SEGMENTATION BASED DATA AUGMENTATION IN DEEP NETWORKS FOR MAGNIFICATION INVARIANT BREAST CANCER DETECTION

[1]Padmaja Jonnalagedda, [2]Daniel Schmolze, [1,3]Bir Bhanu
[1]Department of Electrical and Computer Engineering, University of California, Riverside
[2]Department of Pathology, City of Hope National Medical Center, Duarte CA
[3]Department of Bioengineering, University of California, Riverside
Contact: *sjonn002@ucr.edu*

Introduction: Histopathological images are the most trusted for breast cancer diagnosis compared to all other medical imaging techniques. Despite the advances in digital imaging techniques, a lot of the diagnosis is still performed manually by pathologists. The task is complicated and tedious, with pathologists having to analyze the same tissue at various magnifications for an accurate diagnosis. Furthermore, there are subtle distinctions among breast cancer images in texture, morphology etc., making cancer detection a complex task. Many researchers have proposed approaches for automatic breast cancer detection; however, their methods and results depend on the magnification at which the images were taken. To mitigate these issues, the authors propose a quantitative automated approach to detect cancer, independent of the image magnification level.

Materials and methods: The proposed method utilizes ResNet50, a deep learning architecture, for detection of cancer using the BreaKHis database [1]. It consists of 2480 benign and 5429 malignant tumor images (460x700 pixels, RGB images) at 4 magnification levels from 82 patients. In histopathology images, nuclei and surrounding area are indicators of tumor and its type. The proposed method leverages the location of these indicators. The images in the training set are subjected to clustering using a mixture of Gaussian models with three components. An image displaying one of the clusters contains only nuclei, called the *intermediate image*. In this intermediate image, a patch slides to find the area with the maximum number of nuclei. The patch size used here is 230x350 with 50% overlap. The patch size is chosen to preserve aspect ratio and obtain 2X magnified final image. From the original image, this area of interest is then extracted, resized to the size of original image and added to the training set. This has two advantages: (i) it balances the classes in training set by adding more images to the benign set, (ii) it adds randomly magnified images (zoomed images) to make training of magnification invariant model more robust. The images are divided into 70% train, 15% validation and 15% test sets (61 patients in training/validation, 21 for testing). The ResNet50 model was run on Python 3.6 and trained on an NVIDIA GeForce GTX1080 GPU.

Results and discussion: The performance of the proposed method is compared to a current method as well as random patch zooming to evaluate the impact of using the more informative patches to augment the training set. The accuracy of the detection method by Byramoglu et. al. [2] is 83%. The accuracy with the proposed method is 92.7%, with sensitivity 94.55% and specificity 91.25%. Compared to this, using random patches gives as accuracy of 89.9% with sensitivity 94.3% and specificity 88.3%. The training time is 60-70 minutes and testing time per image is 0.25 seconds. These results are obtained for 10-fold cross validation testing.

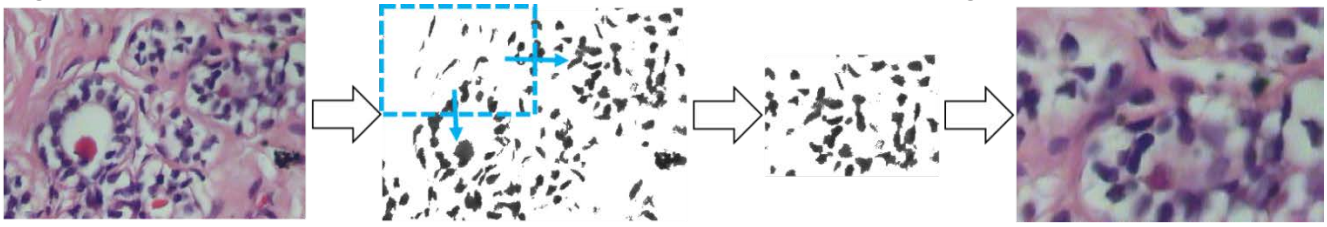

Figure 1: Original image → Intermediate image with sliding patch → Patch (mask) with most nuclei → Resized patch added to train set

Conclusions: The proposed approach provides better performance in magnification independent classification. Gaussian Mixture Model brings attention to areas with tumor nuclei that carry more information in tumor detection than data augmentation using random cropped patches from tissue images.

References:

[1] Spanhol, F., Oliveira, L., Petitjean, C. & Heutte, L. A dataset for breast cancer histopathological image classification. IEEE Transactions on Biomedical Engineering (TBME) 63(7), 1455–1462 (2016).
[2] Bayramoglu, N., Juho K., and Janne H. "Deep learning for magnification independent breast cancer histopathology image classification." In *International Conference on Pattern Recognition (ICPR)*, pp. 2440-2445, 2016.