DEPARTMENT OF COMPUTER SCIENCE

UNIVERSITY OF COPENHAGEN

Unsupervised Deep Learning Applied to Image Segmentation and Mammographic Risk Scoring

Kersten Petersen[1], Mads Nielsen[1,2], Pengfei Diao[1], Nico Karssemeijer[3], and Martin Lillholm[2]

[1]*Image Group, Department of Computer Science, University of Copenhagen, Denmark,* [2]*Biomediq A/S, Copenhagen, Denmark,* [3]*Department of Radiology, Radbound University Nijmegen, Netherlands*

Summary

Motivation:

Breast cancer mortality can be reduced by identifying high risk patients early and treating them adequately. Automated techniques are useful for objectively and efficiently analyzing and scoring large mammographic data. Deep neural networks have been widely used for medical image analysis in recent years and have shown numerous promising results. However, training deep neural networks, such as convolutional neural netwoks, is generally hard and slow for methods that require error being back-propagated through out the whole networks.

Idea: Semi-supervised convolutional neural networks

We propose an efficient semi-supervised method for training convolutional neural networks. Our method builds upon sparse auto-encoders and can be used to automate the breast tissue segmentation, density analysis, texture analysis, and hence the subsequent risk scoring.

Conclusion:

- We presented a novel sparsity regularizer that optimizes population and life time sparsity of the activations.
- The novel convolutional structure allows hierachical feature learning across multiple scales for capturing context.
- Our experimental results suggest that the proposed method is able to learn useful features for each of the considered applications.

Background

Canonical Auto-encoder: Three layer architecture for reconstruction

- *Encoder*: Mapping from input layer to hidden layer

$$f(x) = a(Wx + b) \tag{1}$$

- *Decoder*: Mapping from hidden layer to output layer

$$g(f(x)) = W^T f(x) + c \tag{2}$$

- *Objective*: Minimizing difference between true input and reconstructed input

$$J(\theta|X) = \sum_{i=1}^{N} L(x_i, g(f(x_i))) + \lambda \|W\|_F \tag{3}$$

where $\theta = \{W, b, c\}$ and $x_i \in X$

Sparse Auto-encoder: Over-complete and sparse representation

- *Idea*: In canonical auto-encoders, the number of hidden units is less than the number of input units (*under-complete*). Sparse auto-encoder learns an *over-complete* yet *sparse* representation of the input data by penalizing the overall activations being too high per sample or across the whole datasets.

- *Objective*: Minimizing difference between true input and reconstructed input, and inhibiting overall high activations.

$$J(\theta|X) = \sum_{i=1}^{N} L(x_i, g(f(x_i))) + \lambda \Omega_{sp}(A) \tag{4}$$

where $\theta = \{W, b, c\}$ and $x^n \in X$. The novel sparsity term defined as

$$\Omega_{sp}(A) = \Omega_{psp}(A) + \Omega_{lsp}(A) \tag{5}$$

combines population sparsity $\Omega_{psp}(A)$ and lifetime sparsity $\Omega_{lsp}(A)$ with respect to the activation matrix $A \in \mathbb{R}^{K \times N}$, where the entry $A_{ji} = f_j(x_i)$ denotes activation of j-th unit on the particular input x_i. The population sparsity term is defined as

$$\Omega_{psp}(A) = \frac{1}{K} \sum_{j=1}^{K} \tau(\hat{\rho}_j, \rho)^2 \tag{6}$$

where $\hat{\rho}_j$ defined as

$$\hat{\rho}_j = \frac{1}{N} \sum_{i=1}^{N} A_{ji} \tag{7}$$

denotes the averaged activation of j-th unit over all examples, and the threshold function given by

$$\tau(\hat{\rho}, \rho) = max(\hat{\rho} - \rho, 0) \tag{8}$$

penalizes sparsity values above the pre-specified sparsity parameter ρ to avoid non-specific features. The lifetime sparsity is similarly defined as

$$\Omega_{lsp}(A) = \frac{1}{N} \sum_{i=1}^{N} \tau(\hat{\rho}^{(i)}, \rho)^2 \tag{9}$$

where $\hat{\rho}^{(i)}$ defined as

$$\hat{\rho}^{(i)} = \frac{1}{K} \sum_{j=1}^{K} A_{ji} \tag{10}$$

denotes the averaged activation of all units on i-th example.

Training Convolutional Neural Networks: An unsupervised approach

- A sparse auto-encoder is constructed for each convolutional layer (Fig.1(a)).
- Training examples for auto-encoder are randomly sampled from every possible receptive field of those inputs fed which are fed to the corresponding convolutional layer.
- Input weights learned through auto-encoding is used as the shared weights of the corresponding convolutional layer (Fig.1(b)).
- The networks are trained layer-wisely from bottom to top.

