CAMELYON17 GRAND CHALLENGE

Nicolas Pinchaud, *Martin Hedlund*

AI-lab, ContextVision

ABSTRACT

Grading whole slide images (WSIs) is an important task in digital pathology for treatment planning but it suffers from subjectivity and limited reproducibility. The grading of WSIs is also time consuming and therefore expensive. Designing a robust and automatic solution for decision support is a game changer. We propose a fully automatic pipeline from a set of patient whole slide images to pathologic N-stage prediction. Our approach consists of two steps: i) Segmentation of metastasis in whole slide images, ii) pathologic N-stage predictions from segmentations.

***Index Terms*—** segmentation, FCN, metastasis, breast lymph nodes.

0. INTRODUCTION

This work is focused on the prediction of pN-stage for breast cancer patients. The pN-stage assesment is based on the combined metastatic involvement of several lymph nodes and is one of the most important factors in deciding treatment of breast cancer. Common practice in current clinical settings is to assess metastatic involvement of the each lymph node specimen manually under a microscope. Although this task is routinely performed there is room for improvement since this procedure is highly subjective in nature and the task is difficult and time consuming.

The drawbacks of the current manual practice has created interest in automatic assessment of whole slide images for decision support. Automatic methods aim at reducing the subjectivity of current practices and limiting the time for slide assessment. Recent advances using machine learning methods such as convolutional neural networks for image analysis obtained excellent results for the analysis of histological slides [1][2]. The success and development of machine learning algorithms are largely driven by the availability and quality of annotated data. Due to the increased interest in the field several challenges such as TUPAC, AMIDA and CAMELYON [3, 4, 2] have been created in order to push forward scientific research in the field by providing annotated data.

Automatically assessing whole slide images has several challenging aspects. The nature of the task requires information about the specimen on cellular level, this leads to very high resolution i.e. very large images. Assessing the slide requires attention to image structures of approximately 10^4 pixels (a 100×100 pixel neighborhood) within large WSIs containing 10^{10} pixels. Future challenges include solving the subjectivity of annotations and variations in specimen staining (due to different practices regarding staining and slide preparation).

Recent results of previous challenges were focused on performing in slide predictions such as the detection of mitotic figures or segmentation of tumors. Although the state-of-the-art for these tasks is approaching and even surpassing [2] human-level performance these results have not been focused on the task of patient level assessment.

In the CAMELYON17 challenge the objective is to make a patient level prediction based on information from several whole slide images. Our proposed method computes pixel level segmentation of slides. We then extract geometrical properties from the segmentation maps that are used as features to a classification model predicting the slide level metastasis grading. Patient level grading is then computed accordingly.

We propose an ensemble approach: we combine several U-net [5] learned from different pixel resolutions in a directed acyclic graph (DAG) structure. The U-net being a pixel wise classification model (as to compare with a patch wise classification model), its output and input are in the same domain, this allows to combine different U-nets together by concatenating the output of one (or several) model to the input of another.

Since we can learn the U-net on any slide level resolution, this architecture allows integrate the information available from these different levels. It also have the benefits of ensemble learning, we can learn the individual U-nets with different strategies or hyper-parameters inducing different modeling expressiveness that can be integrated in their combination. Another advantage of such approach is that the resources spent to learn individual models are cumulative. The classical non ensemble approach would spend resources to learn different models in order to choose the best hypothesis among different architectures or hyper parameter sets, and then choose the best candidate and disregard the others. This is resource wise expensive because discarded models only contribute in the choice of the best candidate. In our ensemble approach the models are combined and they contribute by providing statistical information about the joint distribution between slide

pixels and annotation data to another model.

In combination to the ensemble approach, we propose to learn the U-nets using boosting inspired technique. Unlike the standard approach, we do not setup a static dataset on which the model is fitted. Instead we sample patches from the slides in a dynamic setup involving two processes. The first process is the *training process* that samples the patches from slides according to *error map* images that contain, for each slide, the pixel classification error of the model. These *error map images* are computed using another process called the *error map process*. These two processes are synchronized and they can run on different computer and GPUs. This approach allows to speedup the learning because the model focuses on regions where the gradients are largest.

