LEVERAGING GLOBAL DIAGNOSIS FOR TUMOR LOCALIZATION IN DYNAMIC CELL IMAGING OF BREAST CANCER TISSUE TOWARDS FAST BIOPSYING

D. Mandache[a,b], E. Benoit à la Guillaume[b], M.-C. Mathieu[c], J.-C. Olivo-Marin[a], V. Meas-Yedid[a]

[a] BioImage Analysis Unit, CNRS UMR 3691, Institut Pasteur, 25 rue du docteur Roux, 75015, Paris, France

[b] LLTech, 58 rue du dessous des berges, Bâtiment A, 75013, Paris, France

[c] Department of Medical Biology and Pathology, Gustave Roussy Cancer Campus, 114 rue Edouard Vaillant, 94805, Villejuif, France

ABSTRACT

We propose a fast aid-to-diagnosis biopsy assessment method convenient at the point-of-care, on account of both the imaging technique and the algorithm applied. The procedure implies a pipeline of classification and localization of tumors in breast cancer biopsies imaged with a recently developed non-invasive imaging modality: Dynamic Cell Imaging (aka Dynamic Full Field Optical Coherence Tomography). This allows for fast and interpretable extemporaneous cancer detection with high confidence; we obtained a performance of 96% classification accuracy together with a coarse localization of tumors, even so for single isolated invasive cells.

Index Terms— dynamic full field optical coherence tomography, classification, segmentation, convolutional neural network, aid to diagnosis, breast cancer

1. INTRODUCTION

Breast cancer touches 1 in 8 women worldwide, making it the most frequent cancer in women and the second most deadly.

Standard diagnosis procedure consists in mammography screening, where tissue of abnormal density can be suspected through X-rays and is then biopsied and analyzed at microscopic scale to be actually diagnosed. Localizing and reaching the lesion is not a trivial task. The common protocol implies multiple samplings (minimum 5 [1]) of the suspicious mass to ensure correct probing for reliable diagnosis. What is more, the false negative rate caused by poor sampling is currently up to 2% [2].

In the eventuality of a positive diagnosis, the standard treatment involves the surgical removal of the tumor, with partial or total breast ablation. Even after heavy surgery, the risk of recurrence after 5 years is above 10%, suggesting that an imperfect removal of the tumor was performed during surgery. Hence, there is a crucial need to encourage real-time intraoperative assessment of the tumor margins, in order to reduce the ablation of healthy tissue, the surgery time, and the risk of resurgence.

The gold standard procedure for tissue analysis is histology, it consists of fixing, slicing and Hematoxylin & Eosin (H&E) staining of the sample, which is then analyzed by a pathologist under a microscope with up to 40X magnification. The whole procedure takes up to a couple of days. As for its intra-operative variants, frozen section or imprint cytology, even if they are faster $\sim$ 30min, they are still not efficient enough to be widely-employed during surgery.

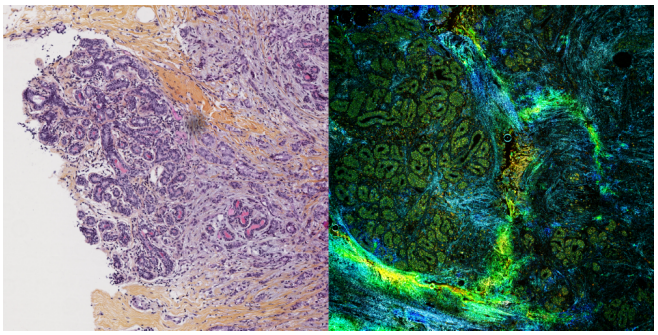

Fig. 1: Healthy breast lobule in H&E and DCI.

We propose the use of an optical sectioning solution that offers histology-like structure appearance in minutes, without requiring any tissue preparation. The technique is Dynamic Cell Imaging (DCI) [3], a time-resolved variant of Full Field Optical Coherence Tomography (FFOCT) [4], which in turn is a *en-face* development of the classical OCT. While FFOCT reveals highly backscattering elements (mostly fibrous structures), DCI captures the movement of scatterers in the cells (presumably vesicles and mitochondria), therefore signal is obtained only in freshly excised living tissue. Both techniques offer an intracellular-level resolution of $1\,\mu\mathrm{m}$, with a 10X objective, compatible with the need of clinicians. However, the attractiveness comes from the fast acquisition, for FFOCT on the order of seconds and DCI a couple of minutes for an average core-needle biopsy.

