Taylor & Francis
Taylor & Francis Group

Computational mammography using deep neural networks

A. Dubrovina[a], P. Kisilev[b], B. Ginsburg[c], S. Hashoul[b,d] and R. Kimmel[a]

[a]Computer Science, Technion, Haifa, Israel; [b]IBM Haifa Research Lab, Haifa, Israel; [c]NVIDIA, Santa Clara, CA, USA; [d]Carmel Medical Center, Haifa, Israel

ABSTRACT

Automatic tissue classification from medical images is an important step in pathology detection and diagnosis. Here, we deal with mammography images and present a novel supervised deep learning-based framework for region classification into semantically coherent tissues. The proposed method uses Convolutional Neural Network (CNN) to learn discriminative features automatically. We overcome the difficulty involved in a medium-size database by training the CNN in an overlapping patch-wise manner. In order to accelerate the pixel-wise automatic class prediction, we use convolutional layers instead of the classical fully connected layers. This approach results in significantly faster computation, while preserving the classification accuracy. The proposed method was tested on annotated mammography images and demonstrates promising image segmentation and tissue classification results.

1. Introduction

The most common cancer in women is breast cancer. It is the second leading cause of death among women. Worldwide research efforts have been devoted to try and find a cure for this disease or any sort of early detection. Medical imaging of the breast by X-rays, also known as the mammography, is often used for diagnosis that leads to better treatment. Automatic classification of such images could play a key role in efficiently monitoring large populations. Tumours in different types of tissues are characterised differently, and require different treatment procedures. Therefore, automatic analysis of breast tissues as captured by mammograms, and accurate segmentation into the known classes is vital for early detection that could take part of the load from radiologists.

Deep Neural Networks (DNNs) have become popular in understanding natural images as well as medical ones. Recent papers demonstrate that in many recognition tasks, features that were automatically extracted by DNN outperform heuristically crafted descriptors, see, for example, Farabet et al. (2013), Sermanet et al. (2013). Deep learning methods applied to medical imaging provide state-of-the-art results, see, for example, Cireşan et al. (2013). In Petersen et al. (2014), the authors describe a problem and its DNN solution, which is closely related to the one we deal with in this paper. The authors learn descriptive features from *unlabeled* mammograms, and use them as an input to a simple classifier that segments the image into different type of tissues, and thereby estimates various characteristics. The framework we propose is different. Here, we present a novel supervised Convolutional Neural Network (CNN)-based method for breast tissue classification from mammogram images.

2. Problem formulation

Given a digital mammography image, we wish to associate each of its pixels with one of the four following classes: pectoral muscle, fibroglandular tissue, nipple, and the general breast tissue, which includes fatty tissue and skin. Our data-set consists of 40 digital mammograms of mediolateral oblique (MLO) view, manually segmented by an expert into the four regions. An illustration of a typical manual segmentation is given in Figure 1. While the images include a significant portion of background pixels, these pixels are easily detected in a pre-processing step, by thresholding image intensity values with zero threshold.

2.1. Breast tissue classification with DNN

The ability to process large data-sets allowed DNNs to produce state-of-the-art results when applied to computer vision, text understanding and speech recognition benchmarks. DNNs can provide discriminative image representations, sometimes referred to as features, by successive application of linear filters, non-linear activation functions, normalisation, and pooling operations, thus, avoiding the need to design such features manually.

The proposed DNN classifier, applied to raw image pixels, provides the probability of each pixel to belong to one of the four classes described above: $Pr(\text{class}(p) = k)$, $k = 0, 1, 2, 3$. The DNN is applied in a patch-wise manner: to classify the pixel p, the DNN is fed with a square image patch of size $w \times w$, centred at p. In our experiments, we set $w = 61$ pixels. The patches are pre-processed prior to training and classification, to have a zero mean, by subtracting from them the mean of all patches in the training set. The classification accuracy is acquired by a multinomial logistic loss function.