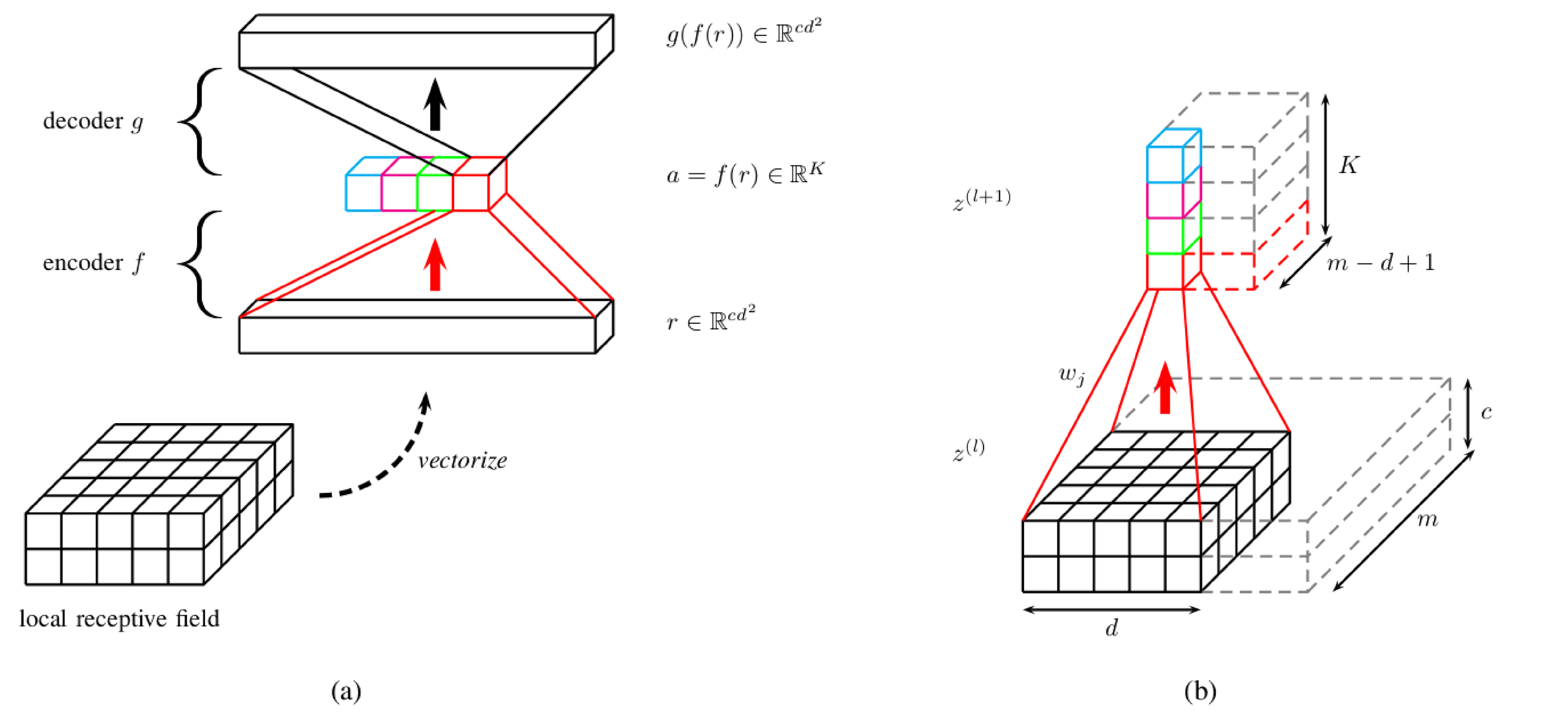

Figure 1: Treating shared weights as the input weights of an sparse auto-encoder

Classifier and multi-scale:

- Parallel convolutional networks are constructed for input sampled at different scales.
- All networks are merged into one common soft-max classifier on top.
- The classifier is trained in a supervised fashion together with the very last convolutional layers.

