2021 18th International Conference on Electrical Engineering, Computing Science and Automatic Control (CCE). Mexico City, Mexico. November 10-12, 2021

Semantic Segmentation of Mammograms Using Pre-Trained Deep Neural Networks

1st Rodrigo Leite Prates
Programa de Engenharia Biomédica
Universidade Federal do Rio de Janeiro
Rio de Janeiro, Brazil
Email: prates@peb.ufrj.br

2nd Wilfrido Gómez-Flores
Cinvestav Tamaulipas
Ciudad Victoria, Mexico
Email: wgomez@cinvestav.mx

3rd Wagner Pereira
Programa de Engenharia Biomédica
Universidade Federal do Rio de Janeiro
Rio de Janeiro, Brazil
Email: wagner@peb.ufrj.br

***Abstract*—Anatomic regions like breast and pectoral muscle are common regions that need to be segmented before determining abnormalities in the mammographic image. For this task, we explore five convolutional neural network models, namely, Resnet50-Unet, Mobilenet-Unet, Vgg-Unet, Unet, and Segnet. A classic technique that uses an MLP network is compared with the convolutional methods. The MIAS and INbreast datasets are used for evaluating Deep Neural Networks on the segmentation of these regions. Several evaluation metrics are used to compare the results of convolutional approach and a traditional method based on hand-crafted features. The results confirm the superiority of the deep networks over the traditional method. The best Jaccard value for Deep Learning algorithms is 0.945 ± 0.068. The Mobilenet-Unet network presents the best trade-off between computational cost and segmentation capacity; therefore, it is the best choice to compose a CAD system.**

***Index Terms*—Mammography, Breast, Pectoral Muscle, Convolutional neural networks, Semantic segmentation, Transfer learning**

I. Introduction

Breast cancer is a disease caused by the uncontrolled multiplication of abnormal cells that form tumors. It is considered a worrying disease because of the possibility of a fatal outcome involving personal, family, and social consequences [1]. To detect breast cancer early, the recommendation of the World Health Organization is to perform mammographic screening. Mammography is an X-ray of the breasts performed by equipment capable of identifying suspicious changes in the normal tissues before the onset of symptoms or any change in the breasts is palpated. The mammogram interpretation can be a challenging task, as the pattern of normality is variable, and the signs of an illness can be subtle, making the analysis susceptible to errors [1]. Computerized mammographic analysis systems have been developed to distinguish between malignant and benign lesions, increase the sensitivity and specificity of the diagnosis [2], indicate suspicious areas and non-palpable abnormalities, and improve the accuracy of radiologists with the reduction of time in the interpretation of images. These systems are also important since they allow a reduction in the number of biopsies performed unnecessarily in benign tumors [3]. In general, these kinds of systems are called computer-aided diagnosis (CAD) systems. They are divided into two groups, those that perform detection of possible anomalous regions (CADe), and those that perform classification (CADx), for example, the distinction between benign and malignant lesions.

Traditional CAD systems use pattern recognition techniques that incorporate prior domain knowledge into hand-crafted features that are specifically tailored to describe suspect structures in the breast [4]. These techniques are considered limited since their performance is lower when more clinical exams are analyzed [4].

An important step is the mammogram segmentation in representative anatomical regions, such as the breast and the pectoral muscle [6]. The remotion of both the pectoral muscle and the image background is expected in the pre-processing of images from mammography software [7]. It is useful in many mammographic analysis areas since the pectoral muscle has similar texture characteristics to the breast parenchyma (as occurs in breasts with excessive glandular tissue) and background due to the presence of artifacts like labels and wedges [5], altering the results of image processing methods in the automatic detection of anomalous regions [2].

For improving and automatizing this process, Deep Learning techniques and specifically deep neural networks can perform semantic segmentation. These neural structures can automatically learn subtle characteristics with a high abstraction level [6], [10].