FFOCT and DCI could potentially reinforce extemporaneous margin assessment by offering a faster and simpler imaging protocol, that would lead to improved outcome for breast-conserving surgery.FFOCT and DCI can also help in quality assessment of biopsies by reducing the number of samples taken per lesion. In such a manner the surgeon or radiologist can decide on-line, in less than 5 minutes, if there

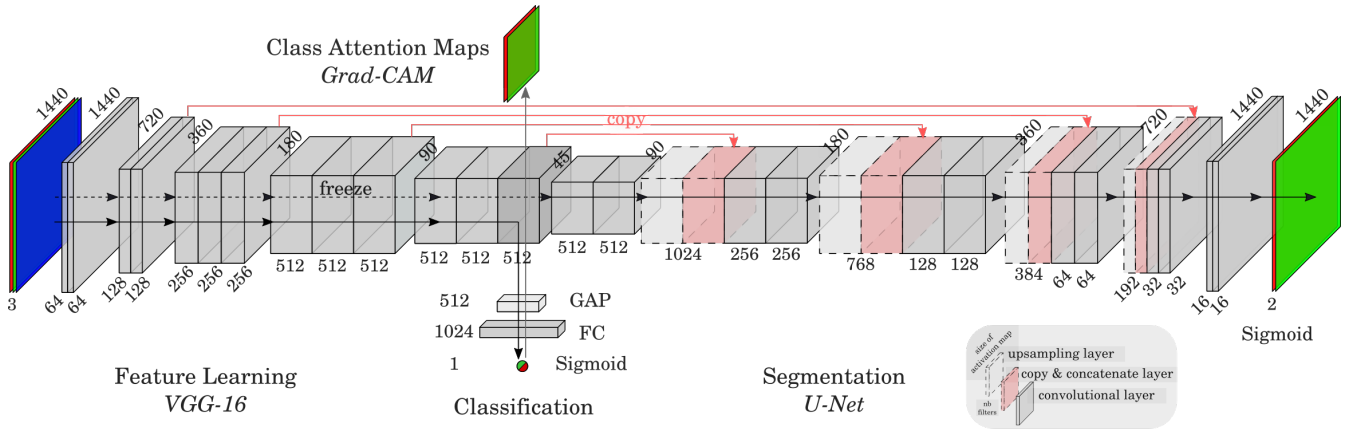

Fig. 2: CNN architecture & 3-step workflow: i) feature extraction and classification with VGG-16 ii) Attention Map computation (offline) with GradCAM iii) Segmentation by building a U-Net architecture having the pre-trained VGG-16 as backbone.

needs to be another incision by imaging the already excised biopsy and quantifying the number of cells or nature of the tissue.

A recent study [5] on 173 breast resections imaged with FFOCT and DCI obtained minimum sensitivity and specificity over 85% from 2 breast surgeons, after a 3-hour training by pathologists, showing promising prospects on the adoption of these novel techniques. However, since the obtained contrasts are significantly different from standard histology (See Fig. 1), FFOCT and DCI images are still difficult to interpret by surgeons or radiologists, let alone to be adopted in current practice. Previously, there have been endeavors on automatic detection of cancerous areas from FFOCT images, in [6] a CNN approach reached 96% accuracy in detecting non-melanoma skin cancer, [7] also detected skin cancer in mice with 80% accuracy. However skin cancer is of less morphological complexity and one of the main features for diagnosis is collagen fiber density and organization, which makes it a very suitable application of FFOCT (and other non-invasive optical techniques as a matter of fact). DCI, on the other hand, produces contrast of individual cells, allowing diagnosing more challenging pathologies. However, due to the novelty of the technique and the difficulty to obtain ground truth, no automatic diagnosis methods have been developed so far.

In this work we propose an aid-to-diagnosis method that could help in training and assisting the clinicians in taking rapid extemporaneous decision. We present a deep-learning approach towards classifying between cancerous and normal breast tissue from DCI images, while remaining in the scope of interpretability and pushing the method further towards segmentation. The method is performed on 47 samples coming from a cohort of 33 patients after lumpectomy or mastectomy.

