Multitask Classification and Segmentation for Cancer Diagnosis in Mammography

Thi-Lam-Thuy Le[1,2] THILAMTHUY.LE@GE.COM
Nicolas Thome[1] NICOLAS.THOME@CNAM.FR
Sylvain Bernard[2] SYLVAIN.BERNARD@MED.GE.COM
Vincent Bismuth[2] VINCENT.BISMUTH@GE.COM
Fanny Patoureaux[2] FANNY.PATOUREAUX@GE.COM
[1] *MSDMA Team, CEDRIC Lab, Conservatoire National des Arts et Metiers, France*
[2] *Mammography Image Quality Team, GE Healthcare, France*

1. Introduction

Deep learning and Convolutional Neural Networks (ConvNets) (Krizhevsky et al., 2012; Durand et al., 2015) recently show outstanding performances for visual recognition. The representation learning capacity of ConvNets has also been successfully applied to medical image analysis and recognition in mammography (Huynh et al., 2016; Lévy and Jain, 2016; Geras et al., 2017; Kooi et al., 2017; Nikulin; Lotter et al., 2017).

A major challenge in the medical domain relates to the collection of a large amount of data with clean annotations, which is especially difficult due to the strong level of expertise required to perform data labelling. Regarding mammography image analysis for Computer-Aided Diagnosis (CAD), annotations can be very diverse, from global binary image labels (presence / absence of a cancer), *e.g.* DDSM (Heath et al., 2000), to finer-grained and pixel-level annotations (mass, calcification, both benign or malignant labels), to BI-RADS labeling, *e.g.* INbreast (Moreira et al., 2012).

In this work, we propose to leverage these heterogeneous but correlated forms of annotations to improve performances of deep ConvNets. To this end, we introduce a Multi-Task learning (MTL) scheme, which combines pixel-level segmentation and global image-level classification annotations. The proposed architecture is based on a Fully Convolutional Networks (FCN), which enables efficient feature sharing between image regions and fast prediction. We evaluate our model on the DDSM database (Heath et al., 2000), with cancer classification and pixel segmentation with five classes. We show that the joint training is able to learn shared representations that are beneficial for both tasks. Our method can be seen as a generalization of approaches relying on detection annotations to pre-train deep model for a classification purpose, *e.g.* Nikulin; Lotter et al. (2017). We show that our joint training of classification and segmentation enables a better cooperation between tasks.

© 2019 T.-L. Le, N. Thome, S. Bernard, V. Bismuth & F. Patoureaux.

2. Proposed Multi-Task Classification and Segmentation Architecture

As shown in Figure 1, the proposed network is based on a Fully Convolutional Network (FCN) (Chen et al., 2018) with a ResNet backbone (He et al., 2016) to extract local features. The FCN enables efficient feature computation in each region but also sharing computation from all regions in the whole image in a single forward pass. In addition, we can still process input images with high spatial resolution (e.g. 1152×832 in our experiments). We combine segmentation and classification, thus heterogeneous annotations can be leveraged to jointly optimize both tasks:

$$\mathcal{L}_{total} = \lambda \cdot \mathcal{L}_{cls} + (1 - \lambda) \cdot \mathcal{L}_{seg} \tag{1}$$

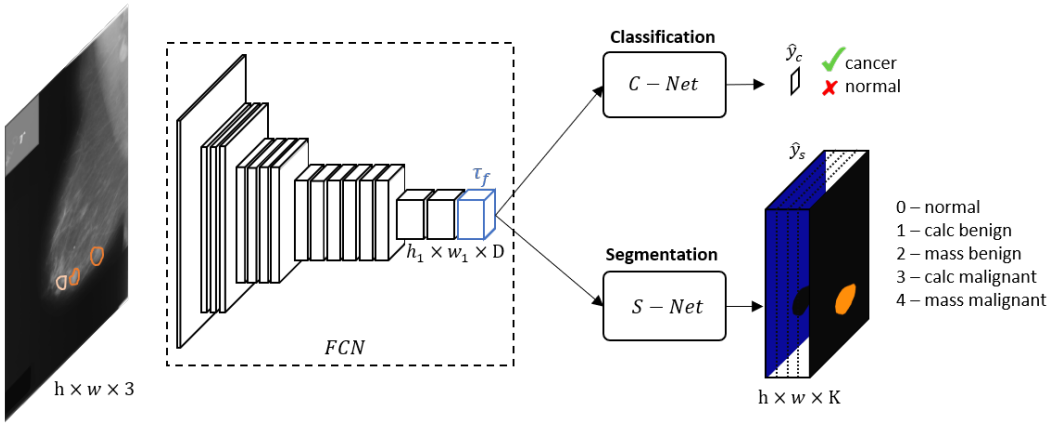

Figure 1: Overall architecture. It is based on a FCN to extract local features from the whole input images. Local features are then encoded by S-Net to yield lesion segmentation, and aggregated by C-Net to yield cancer classification.

Segmentation Network (S-Net) The segmentation network aims at classifying each pixel in the input image into a set of K pre-defined classes. More precisely, from the output tensor T_f of the backbone FCN, S-Net first consists in adding a transfer layer of 1×1 convolution to transform the output of the feature extraction network from large scale dataset to our K-class target dataset. Then, we perform an upsampling to create the semantic segmentation. On top of that, we define a weighted cross-entropy loss $\mathcal{L}_{seg}$ to address the class-imbalance issue ('healthy' regions are dominant compared to other lesion classes).

Classification Network (C-Net) The classification branch takes as input the shared tensor T_f and outputs a single class label assessing the presence or absence of cancer in the input image. C-Net is composed of two steps. Firstly, local features are aggregated with a global average pooling (Lin et al., 2013) (GAP) to yield a single score per modality, over all regions in the input image. The second step consists in the last fully connected layer to get the final probability of cancer. We define the classification loss $\mathcal{L}_{cls}$ as a standard binary cross entropy as is common in classification tasks.