The semantic segmentation of mammographic regions can be addressed from three different approaches [11]:

- Traditional approach: A classifier (e.g., Multilayer Per-

978-1-6654-0029-9/21/$31.00 ©2021 IEEE

ceptron, MLP) receives expert-defined hand-crafted features.
- Convolutional approach: A Convolutional Neural Network (CNN) automatically learns features through stacked convolution and pooling layers from raw images.

Some CNN architectures like Unet and Segnet are commonly used for segmenting biomedical images [7]. There are different implementations of both networks, and each variation can present different results depending on the application.

The objective of this work is to improve the performance of CAD tools with the application of deep learning techniques in the segmentation of mammographic images.

The novelty of our approach lies in the exploration of different versions of a well-known deep network, Unet, by changing its backbone, which as far as the authors have been able to verify, it has not yet been done.

Therefore, in this paper, we evaluate five CNN architectures to segment anatomical regions in mammographies, including four variations of Unet and one for Segnet. These deep networks are compared with a traditional approach based on an MLP network. Two public datasets are combined for the evaluation of each model. These pre-trained networks are available to the scientific community to carry out transfer learning in new breast segmentation problems.

The rest of this paper is organized as follows. Section 2 reviews recent CNN-based approaches for mammography segmentation. Section 3 describes the mammography datasets used in this study, the evaluated CNN architectures, the training procedure, the segmentation performance assessment, and a brief description of the MLP-based method, which will be compared with the CNN-based methods. The experimental results and discussions are presented in Section 4. Finally, Section 5 gives some concluding remarks.

II. Literature review

The section reviews the state-of-the-art of classical and CNN-based methods for the segmentation of mammographic images.

Carvalho et al. [2] proposed an automatic segmentation method based on morphological operators to detect and outline the pectoral muscle contour in mammograms. From 264 images of the Mammographic Image Analysis Society (MIAS) dataset, 0.917 of them resulted in overlap ratio (RS) greater than or equal to 0.5, and the mean values of False Positive (FP) and False Negative (FN) were 0.043 and 0.109, respectively. High false negative values were found for almost half of the images.

The use of an Unet network to segment both the breast region and the pectoral muscle of digital mammograms (DM) and breast tomosynthesis (3D mammo) (DBT) was proposed by Wang et al. [6]. A total of 2825 raw images from different X-ray manufacturers were used: 2671 DM and 154 DBT. From 825 test images, the best results was a dice-similarity coefficient of 0.887 and 0.991 for pectoral and breast segmentation, respectively. It was found that the technique performed poorly in input images with noise and low lighting.

Oliveira et al. [7] investigated the Fully Convolutional Networks (FCN), Unet, and Segnet to segment mammography images into three regions: breast, pectoral muscle, and background. A total of 522 images were used, 322 from MIAS and 200 from INBreast datasets, where for the latter, only the pectoral muscle was segmented. Regarding the MIAS dataset, the best Jaccard values were 0.897, 0.984, and 0.970 for pectoral muscle, background, and breast region segmentation, respectively, whereas for the INbreast datasets, a Jaccard value of 0.908 was obtained.

Rampun et al. [12] proposed an automatic contour-based approach to segment pectoral muscle in MLO mammograms using a holistically-nested Edge Detection (HED) CNN architecture. Results were obtained for four different datasets: MIAS, INbreast, Breast Cancer Digital Repository (BCDR), and Curated Breast Imaging Subset of Digital Database for Screening Mammography (CBIS-DDSM). The best values were 0.948 and 0.975 for the Jaccard and Dice indices, respectively. The main limitation is the downsampling of the original image size, which may affect the representation and accuracy of the pectoral muscle boundary.

Considering the approaches and limitations in all works reviewed, we propose to apply different Unet architectures and a Segnet to the problem of anatomical region segmentation such as breast and pectoral muscle using MIAS and INBreast mammograms.