The motivation for using patch-wise classification is twofold. In medical imaging applications in general, data, manually annotated by experts, is rarely available. This is in contrast to general computer vision tasks, such as natural image classification, segmentation and object boundary detection, for which there exist data-sets with thousands and even millions of annotated
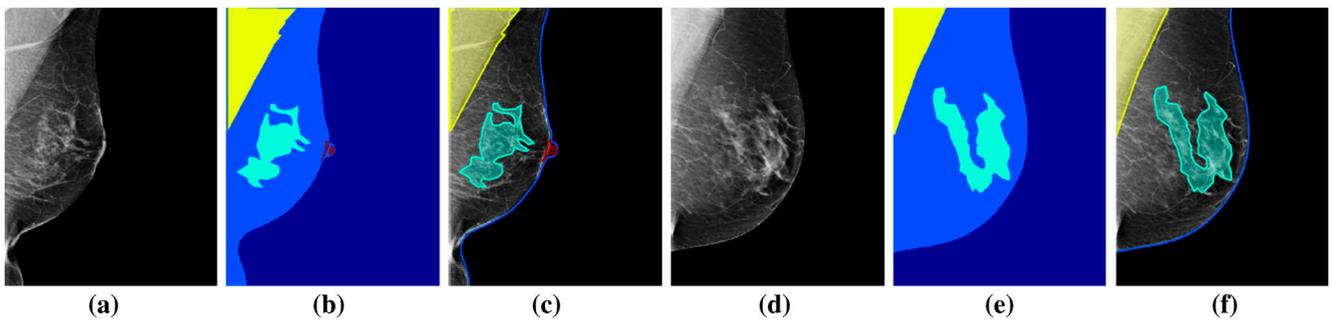
Figure 1. Manual segmentation. (a), (d) Original mammography images. (b), (e) Manual labelling into the pectoral muscle (yellow), fibroglandular tissue (cyan), nipple (bordo), breast tissue (light blue), and background (dark blue). (c), (f) Manual labelling superimposed over the mammography image.

Table 1. Initial deep neural network architecture.

Layer	1	2	3	4	5	6	7 – Output
Stage	conv+relu+max	conv+relu+max	conv+relu+max	dropout	full+relu	full+relu	full
# Channels	16	16	16	16	128	16	4
Filter size	7×7	5×5	5×5	–	–	–	–
Pooling size	3×3	3×3	3×3	–	–	–	–
Pooling stride	2	2	2	–	–	–	–
Dropout factor	–	–	–	0.5	–	–	–
Spatial input size	61×61	27×27	11×11	3×3	3×3	1×1	1×1

Notes: The first three stages are comprised of convolutional layers, followed by ReLU and max pooling layers. Dropout layer with dropout factor of 0.5 is placed before the first fully connected layer, which is followed by two additional fully connected layers. The last fully connected layer acts as a pixel classifier.

Table 2. Conversion of fully connected layers into convolutional layers, for fast full image pixel classification.

Layer	5	6	7		5	6	7
Stage	full+relu		full			conv+relu	conv
# Channels	128	16	4		128	16	4
Filter size	–	–	–	$\Rightarrow$	3×3	$128 \times 1 \times 1$	$16 \times 1 \times 1$
Pooling size	–	–	–		–	–	–
Pooling stride	–	–	–		–	–	–
Dropout factor	–	–	–		–	–	–
Spatial input size	3×3	1×1	1×1		Depends on the network input size		
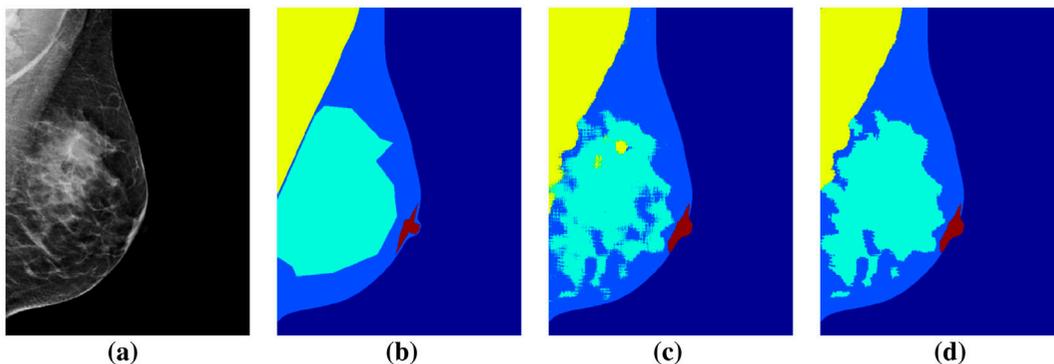
Figure 2. DNN output post-processing example. (a) Original image, (b) manual segmentation, (c) DNN output, (d) post-processed DNN output. Colour coding as in Figure 1.

