Paper Reading No.4

Multi Scale Curriculum CNN for Context-Aware Breast MRI Malignancy Classification

Sheng Lian

June 2019

1 Brief Paper Intro

- ***Paper ref: MICCAI 2019*** , http://arxiv.org/abs/1906.06058 code will be available on github.
- ***Authors: see Figure 1***

Christoph Haarburger[1], Michael Baumgartner[1], Daniel Truhn[2,1], Mirjam Broeckmann[2], Hannah Schneider[2], Simone Schrading[2], Christiane Kuhl[2], Dorit Merhof[1]

[1]Institute of Imaging and Computer Vision, RWTH Aachen University, Germany
[2]Dpt. of Diagnostic and Interv. Radiology, University Hospital Aachen, Germany
christoph.haarburger@lfb.rwth-aachen.de

Figure 1: authors' brief intro.

- ***Paper summary:*** For caner malignancy classification task, the general way is to tackle it as an object detection problem. Individual lesions are first localized and then classified with respect to malignancy. This way has the drawback that global medically relevant information are disregarded. This paper propose a 3D CNN and a multi scale curriculum learning strategy to classify malignancy globally based on an MRI of the whole breast, which is not rely on lesion segmentation.

- ***Reading motivation:*** I want to learn something about curriculum learning and found this MICCAI 2019 paper released on arxiv. But, compared with other 3D medical image segmentation methods, the proposed method does not have many highlights, and is not so 'curriculum learning', in my opinion, so there isn't much to introduce about this method. However, there are many places worth to learn from this paper's writing. And I want to understand the role of active learning and curriculum learning in medical image analysis.

2 Data Intro

This paper proposed a breast MRI dataset which consists of dynamic contrast-enhanced (DCE) MR images of 408 patients from clinical routine at their institution. Out of the 408 patients, 305 had malignant and 103 had benign findings. And the overall ratio of malignant and benign samples at breast level (rather than patient level) in the whole dataset is 40.4%/59.6%. All images were resampled to 512*512*32 voxels.

As introduced above, the proposed approach does not rely on lesion segmentations. However, for comparison with segmentation based approaches such as Mask R-CNN, all lesions were manually segmented on every slice by a radiologist with 13 years of experience in breast MRI.

3 Method

The framework of the proposed method is shown in Figure 2. For the backbone model, several state-of-the-art architectures are evaluated, including ResNet, DenseNet, FPN, U-Net, et al. But the choice of the backbone architecture did not have a significant impact on the overall performance, so authors focus on ResNet18.

The proposed **Multi Scale Curriculum Training** contains 2 stages:

1) **Classification of 3D lesion patches** with size 64*64*64. Each patch contains at least one lesion. If at least one malignant lesion is contained in the patch, the patch label is set to malignant and benign otherwise.

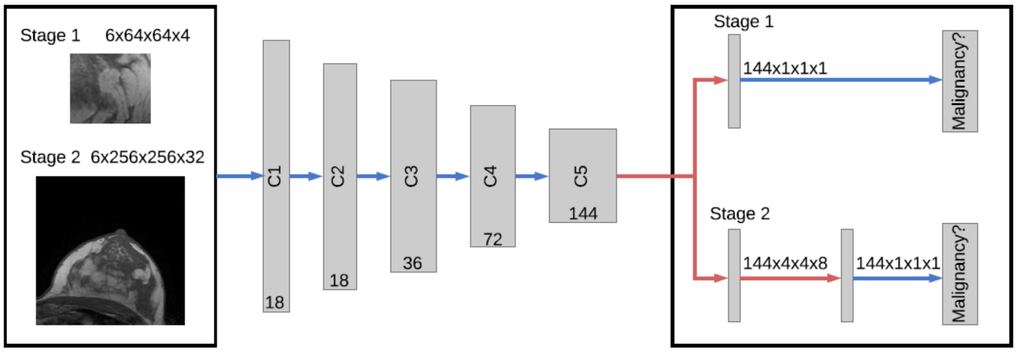

Figure 2: Network architecture: Residual Blocks are named C_i, convolutions are indicated in blue, pooling operations in red.