III. Materials and methods

A. Mammography dataset

The mammography images used in this study are from the following public datasets:

- The MIAS dataset [13] is formed by 322 scanned mediolateral oblique (MLO) images of 1024×1024 pixels, which the UK National Breast Screening Programme distributes[1]. The dataset has annotations of the central coordinates of anomaly regions, such as calcifications, well-defined circumscribed masses, and spiculated masses. Also, biopsy results (benign and malignant classes) and density types (fatty, fatty-glandular, and dense-glandular) are provided.
- The INbreast [14] dataset is composed by 410 digital images of 3328×4084 and 2560×3328 pixels made available by the Sao João Hospital, Breast Center, Porto, Portugal[2]. The dataset contains 201 MLO images and 209 bilateral craniocaudal (CC) images. ROIs of abnormalities (masses, calcifications, asymmetries, and distortions), pectoral muscle coordinates, and malignancy classification in BIRADS format are also included. In this work, only MLO images were used.

B. Deep neural network architectures

Five state-of-the-art approaches for building semantic segmentation models are used in this study, namely: Mobilenet-

[1] http://peipa.essex.ac.uk/info/mias.html

[2] http://medicalresearch.inescporto.pt/breastresearch/index.php

Unet, Resnet50-Unet, Vgg-Unet, Vanilla-Unet(Unet), Segnet. The characteristics of these CNNs are shown in Table I.

Unet was proposed by Ronneberger et al. [15] to segment biomedical images, where training data are commonly scarce. An encoder network followed by a decoder network gives its U-shaped architecture. The spatial information is reduced by convolution/max-pooling layers in the encoder path, while feature information increases. In the decoder path, feature maps and spatial information are combined by a sequence of up-convolutions and concatenations with high-resolution features from the decoder path. Here, four different implementations of Unet were used. We used the following classification networks for the encoder: Resnet50, Mobilenetv1, Vgg-16, and Vanilla Encoder.

He et al. [16] proposed a residual learning framework to ease the training of networks that are substantially deeper than those used previously. This architecture uses several stacked convolutional layers with skip connections (double- or triple-layer skips) with few extra parameters. It is expected that this network architecture could achieve good classification accuracy.

Howard et al. [17] presented Mobilenet models for mobile and embedded vision applications. They are based on a streamlined architecture that uses depthwise separable convolutions to build lightweight deep neural networks. This network architecture can be very deep with few trainable parameters. It is expected that this model can achieve high precision with a low computational cost.

Simonyan et al. [18] introduced the Vgg network that uses convolution layers with 3×3 filters with a stride of 1 and a max-pooling layer of 2×2 with a stride of 2. This arrangement is consistently followed throughout the whole architecture.

These 3 classifiers were selected in order to assess the impact of variables such as depth and number of parameters on the performance of segmentation networks.

Segnet is an encoder-decoder architecture based on the Vgg family proposed by Badrinarayanan et al. [19]. The encoder network consists of several convolution/max-pooling layers used to produce low-level feature maps. The decoder network upsamples low-level feature maps using the max-pooling indices from the corresponding encoder feature maps. Here, Vgg16 is used in the Segnet implementation.

The input layer of CNN models requires a 3-channel image because they were pre-trained using color images from the ImageNet database [20]. The output layer is a Softmax layer used to obtain each pixel probabilities belonging to a class. Considering zero padding, batch normalization, concatenation, convolutions, and up-sampling operations, we have a total of 82 layers for Mobilenet encoder, 144 for Resnet50 encoder, 15 for Vgg encoder and 21 layers for Unet decoder, whose total is shown in Table I. All models are available on the internet [21].

TABLE I
SEMANTIC SEGMENTATION ARCHITECTURES BASED ON DEEP NEURAL NETWORKS.

CNN	Number of Layers	Number of Parameters
Mobilenet-Unet	103	6,316,099
Resnet50-Unet	165	16,374,019
Vgg-Unet	36	12,323,523
Unet	42	4,472,323
Segnet	39	3,698,179

C. Deep neural network training

Initially, image data are normalized by adjusting the distribution of pixel values in input images to approximate a normal distribution with zero mean and unit variance. In addition, in all images the pectoral muscle and the breast are aligned to the left.