examples, see for example Russakovsky et al. (2015). Since a typical DNN has between tens of thousands to millions of parameters, it requires a large amount of manually annotated examples to properly train all the parameters and avoid overfitting to an insufficient training data. Therefore, by working with separate, possibly overlapping image patches, we can obtain a training set large enough to overcome these limitations – in our experiments, we used training sets of approximately $8 \cdot 10^5$ training

examples. In addition, we expect that large enough patches would capture a sufficient part of the local information required to correctly classify pixels belonging to different breast tissues.

By applying the DNN to separate image patches, we loose the spatial dependency between neighbouring patch labels, as well as information about the spatial pixel location. Clearly, $Pr(\text{class}(p))$ and $Pr(\text{class}(q))$ are dependent for neighbouring patches p and q. Although, using overlapping patches implicitly

imposes some smoothness in the output label space, it is not trivial to add such information to the proposed patch-based classifier. Alternative solutions include switching to recurrent neural networks, or changing the network, so that it would act on larger image patches and produce an image of class probabilities, as opposed to a single class probability vector produced by the proposed network. The latter would allow to condition the class probabilities of neighbouring pixels, and improve the classification results. Such architectures were successfully deployed to perform dense natural image labelling in Farabet et al. (2013), Long et al. (2015), Chen et al. (2014). We plan to explore this line of thought in the future research.

In contrast, the spatial pixel location information is easier to include into the proposed network, for instance, in the form of x and y pixel coordinates, given that the images are identically aligned, similarly to Bar et al. (2015). To normalise the location information across images of breasts of different sizes, the x pixel coordinate is divided by the x coordinate of the rightmost image pixel which does not belong to the background. The y coordinate is normalised to the range [0,1]. The coordinates can then serve as an additional input for the proposed network, and added after the first fully connected layer by concatenating them to the first fully connected layer output, and then transferring the results to the next fully connected layer. While this simple approach allows us to incorporate the spatial information in our learning scheme, it produces equivalent or inferior results than only intensity-based classification – see the results in Table 3. Thus, more research is required to devise a correct way to use the spatial information.

2.2. DNN architecture

The architecture of the proposed network is summarised in Table 1. It consists of three stages of convolutional layers, ReLU (rectified linear unit) activation layers, and max pooling layers, followed by three fully connected layers. To prevent overfitting, a dropout layer (Srivastava et al. 2014) with dropout factor of 0.5 was added between the convolutional and the fully connected layers. The image intensity and the normalised coordinate information can be combined after the first fully connected layer in the following manner: the two normalised patch centre coordinates and the 128-dimensional fifth layer output are concatenated into a 130-dimensional vector, passed to the second fully connected layer.

2.3. Fast full image pixel classification

During the classification stage, the network must be applied separately to all the overlapping image patches. This introduces a significant computation overhead, since both the convolutional layers, and the fully connected layers are applied multiple times to overlapping regions. Inspired by Sermanet et al. (2013) and Long et al. (2015), we converted the proposed classification network into a fully convolutional network. That is, we converted the fully connected layers #5, 6, 7, into convolutional layers, as shown in Table 2. The new network is able to output dense predictions for input images of arbitrary sizes. Specifically, for the 829×640 images we used in our experiments, the new network output was 97×73. To obtain dense prediction for the

whole image, we adopt the shift-and-stitch method of Sermanet et al. (2013).