2. TUMOR CLASSIFICATION

The dataset used is the product of a prior clinical study conducted in Gustave Roussy cancer center to test the feasibility of DCI-based cancer diagnosis, namely breast cancer, with 90% accuracy and a minimum of 89% sensitivity and 80% specificity per-pathologist. From a cohort of 33 patients, there have been obtained 47 samples issued of surgical waste from full or partial breast ablations (34 samples containing tumor and 13 normal tissue). As a consequence of the early stage of the technical development process of the DCI technique at the moment of the clinical study, only individual fields-of-view (FOVs) have been acquired from different locations of each sample. Each FOV has $1440 \times 1440px$, corresponding to a physical size of $1.3 \times 1.3mm$. Although sufficient to cover specific structures like lobules and ducts in their entirety, a global overview of the tissue architecture (provided by FFOCT) was needed. There was a median number of 9 FOVs per sample, ranging up to 15, resulting in a dataset of 396 total individual FOV (260 containing tumor on at least 5% of their area and 136 with no tumor). The per FOV interpretation which served as ground truth was done by a pathologist trained on DCI images and having the corresponding standard H&E slide as reference for diagnosis.

To obtain the train and test sets, we separated the samples with $80\% : 20\%$ proportionality in a random stratified manner, meaning that the two sets were constructed to respect the class distribution of the global set. This results in 37 samples in the train set (28 malignant and 9 healthy samples, 226 tumor-positive (P) and 97 tumor-negative (N) FOVs) and the remaining 10 samples in the test set (6 malignant and 4 healthy, 39 P and 34 N FOVs).

Due to the very small and heterogeneous dataset, we opted for a pre-trained network, namely VGG16 [8] for its relatively small number of parameters, pre-trained on ImageNet[9] (note that training from scratch failed to converge.). Other architectures were tested, namely InceptionV3 and ResNet50 which were very fast to overfit, most likely due to their higher number of parameters: $\sim 25M$ vs. $\sim 15M$ for VGG16. Following the same philosophy of reducing the number of parameters to a minimum, after the convolutional feature extraction layers was added a Global Average Pooling Layer (GAP)[10] which produces a feature vector of size 512

representing the average activation of each filter of the last convolutional layer of VGG16. Intuitively, a narrow bottleneck forces compression therefore leading to generalization. Another advantage of GAP is that it makes the model more versatile in the sense that it can accept inputs of different sizes, because its size is only dependent on the number of filters. After pooling, the actual classifier was also kept to a minimum of complexity with only one hidden layer of size 1024, followed by the binary output neuron with sigmoid activation.

Here-described configuration was fine-tuned using the Stochastic Gradient Descent (SGD) optimizer with a learning rate of $1e-4$ and 0.8 momentum by minimizing the weighted binary cross-entropy loss on mini-batches of size 3 (dictated by memory constraints). The weight for each class is inversely proportional to its frequency, resulting in $w_{c=0} = 1.7$ and $w_{c=1} = 0.7$ computed as $w_c = \frac{N}{n_{classes} * N_c}$ where N is the total number of samples and N_c is the number of samples belonging to class c. The best model was found after 104 training epochs, after which the model started to overfit.

3. CLASSIFICATION RESULTS AND INTERPRETATION

In terms of results, we obtained : accuracy 95.89%, sensitivity 91.18% and specificity 100% (3/39 missed tumoral FOVs) AUC 1. Aggregating the per-FOV predictions to obtain a global per-sample diagnosis was done using the 95^{th} quantile of the FOV prediction distribution, this approach ensures more robustness to outliers than the maximum and it is proven by the correct classification of all the samples in the test-set.

In order to validate the presented hyperparameter selection we have performed a 5-fold cross validation with the following results : mean accuracy of $89 \pm 4\%$ (sensitivity $88 \pm 4\%$ and specificity $86 \pm 6\%$) and the area under the ROC curve (AUC) of 0.92 ± 0.02 at the FOV level. Bearing in mind that the folds where defined by sample, the number (and class proportion) of actual images (FOVs) in train and test sets fluctuates from one fold to the other, which together with the heterogeneous nature of the data explain the slight variation in performance.

In a quest for interpretability and confidence in the trained model, we worked on two approaches: visualizing the learned features and visualizing the per-sample attention with respect to the class. For the first endeavor, we obtain synthetic inputs through gradient ascent by maximizing the activation of each convolutional filter iteratively, using the method in [11], and we obtain the textures learned from the data. See Fig. 3 for an example of some filters of the deepest convolutional layer. Doing so we make sure that our filters are different from the ImageNet filters, not noisy and also true to the dataset at hand, so indeed we can conclude that the learning has also been extended to the feature extractor, not only limited to the classifier.