The CNN models are trained using a stochastic gradient descent method based on per dimension adaptive learning rate (Adadelta) with a learning rate of 0.001 for 60 to 100 epochs. The number of epochs is controlled by a parameter that informs how many epochs the model can continue to be trained without improving the validation error. This parameter was adjusted empirically to 15. As a cost function, cross-entropy is initially used. The model with the lowest validation loss value is selected as the final model.

The computing platform consisted of an Intel i7 processor with eight cores at 3.50 GHz, 16 GB of RAM, and a graphic card NVIDIA GeForce GTX 960M with 4 GB of VRAM. All programs were developed in Python 3.7 with Tensorflow and Keras.

D. Performance metrics

The segmentation performance is evaluated by comparing the similarity between the ground-truth outlined by the specialists and the CNN outcome. The following segmentation performance metrics are calculated:

- Dice similarity coefficient (Dic). It is also known as F_1-Score
- Jaccard similarity coefficient (Jac). It is also known as Intersection over Union (IoU).
- False Positive Rate (FPR).
- False Negative Rate (FNR).

Dice and Jaccard are two indices that quantify the overlap ratio between the ground-truth mask and the segmented image. Both indices return values in the range $[0, 1]$, where values toward unity indicate an adequate segmentation performance. FPR and FNR indicate the rate of incorrectly classified pixels, where small values close to zero are the best.

E. Segmentation performance assessment

The generalization capability of CNN models is measured using independent training-test experiments based on the k-fold cross-validation method [4]. For the combination of both datasets, it is randomly divided into five disjoint groups. In every experiment, one group is the test set, and the remaining four groups comprise the training set. Each training set is randomly divided into 70% for learning and 30% for validation. Performance metrics are calculated on the test set.

Performance indices are calculated as a weighted average of all three classes (breast, pectoral muscle, and background). Statistic differences between segmentation methods are determined by one-way analysis of variance (ANOVA) test ($\alpha = 0.05$), where the null hypothesis is that the means of the groups are equal, assuming that the groups come from the same distribution [22]. Because ANOVA does not specify which groups were statistically significantly different from each other, a second test is necessary. The Tukey test is used to compare the means of two groups, assuming that data groups are independent, approximately normally distributed, and have a similar variance within each group being compared.

F. MLP-based method

In this paper, CNN-based methods are compared with a traditional MLP-based method that uses pixel intensity and texture characteristics, as described in [8]. This technique was originally applied for semantic segmentation of breast tumors on ultrasound images (BUS). This network is trained based on hand-crafted features extracted from texture, intensity and spatial information. The steps below summarize the approach:

1) Image pre-processing for noise reduction and contrast enhancement. For the first task, a low-pass Gaussian filter is used, and for the second task, the dynamic range of the image is stretched.
2) Hand-Crafted features extraction where techniques for texture description are used to generate 240 features for each pixel in the BUS images. In addition to texture features, intensity and spatial adjacency information are included totaling 243 features. Dimensionality reduction techniques are applied to discard irrelevant features.
3) MLP network training where the number of input neurons is defined by the number of features after dimensionality reduction process and the number of hidden neurons is given by the approximation order concept from [9]. The activation function in the output layer is a Softmax function and the stochastic gradient descent performs the MLP weight updating to minimize the cross-entropy loss function.
4) Image post-processing techniques like hole- filling and active contours are applied to reduce the number of false positives and to refine the segmentation.

IV. Results and Discussion

The segmentation results of the CNNs architectures are presented in this section. For convenience, our discussions are held in this section to help interpret the results.

Figure 1 show a comparison between segmentation quality and time required for training a CNN using our computing platform. As expected, the network with more weight parameters or layers will have a longer training time. The only exception is the Mobilenet-Unet network, which has the shortest training time despite being the second model with more layers. This effect is due to the use of depthwise separable convolution, as explained in Section III-B. The Vgg-Unet training time was the longest in almost all scenarios.