In the proposed network, the outputs were downsampled by a factor of 8 with respect to inputs. Hence, by feeding the new network shifted versions of the input, by $i \in \{0, 1, \ldots, 7\}$ pixels right and $j \in \{0, 1, \ldots, 7\}$ pixels down, and interlacing the obtained 64 output images, we obtain a dense prediction for the whole image. The classification time of the new network is approximately 1.8 seconds as opposed to 114 seconds for the per-patch neural network application. Conversion to the fully connected network requires the following adjustment. Previously, the mean of all training examples was subtracted from the network input during the training and the classification. Now, a single value is subtracted from the training examples at the training stage, and from the entire image at the classification stage, allowing a full-image classification. In our experiments, this change had a minor effect on the classification accuracy. We used a single mean intensity value of the pixels in the mean image, computed over all the training examples.

2.4. Classifier output post-processing

The raw DNN classifier output obtained for one of the images in our data-set is shown in Figure 2(c). Since the proposed patch-based classifier cannot incorporate constraints in relative spatial locations of different tissues, it may produce fragmented regions, as shown in Figure 2(c). Therefore, during the post-processing step, as dictated by the physiological breast structure, (i) the interior of the large pectoral muscle region adjacent to origin of the image is filled with its corresponding label, while small unconnected components of the pectoral muscle region are given the label of the general breast tissue; (ii) the outer boundary of the fibrogladular tissue region is morpholocally filled to contain the fibrogladular tissue label, and its connected components smaller than some predefined threshold, are removed; (iii) a single connected component of the nipple region, closest to the centre of the image, is retained.

3. Segmentation results

3.1. Pre-processing

In our experiments, we used a data-set with 40 manually segmented MLO views. All images were aligned so that the pectoral muscle appeared on the left side of the image, for spatial consistency. A leave-one-subject-out cross validation procedure was used. We considered different 40 image sets with 39 training images, and the remaining image was used to evaluate the classification performance. The results presented below were averaged over all 40 possible training and test image combinations. Approximately 800,000 patches were extracted to form the training set, containing an equal number of patches centred at pixels belonging to the four different regions.