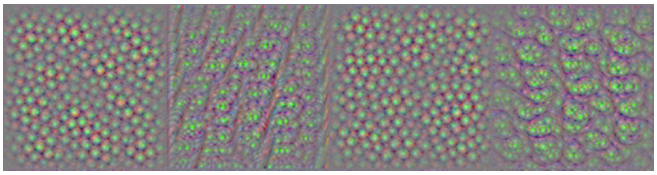

Fig. 3: Some learned filters (4/512) of the last convolutional layer (whose activation maps where used to compute the attention maps) hinting to different fiber and cell organization, including also multiple cell sizes.

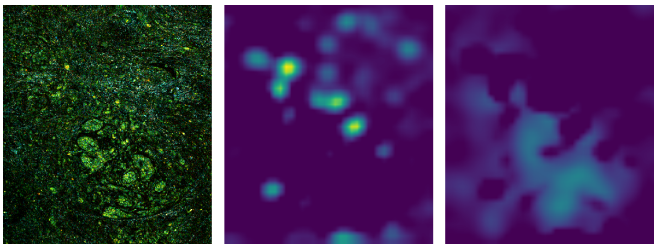

Fig. 4: a) crop of an image showing healthy breast lobule surrounded by isolated infiltrating cancerous cells, correctly predicted as cancerous with 97% confidence, b) tumor-positive attention map, c) tumor-negative attention map.

The second approach consists in displaying the class activation maps of several inputs using the GradCAM [12] method, which reveals the "important" areas in an input indicating towards a certain class. The averaged gradients flowing back from a chosen class output neuron to a previous layer (usually last convolutional layer) act as weighting factors for each activation map, the final result being a linear combination between the weights and filter activation maps.

This results in a coarse localization of the class presence in the input which can serve numerous purposes, an important one is verifying that the model is not biased (e.g. higher importance to context, rather than the actual object of interest or, on the other hand, a very focused attention on a small part of the object).

Common practice is to pass the grad through a ReLU (in other words, keep only positive values) and then scale $[0, 1]$, however since we deal with binary classification we also extract the (absolute value of) negative gradient too as an indicator for the absence of the tumor class (i.e. healthy). scale to $[-1, 1]$ range with respect to the absolute maximum activation value. See Fig. 4 for an example of the positive GradCAM (localizing cancer cells) and the negative GradCAM highlighting a normal lobule (confirmed by pathologist).

4. SELF-SUPERVISED SEGMENTATION

The obtained attention maps were confronted against the previous interpretation of the pathologist on several interesting FOVS (i.e. that contained both normal and cancerous structures or on which they had doubts) and we have thus decided that is pertinent to leverage this coarse localization to guide a segmentation model. To pursue this we transformed the attention maps into segmentation mask which would serve as

ground truth for training a U-Net [13] built by merging the network already trained on the classification task and adding a decoder branch. See Fig. 2 for more details on the architecture. The pre-trained branch is "frozen" meaning that we are building upon the classification features and there are only the parameters of the decoder left to train ($\sim 9M$ parameters).

Noting that there is no high-confidence ground truth available for segmentation, the processing steps of converting the attention maps into segmentation mask, as well as the choice of the loss to optimize were guided by two aspects: i) the 16X lower resolution of the GradCAM i.e. 90×90 px as opposed to 1440×1440 px image resolution and ii) GradCAM's documented weak point that it usually captures only the most discriminative part of the classified object or only one instance of the object. The attention maps were upscaled to the input size using bilinear interpolation, followed by Otsu thresholding, morphological dilation with a circular structuring element of radius $r = 15$, and Gaussian filtering with $\sigma = 15$ to account for the uncertainty on the boundaries. We also zeroed out the positive attention maps for normal samples, knowing there are no tumor cells present in normal FOVs, but there could be several healthy structures present in cancerous FOVs.