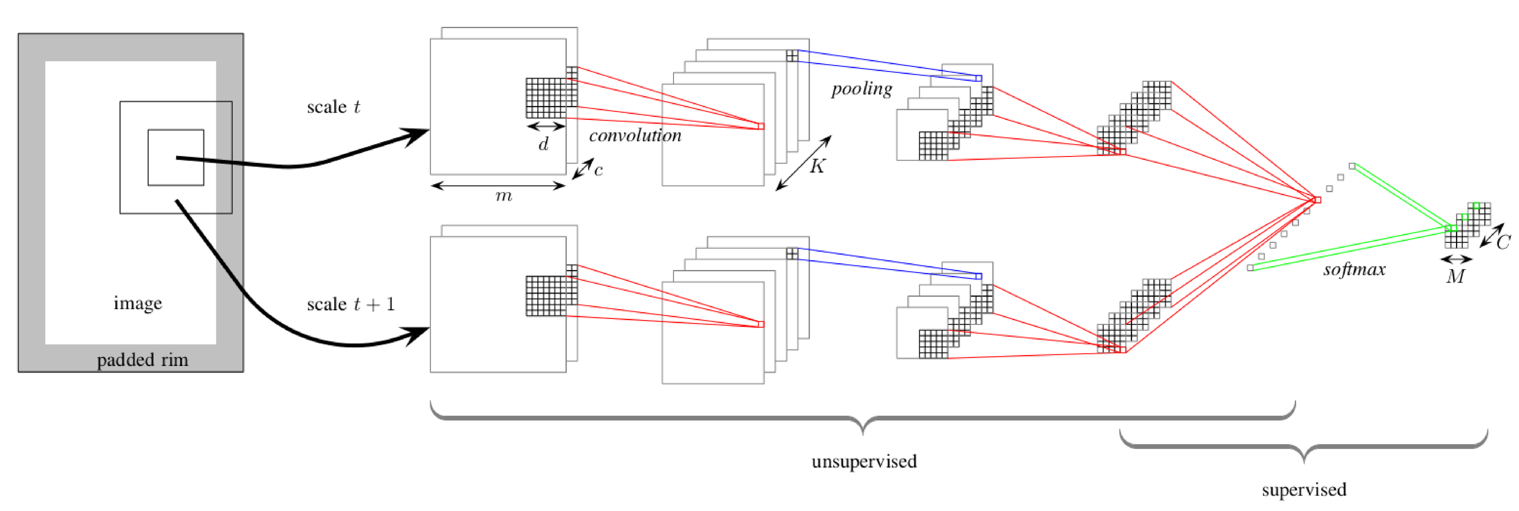

Figure 2: Multiscale convolutional neural networks

Experients

We used our method to perform segmentation, density and texture scoring on mammographic datasets. The neural networks consists of 6 convolutional layers and a soft-max classifier built on top.

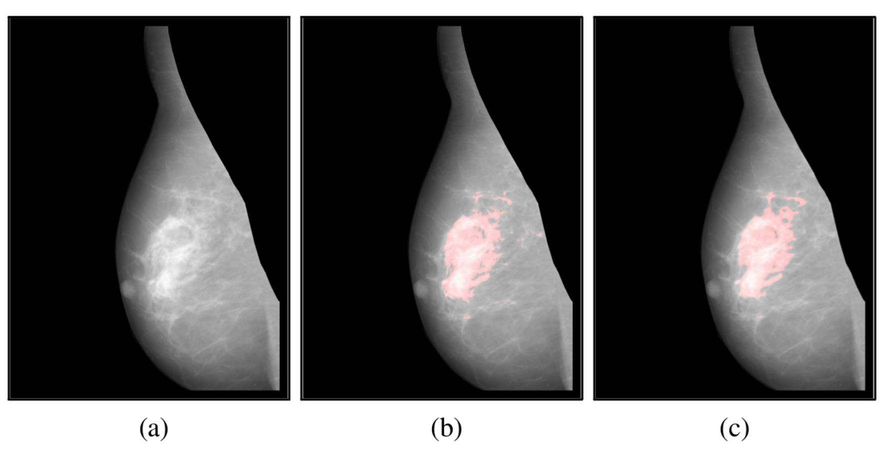

Figure 3: Densee tissue segmentation: (a)original image,(b)dense tissue (in red) according to human-expert, and (C)dense tissue classified by neural networks

Method	Case	Control	R_{PMD}	AUC (95% CI)
PMD	0.20 ± 0.13	0.18 ± 0.13	-	0.56 (0.51, 0.61)
BI-RADS	2.23 ± 0.72	2.10 ± 0.76	0.87	0.55 (0.50, 0.60)
$\text{PMD}_{\text{CSAE-}\sigma}$	0.19 ± 0.11	0.18 ± 0.13	0.69	0.54 (0.49, 0.59)
PMD_{CSAE}	0.21 ± 0.11	0.18 ± 0.12	0.87	0.57 (0.52, 0.62)

Table I: Comparison of automated with human-expert's scores. PMD - density score based on dense tissue segmented by human. BI-RADS - Radiologist score. $\text{PMD}_{\text{CSAE}-\text{œ}}$ - networks trained using canonical regularizer. PMD_{CSAE} - networks trained using novel sparsity term.

Method	R_{PMD}	AUC (95% CI)
KNN using 3-jet [16]	N/A	0.63
Static Histograms [15]	0.13	0.52 (0.47, 0.57)
$\text{MT}_{\text{CSAE-}\sigma}$	0.07	0.63 (0.58, 0.68)
MT_{CSAE}	0.11	0.65 (0.60, 0.70)

Table II: Comparison of automated texture scores. R_{PMD} - correlation with density score.