3.2. Network training

The DNN was trained by stochastic gradient descend with momentum. We used minibatches of 256 image patches, and learning rate of 10^{-3}, reduced by a factor of 10 twice, after 30,000 and 60,000 training iterations (approximately 10 and 20 epochs).
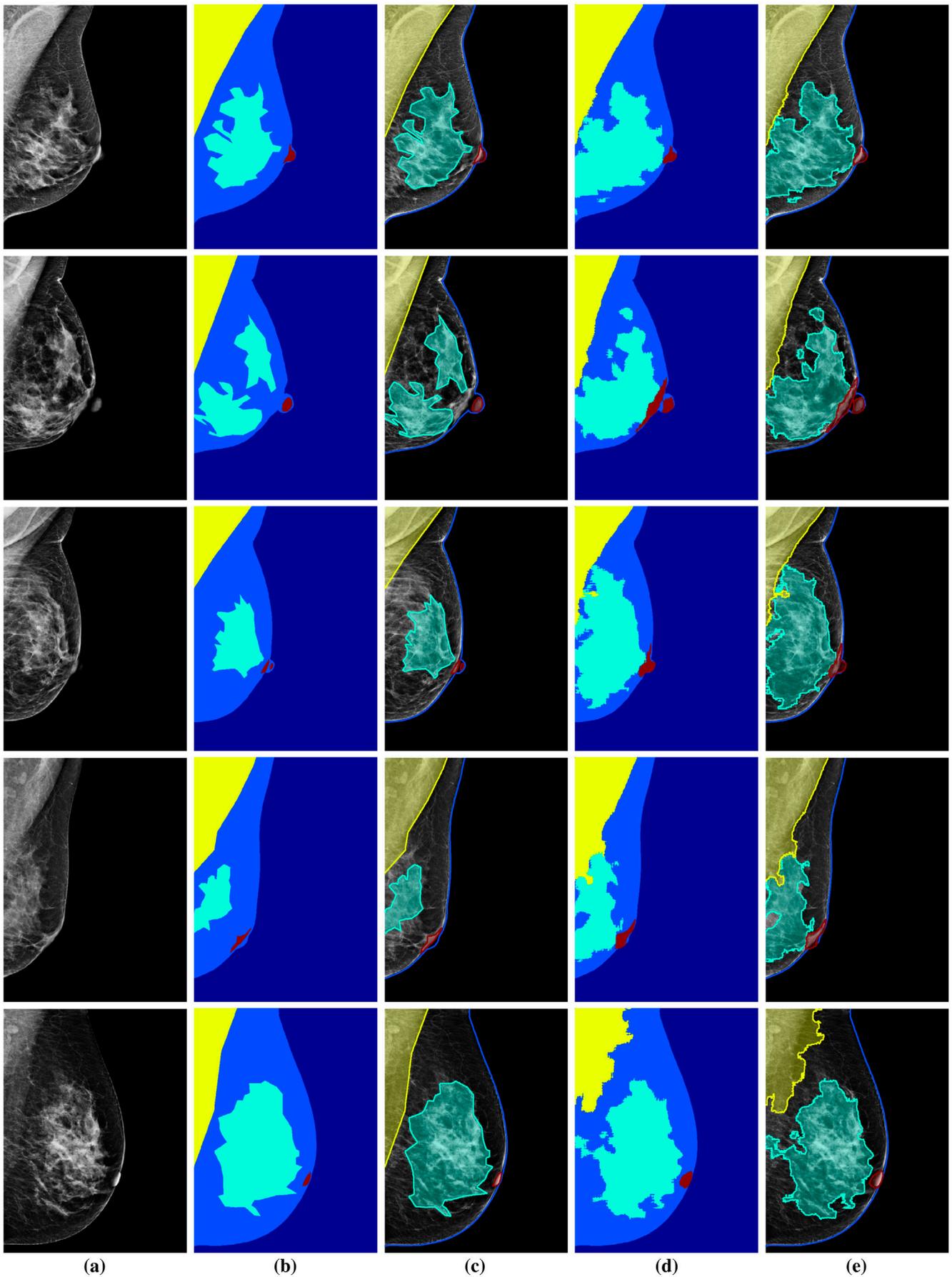
Figure 3. Segmentation results. (a) Original image. (b), (c) Manual labels. (d), (e) Labels obtained using the proposed method. Colour coding is as in Figure 1.

Table 3. Average Dice coefficients obtained using the proposed methods for different breast tissues, and average coefficients over the four tissues.

	Pectoral muscle	Fibroglandular tissue	Nipple	Breast tissue (fat, skin)	Average over four tissues
Raw DNN output	0.78	0.60	0.47	0.80	0.66
Post-processed DNN output	0.85	0.61	0.56	0.81	0.71
Raw DNN output, with (x, y) coordinates	0.78	0.60	0.56	0.77	0.68
Post-processed, with (x, y) coordinates	0.79	0.61	0.57	0.77	0.69

We used momentum 0.9, weight decay $5 \cdot 10^{-4}$, and initialised the net weights randomly, using normal distribution with zero mean and 0.01 and 0.1 variances for the convolutional and the fully connected layers, respectively. During testing, the dropout layer was removed, and the outputs of the layer #3 were multiplied by a factor of 0.5.

3.3. Quantitative evaluation

The quality of the proposed classification method was evaluated using the Dice coefficient (DC), which denotes the volume overlap between two sets of pixels A and B

$$DC = \frac{2|A \cap B|}{|A| + |B|}.$$

Dice coefficients were computed for each of the four regions, and averaged over the 40 test images. The results are presented in Table 3. The pectoral muscle has the highest DC, of 0.85, the fibrogladular tissue and the nipple have DC equal to 0.61 and 0.56, respectively. Adding spatial information in the form of normalised patch centre coordinates does not improve the segmentation accuracy, contrary to our expectations. The latter is probably due to large variations in the breast spatial structure and size, and the correct way to use the spatial information is not yet clear.