2) **Classification of 3D volumes** containing a whole breast with size 256*256*32. In order to unify the scale of the input, an additional adaptive average pooling layer is introduced. The five time points from the dynamic T1-weighted series and the T2-weighted series were fed into the CNN in the channel dimension, which leads to an input volume of 6*256*256*32 (channels, x, y, z).

Furthermore, a series of data augmentation operations including rotation,mirroring are introduced.

4 Experiments

This paper choose naive 3D ResNet18, Mask R-CNN, and Retina U-Net for comparison. Among these methods, naive 3D ResNet18 is trained without curriculum scheme (stage 2 only). The comparisons between the methods are indicated in Figure **??**. We can find that the highest AUROC is achieved by the radiologist, followed by ResNet18 Curriculum and Retina U-Net. The naive image classification approach achieved a very poor performance.

	AUROC	Accuracy	#Parameters
Mask R-CNN [3]	0.88 ± 0.01	0.77 ± 0.03	3.91M
Retina U-Net [6]	0.89 ± 0.01	0.82 ± 0.02	3.90M
ResNet18 Naive	0.50 ± 0.04	0.45 ± 0.05	2.66M
ResNet18 Curriculum	0.89 ± 0.01	0.81 ± 0.02	2.66M
Radiologist	0.93	0.93	-

Figure 3: Test performance of the comparison methods and proposed approach over a 5-fold cross validation.

5 Some thoughts on curriculum learning and active learning

Curriculum learning is first proposed by Bengio et al. on ICML2009 [1]. Beyond curriculum learning (CL) introduced in this paper, active learning (AL) also help a lot in the field of medical image analysis, especially when the data is limited. I have read a classic paper about active learning in the field of medical image analysis in CVPR2017 [2]. Maybe I will take a note on this paper next time.

Actually, these two methods have similar places. The commonality between the two methods is that they all start from the perspective of adjusting the learning sample, but the purpose is not the same. AL tries to answer the question that if machine can ask questions, can they complete the training with fewer samples? While CL tries to find the answer that for machine, is it better to learn from easy to difficult, like human being do? The former is for less data, while the latter is for better and faster training. Generally, AL tries to pick informative samples (usually hard samples) for training, which will lead for faster model convergence. But, hard samples are sometimes noisy, sometimes cannot be learnt.

Just like human education, our goal is to achieve a high level of mathematics for a student. But we don't teach primary school students calculus directly, but start with the simplest '1+1'. This is also the starting point for CL. To train a model, maybe you need to start from a simple beginning.

AL is a direction that has been learn a lot. AL is suitable for situations where the amount of data is small or the cost of labeling is expensive. But for specific scenario, but it's still questionable for the way choosing training data.

6 Some sentences that is useful

In this section, I collect some sentences that is useful for future paper writing.

1) *'However, there is not always a clear consensus among clinicians on which object is considered a suspicious lesion that should be segmented.'*

2) *'The degree of natural variation cannot be captured by interpreting tumors as objects with hard boundaries.'*

3) *'As a consequence, most datasets are rather small, preventing that deep learning algorithms unfold their full potential and limiting the power of evaluation results.'*

4) *'However, this approach has produced poor results so far because it is a needle-in-haystack kind of problem'*

References

[1] Yoshua Bengio, Jérôme Louradour, Ronan Collobert, and Jason Weston. Curriculum learning. In *Proceedings of the 26th annual international conference on machine learning*, pages 41–48. ACM, 2009.

[2] Zongwei Zhou, Jae Shin, Lei Zhang, Suryakanth Gurudu, Michael Gotway, and Jianming Liang. Fine-tuning convolutional neural networks for biomedical image analysis: actively and incrementally. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 7340–7351, 2017.