3. Results and Discussion

Quantitative results To highlight the relevance of our method, we perform classification and segmentation in three ways with the labeling scheme depicted in Figure 1: (1) train individually classification and segmentation models to get baseline performances on each task, (2) sequentially train the segmentation model and finetune local features for the classification task, and (3) jointly train both tasks. We evaluate segmentation and classification performances using respectively the mean Dice score over five classes and the Area Under Curve (AUC) of Receiver Operating Characteristic (ROC). As shown in Table 1, by pretraining the model on segmentation, our classification performance is slightly improved by ~ 1 pt to $AUC_{(2)} = 81.37\%$, compared to the pure classification with $AUC_{(1)} = 80.54\%$. As for our joint method, we achieved a significant gain both in segmentation and classification of ~ 3.5 pts ($meanDice_{(3)} = 38.28\%$) and ~ 2.5 pts ($AUC_{(3)} = 84.02\%$) respectively, compared to its sequential counterpart.

Table 1: Segmentation and classification performances on DDSM (Heath et al., 2000).

Training strategy	Cls baseline$_{(1)}$	Seg baseline$_{(1)}$	Seg-cls sequential$_{(2)}$	Seg-cls joint$_{(3)}$
Seg perf (mean Dice)	-	34.98	34.98	**38.28**
Cls perf (AUC)	80.54	-	81.37	**84.02**

Qualitative results Figure 2 illustrates classification and segmentation predictions for two images from DDSM, one normal and one cancer. In both cases, our joint method outperforms the sequential one for both classification and segmentation, and succeeds to capture lesions with highly precise localisation capability.

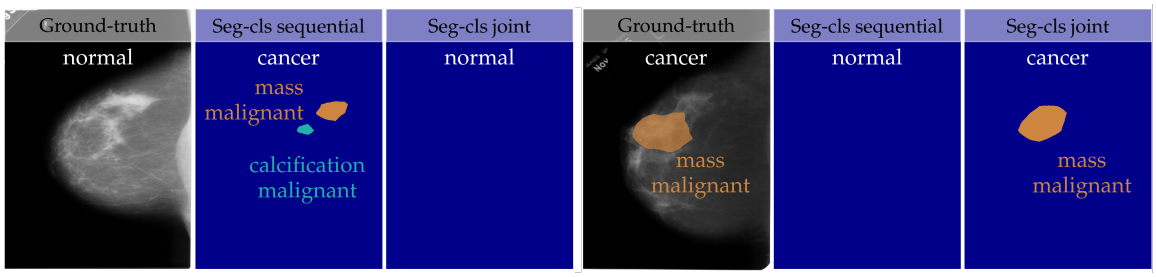

Figure 2: Segmentation and classification examples on DDSM (Heath et al., 2000). From left to right, the first three images show a normal case with ground-truth annotations and results from the segmentation-classification sequential and the segmentation-classification joint methods. The next three images show the same things for a cancer case.

Conclusion

In this work, we proposed a multitask learning scheme which combines segmentation and classification for cancer diagnosis in mammography. The achieved performances have shown the effectiveness of our method on both recognition tasks, lesion segmentation and cancer classification. Future works include investigating more powerful models and leveraging heterogeneous datasets to improve predictive performances.

References

Liang-Chieh Chen, George Papandreou, Iasonas Kokkinos, Kevin Murphy, and Alan L. Yuille. Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs. *IEEE Trans. Pattern Anal. Mach. Intell.*, 40(4): 834–848, 2018.

Thibaut Durand, Nicolas Thome, and Matthieu Cord. Mantra: Minimum maximum latent structural svm for image classification and ranking. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 2713–2721, 2015.

Krzysztof J Geras, Stacey Wolfson, S Kim, Linda Moy, and Kyunghyun Cho. High-resolution breast cancer screening with multi-view deep convolutional neural networks. *arXiv preprint arXiv:1703.07047*, 2017.

Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016.

M Heath, K Bowyer, D Kopans, R Moore, and P Kegelmeyer. The digital database for screening mammography. *Digital mammography*, pages 431–434, 2000.

Benjamin Q Huynh, Hui Li, and Maryellen L Giger. Digital mammographic tumor classification using transfer learning from deep convolutional neural networks. *Journal of Medical Imaging*, 3(3):034501, 2016.

Thijs Kooi, Geert Litjens, Bram van Ginneken, Albert Gubern-Mérida, Clara I Sánchez, Ritse Mann, Ard den Heeten, and Nico Karssemeijer. Large scale deep learning for computer aided detection of mammographic lesions. *Medical image analysis*, 35:303–312, 2017.

Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. Imagenet classification with deep convolutional neural networks. In *Advances in neural information processing systems*, pages 1097–1105, 2012.

Daniel Lévy and Arzav Jain. Breast mass classification from mammograms using deep convolutional neural networks. *arXiv preprint arXiv:1612.00542*, 2016.

Min Lin, Qiang Chen, and Shuicheng Yan. Network in network. *arXiv preprint arXiv:1312.4400*, 2013.

William Lotter, Greg Sorensen, and David Cox. A multi-scale cnn and curriculum learning strategy for mammogram classification. In *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support*, pages 169–177. Springer, 2017.

Inês C Moreira, Igor Amaral, Inês Domingues, António Cardoso, Maria João Cardoso, and Jaime S Cardoso. Inbreast: toward a full-field digital mammographic database. *Academic radiology*, 19(2):236–248, 2012.

Yaroslav Nikulin. DREAM Challenge Yaroslav Nikulin (Therapixel) write up.