1. DATA

1.1. Available data

Two sources of data have been used during training of the model, the data from CAMELYON16 and CAMELYON17. The CAMELYON16 dataset was originally used in the 2016 edition of the Camelyon competition with the objective to detect and localize tumors in the whole slide images. The CAMELYON17 dataset was created for use in the 2017 edition of the Camelyon competition.

The CAMELYON16 dataset contains whole slide images with corresponding annotations of metastatic areas. A small subset of the data has not been exhaustively annotated.

The CAMELYON17 dataset contains whole slide images categorized by patient and clinical center. The annotations available are patient pN-stage and the largest tumor class for each slide. Additional annotations of metastatic areas are available for a subset of the data from each clinical center. The metastatic area annotations are of the same type as from the CAMELYON16 dataset.

1.2. Training and validation sets

For training we used all the slides from CAMAELYON16 and all the patients from CAMELYON17 that had at least one pixel level annotated slide. We had 57 patients remaining for the validation set used to design the architecture of the DAG.

2. METHOD

2.1. Dynamic sampling

2.1.1. Patch wise dynamic sampling

We trained our models with patches dynamically extracted from slides using a pixel-level probability density function inferred from the error of the model on the pixel level classification. The probability density function is computed on a separate process synchronously with the *train process* as illustrated on [Figure 1]. While the *train process* train on a sampled patches set, the *error map process* compute the error map of the next slide. When it has finished, the train process samples a new set of patches. The size of this set is automatically adjusted to avoid waiting state between the two processes. This shorten the duration of cycles and ensures that the sampling of patches is made on a *error map* computed with up-to-date model parameters.

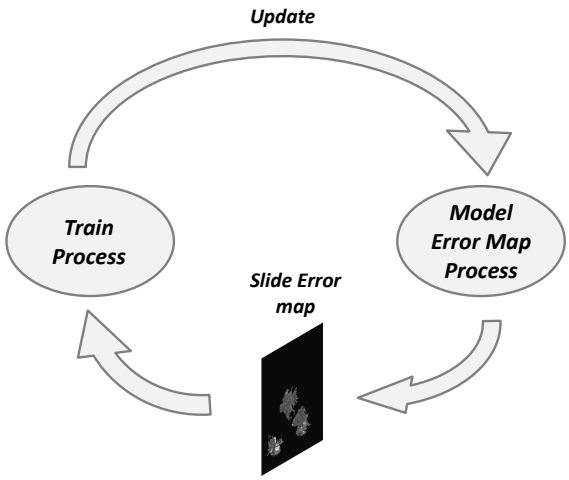

Fig. 1. **Dynamic sampling cycle.** ***Train Process*** samples dynamically from Patch level. ***Model Error Map Process*** dynamically sample on Slide level according to different strategies

2.1.2. Slide wise dynamic sampling

The *error map process* chooses the next slide to sample patches from. This is done using a slide level sampling distribution that gives more probability on slides that contains *greater errors*. This can be designed according to several optimizations strategies. Among them, we used probability distribution that emphasize the optimization of the recall, such that slides that on which the model misses many cancer regions are prioritized. Similarly, we used distribution that optimize the slide level grading classification, slide whose segmentation induces wrong grading classification are sampled more often. As a third strategy we used distribution that optimized informedness score of slides. These different strategies allowed to learned models with wider expressivity. This is a desirable property in the context of model ensembling where we want reduced redundancy between models.

2.2. Model for Metastasis segmentation

Our segmentation model combines 9 U-nets in a DAG structure illustrated in [Figure 2]. Each U-net is learned with variation of hyper parameters, slide sampling distribution and input resolution *mpp (micrometer per pixel)* . Models were stacked according to increasing stages. At first stage (stage 1) we independently trained 3 U-nets. On later stages models are combined with those from previous stages. Their input were augmented by adding extra channels containing the outputs of previous models [Figure 3].

