2, a new Convolutional Neural Network (CNN) is trained. The selected network is, again, the state-of-art AlexNet [1], using a fine-tuning approach. This way, a pretrained version of the model is used to slightly modify the weights to adapt them to this problem. The architecture of this network is shown in Figure 2.

The hyperparameters of the training process are slightly different from the previous model, due to the different number of classes and size for the input images. The learning

rate is started out with an initial value of 0.001, decreasing it with drop factor of 0.001 with a period of 6. For back propagation, the Stochastic Gradient Descent was used, establishing 0.004 for L2-Regularization. With the previous parameters, the training reaches the best results with 10 epochs (after that, the loss value and accuracy do not improve). It takes around 6 hours to perform the training process with this configuration, using the GPU NVIDIA Quadro M4000 with 8 GB of VRAM each.

2.4. WSI preprocessing

In order to process only the significant regions of the WSIs, binary masks are developed to determine where the tissue is placed.

This is performed using a RGB colour thresholding at 10x. This threshold is useful to build the mask, in which the tissue is represented as a Boolean value of one, leaving the zero value for non-tissue regions.

Then, a second processing is performed, employing basic mathematical morphology processing based on erosion, and dilation with a disk of 8 pixels. Moreover, regions with an area smaller than 1500 pixels are deleted. After this process, the masks of every tissue from every slide are obtained.

2.5. Classification workflow

The classification of a new WSI takes the preprocessing mask as input. Each tissue region extracted from the information of the mask is divided into 91 x 91 pixel patches. These patches are fed into the developed CNN model, obtaining the class as output.

Once the individual patches from the slide are classified into benign (0) and malignant (1), an output mask with the same magnification as in the first step (10x), is obtained. This mask makes a spotlight in the regions that are potentially malignant.

After that, the regions from those masks are extracted in 181 x 181 pixel patches, using the second CNN model to classify them in the 5 classes that this model has learnt. Only when the patch is classified as class 4 or 5, this region is highlighted in the final output mask.

2.6. Diagnosis

Finally, the resulting masks are used to classify the nodes and patients in macro, micro, ITC or negative. If at least one of the region of the slide has the major axis length bigger than 2 mm or 125 pixels in the slide at 10x it will be classified as macro-metastases. On the other hand, if the major axis length of a region is between 2 mm (125 pixels) and 0.2 mm (12 pixels), it is classified as micro-metastases. If the mask has regions smaller than 0.2 mm, it is stated to be ITC. Finally, if there is no malignant region, it is considered as negative.

Once every node of every patient is classified, the final decision over the patient is stated among pN0, pN0(i+), pN1mi, pN1 and pN2. This decision is based on the guidelines provided by the evaluation rules of the challenge. Finally, these results are formatted in a submission CSV file.

3. RESULTS AND CONCLUSION

The 10% of the datasets used for CNNs training were dedicated to testing purposes, so it is possible to evaluate the performance of these models on the datasets before they are used in WSIs.

The CNN model 1 achieves a 99.64% of accuracy, with the following confusion matrix:

	Benign	Malignant
Benign	36,230	88
Malignant	139	25,886

The CNN model 2 achieves 95.28% of accuracy, with the following confusion matrix:

	BTP1	BTP2	BTP3	MTP1	MTP2
BTP1	2297	5	55	8	2
BTP2	1	9414	127	3	5
BTP3	120	543	2244	28	19
MTP1	20	8	25	5986	35
MTP2	5	4	25	62	2262

As it is stated, both models have a great performance over the datasets. Whether these datasets are representative enough of the WSIs variability would be the key to obtained good diagnosis performance.

REFERENCES

[1] A. Krizhevsky, I. Sutskever and G. Hinton, "Imagenet classification with deep convolutional neural networks", Advances in neural information processing systems, 1097-1105, 2012

Acknowledgments

The authors acknowledge financial support from the EC Marie Curie Actions, AIDPATH project (Contract No.612471). <http://aidpath.eu/>

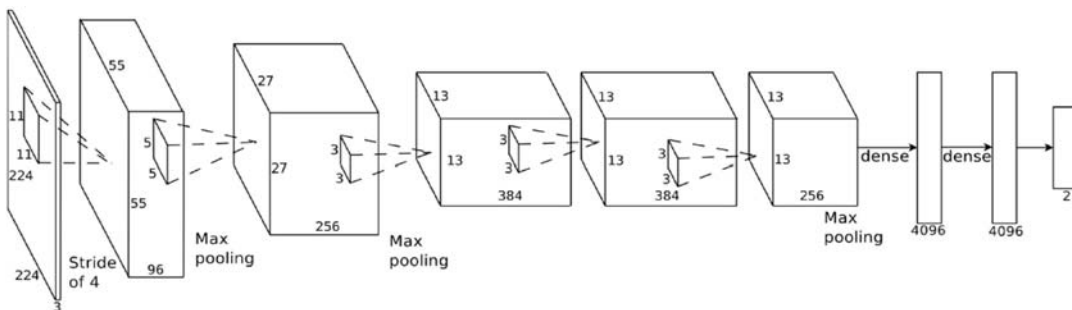


Figure 2 Architecture of the AlexNet CNN