We train the decoder by minimizing Tversky loss [14] using Adam optimizer with a rate of $1e-4$. The loss, defined as $\mathrm{L} = 1 - \frac{TP}{TP + \alpha FN + \beta FP}$, which is a generalization of the more popular Dice loss that introduces unbalanced penalization of classes vs background. The penalization parameter $\alpha = 0.6$ (chosen from literature), meaning that false positives (FP) are penalized higher than false negatives (FN) i.e. modeling the fact that we have high confidence in the "pre-segmentation" already obtained through GradCAM, but we encourage an extended segmentation of the entire areas of interest; by extension $\beta = 1 - \alpha = 0.4$ relaxes the penalization on adding "new" pixels to the segmentation. Less parameters to train allow for a sightly bigger batch size of 5. Knowing that U-Net is generally fast to converge and our generated ground truth is not of 100% confidence, we stop training when the loss is stabilized, after 15 epochs. Visually, the segmentation obtained is slightly finer and indeed including more cells, but we can not give a quantitative result at this point in the study.

5. DISCUSSION AND CONCLUSION

To sum up, we introduced a method where classification and segmentation stream-lined together to obtain a high confidence classification jointly with a coarse segmentation of DCI breast specimens. Qualitative analysis of the model shows that the CNN discriminates between normal and cancerous area in the same sample without explicit training and learns textures consistent with the data. It can also detect isolated cells which include also low contrast cells that can be difficult to spot otherwise. It is difficult to quantify the improvement brought by the added decoder to the fidelity of the already computed segmentation masks, however a clear advantage of the used approach is that the model (i.e. weights) can be easily deployed and be used in a plug-and-play manner with dedicated software, like Cytomine [15] or Icy [16]. Henceforth, its allows the pathologist to give their feed-back and corrections with little throughput, which would lead to improved expert annotations.

6. REFERENCES

[1] G Sauer, et al., "Ultrasound-guided large-core needle biopsies of breast lesions: analysis of 962 cases to determine the number of samples for reliable tumour classification," *British Journal of Cancer*, vol. 92, pp. 231–235, 2005.

[2] Hyun Youk Ji, et al., "Missed breast cancers at us-guided core needle biopsy: How to reduce them," jan 2007.

[3] C. Apelian et al., "Dynamic full field optical coherence tomography: subcellular metabolic contrast revealed in tissues by interferometric signals temporal analysis," *Biomedical Optics Express*, vol. 7, no. 4, pp. 1511–24, 2016.

[4] A. Dubois et al., "High-resolution full-field optical coherence tomography with a linnik microscope," *Applied Optics*, vol. 41, no. 4, pp. 805–12, 2002.

[5] Houpu Yang, et al., "Use of high-resolution full-field optical coherence tomography and dynamic cell imaging for rapid intraoperative diagnosis during breast cancer surgery," *Cancer*, vol. 126, no. S16, pp. 3847–3856, aug 2020.

[6] D. Mandache, et al., "Basal cell carcinoma detection in full field OCT images using convolutional neural networks," in *Proceedings - International Symposium on Biomedical Imaging*, 2018, vol. 2018-April.

[7] Chi-Jui Ho, et al., "Detecting mouse squamous cell carcinoma from submicron full-field optical coherence tomography images by deep learning," *Journal of Biophotonics*, sep 2020.

[8] Karen Simonyan and Andrew Zisserman, "VGG-16," *arXiv preprint*, 2014.

[9] J. Deng, et al., "ImageNet: A Large-Scale Hierarchical Image Database," in *CVPR09*, 2009.

[10] Min Lin, et al., "Network In Network," *arXiv preprint*, p. 10, dec 2013.

[11] D. Erhan, et al., "Visualizing Higher-Layer Features of a Deep Network," 2009.

[12] Ramprasaath R. Selvaraju, et al., "Grad-CAM," *ICCV*, 2017.

[13] Olaf Ronneberger, et al., "Unet," *MICCAI2015*, 2015.

[14] Seyed Sadegh, et al., "Tversky loss function for image segmentation using 3D fully convolutional deep networks," Tech. Rep.

[15] Raphaël Marée, et al., "Collaborative analysis of multi-gigapixel imaging data using Cytomine," *Bioinformatics*, vol. 32, no. 9, pp. 1395–1401, may 2016.

[16] Fabrice de Chaumont, et al., "Icy: an open bioimage informatics platform for extended reproducible research.," *Nature Methods*, vol. 9, no. 7, pp. 690–6, 2012.

This study was performed in line with the principles of the Declaration of Helsinki. Informed consent was obtained from all individual participants involved in the study.

The authors acknowledge financial support of the ANRT (grant CIFRE no 2018/0139) and the Inception program (Investissement d'Avenir grant ANR-16-CONV-0005) for providing the computing resources, as well as the help of Quang tru Huynh on using them.