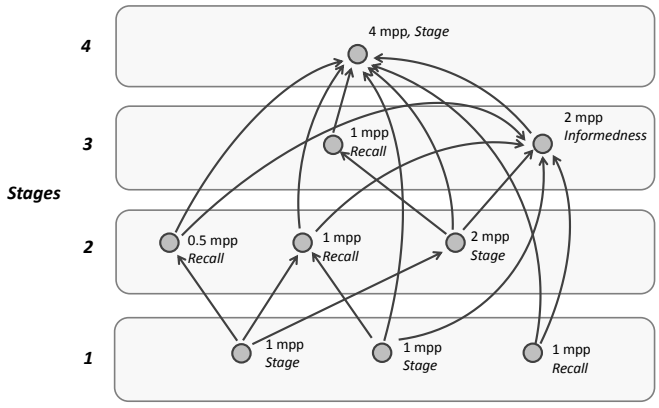

Fig. 2. **DAG structure of the proposed model** *Different slide sampling strategies used: Stage grading error, Recall and Informedness. Different resolutions 0.5-4 mpp.*

Fig. 3. **Combination of Models** *RGB image, Model 1,2..n predictions and Label image.*

For all models we used the same data augmentation, random rotation and mirroring, and color perturbation (hue, saturation, brightness, contrast). The channels corresponding to the connected models were not color augmented.

2.3. pN-stage prediction

For each slide we extracted the diameter of the largest tumor after applying a dilation filter on the segmentation map with a size of 300 micrometers, we did this for nine different thresholds ranging from 0.1 to 0.9, resulting to 9 features per slides. These features are feed into a random forest classifier to predict the slide level grading. The model parameters are optimized using cross validation on the validation set. The final PN-Stage is computed according to the rules.

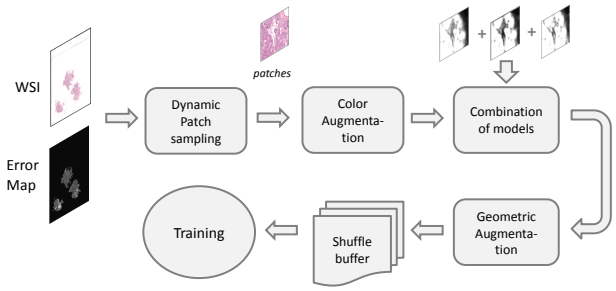

Fig. 4. **Patch Processing pipeline.** *All processing are applied dynamically in the pipeline and are randomized except for the combination of models.*

3. REFERENCES

[1] Dan C Cireşan, Alessandro Giusti, Luca M Gambardella, and Jürgen Schmidhuber, "Mitosis detection in breast cancer histology images with deep neural networks," in *International Conference on Medical Image Computing and Computer-assisted Intervention*. Springer, 2013, pp. 411–418.

[2] Babak Ehteshami Bejnordi, Oscar Geessink, Meyke Hermsen, Geert Litjens, Jeroen van der Laak, Quirine Manson, Mitko Veta, and Nikolaos Stathonikos, "Camelyon16 ISBI grand challenge," 2016.

[3] Mitko Veta, Josien P.W. Pluim, Nikolaos Stathonikos, Paul J. van Diest, Francisco Beca, and Andrew Beck, "Tumor proliferation assessment challenge 2016 , MICCAI grand challenge," 2016.

[4] Mitko Veta, Paul J Van Diest, Stefan M Willems, Haibo Wang, Anant Madabhushi, Angel Cruz-Roa, Fabio Gonzalez, Anders BL Larsen, Jacob S Vestergaard, Anders B Dahl, et al., "Assessment of algorithms for mitosis detection in breast cancer histopathology images," *Medical image analysis*, vol. 20, no. 1, pp. 237–248, 2015.

[5] O. Ronneberger, P.Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *Medical Image Computing and Computer-Assisted Intervention (MICCAI)*. 2015, vol. 9351 of *LNCS*, pp. 234–241, Springer, (available on arXiv:1505.04597 [cs.CV]).