For comparison, in a related work, Oliver et al. (2014) presented an algorithm for automatic segmentation of digitised mammography images into breast, pectoral muscle and background. They reported a DC of 0.83 for the pectoral muscle region, obtained using intensity, texture and atlas (spatial) information. Without the atlas information, best pectoral muscle DC reported in Oliver et al. (2014) was 0.54 – significantly lower than the result obtained using the proposed method.

The segmentation results obtained using the proposed method are presented in Figure 3. We notice that the pectoral muscle is detected with relatively high accuracy. The fibrogladular tissue and the nipple are detected with a lower accuracy. It could be explained by the high appearance variability across different regions. In our future research, we will explore how the classification can be improved using a larger annotated data-set, together with breast structure priors and smart data normalisation.

4. Conclusions

We proposed a new deep learning-based framework for tissue classification with application to the problem of mammogram image segmentation. The proposed method deploys CNN to learn discriminative features automatically during the classifier training. We train the CNN using the patch-wise approach; this

ensures sufficient number of training examples. To speed up pixel-wise class prediction, we use convolutional layers instead of the fully connected ones. This approach yields nearly two orders of magnitude faster computation, while maintaining the same classification accuracy. In the future, we plan on integrating spatial smoothness into the classification process, and further improve the classification rates. To that end, we plan to extend the proposed framework and explore the use of recurrent neural networks instead of the fully convolutional set-up.

References

Bar Y, Diamant I, Wolf L, Greenspan H. 2015. Chest pathology detection using deep learning with non-medical training. In: IEEE International Symposium on Biomedical Imaging (ISBI). New York (NY).

Chen LC, Papandreou G, Kokkinos I, Murphy K, Yuille AL. 2014. Semantic image segmentation with deep convolutional nets and fully connected CRFs. arXiv preprint arXiv:14127062.

Cireşan DC, Giusti A, Gambardella LM, Schmidhuber J. 2013. Mitosis detection in breast cancer histology images with deep neural networks. In: Medical Image Computing and Computer-Assisted Intervention (MICCAI); p. 411–418. Nagoya, Japan.

Farabet C, Couprie C, Najman L, LeCun Y. 2013. Learning hierarchical features for scene labeling. IEEE Trans Pattern Anal Mach Intell. 35:1915–1929.

Long J, Shelhamer E, Darrell T. 2015. Fully convolutional networks for semantic segmentation. IEEE computer society conference on computer vision and pattern recognition (CVPR), Boston, MA; p. 3431–3440.

Oliver A, Llado X, Torrent A, Marti J. 2014. One-shot segmentation of breast, pectoral muscle, and background in digitised mammograms. In: 2014 IEEE International Conference on Image Processing (ICIP), Paris; p. 912–916.

Petersen K, Nielsen M, DiaoP, Karssemeijer N, Lillholm M. 2014. Breast tissue segmentation and mammographic risk scoring using deep learning. In: Breast imaging, 12th International Workshop, IWDM 2014, Gifu City, Japan, June 29–July 2. Springer; p. 88–94

Russakovsky O, Deng J, Su H, Krause J, Satheesh S, Ma S, Huang Z, Karpathy A, Khosla A, Bernstein M, et al. 2015. ImageNet large scale visual recognition challenge. Int J Comput Vision. 115:211–252

Sermanet P, Eigen D, Zhang X, Mathieu M, Fergus R, LeCun Y. 2013. Overfeat: integrated recognition, localization and detection using convolutional networks. arXiv preprint arXiv:13126229.

Srivastava N, Hinton G, Krizhevsky A, Sutskever I, Salakhutdinov R. 2014. Dropout: a simple way to prevent neural networks from overfitting. J Mach Learn Res. 15:1929–1958.