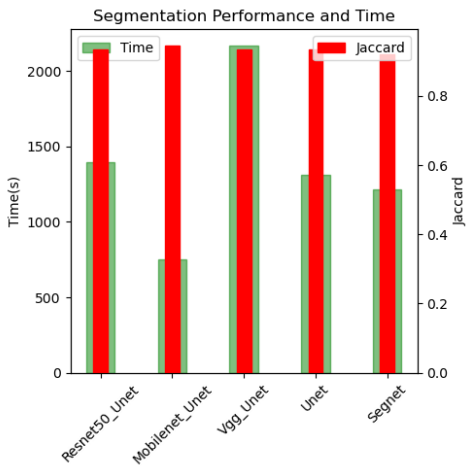

Fig. 1. Comparison of CNN models in terms of segmentation performance and time to train a CNN.

TABLE II
Overall segmentation performance of CNN architectures.

Models	Jac	Dic	FPR	FNR
Resnet50-Unet	0.933 ± 0.079	0.959 ± 0.066	0.010 ± 0.014	0.038 ± 0.061
Mobilenet-Unet	$\mathbf{0.945 \pm 0.073}$	$\mathbf{0.966 \pm 0.061}$	$\mathbf{0.007 \pm 0.012}$	$\mathbf{0.028 \pm 0.050}$
Vgg-Unet	0.935 ± 0.079	0.960 ± 0.063	$\mathbf{0.007 \pm 0.012}$	0.031 ± 0.050
Unet	0.933 ± 0.086	0.958 ± 0.072	0.008 ± 0.016	0.038 ± 0.061
Segnet	0.921 ± 0.089	0.952 ± 0.071	0.010 ± 0.016	0.037 ± 0.059

Table II shows the overall segmentation performance of CNN architectures calculated on the test set. Based on values from Jaccard and Dice metrics, all models obtained values greater than 0.9. In general, the Unet models showed slightly better values in most metrics, with the Mobilenet-Unet architecture being the one with the highest value on all metrics. This result confirms that network depth and the number of weight parameters are relevant to achieve better segmentation quality because more subtle features are extracted as the network depth grows.

Regarding the results obtained by the Mobilenet-Unet network and those presented in the literature, a direct comparison is not possible since the results obtained in this study are on the combined MIAS and INbreast datasets. Even so, the metrics obtained here had values similar to those obtained by Carvalho et al. [2] and Oliveira et al. [7].

The statistical differences between CNN architectures were determined by the Tukey test, as shown in Table III. The statistical analysis revealed that the networks pairs Resnet50-Unet and Mobilenet-Unet, Resnet50-Unet and Vgg-Unet, Resnet50-Unet and Unet, Mobilenet-Unet and Vgg-Unet, Mobilenet-Unet and Unet, Vgg-Unet and Unet, Unet and Segnet did not present statistically significant differences between them. All Unet architectures except Unet itself were better than Segnet. In general, it is observed that these networks obtained the highest performance parameter values.

A Multilayer Perceptron (MLP) and Mobilenet-Unet networks were selected to compare a CNN-based model with

TABLE III
PAIRWISE COMPARISONS BY TUKEY TEST. THE SYMBOL "+" INDICATES THAT MODEL A IS STATISTICALLY BETTER THAN MODEL B ($p < 0.05$), AND "=" INDICATES THAT MODEL A AND MODEL B ARE STATISTICALLY SIMILAR ($p > 0.05$).

Model A	Model B	Jaccard
Resnet50-Unet	Mobilenet-Unet	=
Resnet50-Unet	Vgg-Unet	=
Resnet50-Unet	Unet	=
Resnet50-Unet	Segnet	+
Mobilenet-Unet	Vgg-Unet	=
Mobilenet-Unet	Unet	=
Mobilenet-Unet	Segnet	+
Vgg-Unet	Unet	=
Vgg-Unet	Segnet	+
Unet	Segnet	=

TABLE IV
OVERALL SEGMENTATION PERFORMANCE BETWEEN MOBILENET-UNET AND MLP.

Index	Mobilenet-Unet	MLP
Jac	$\mathbf{0.945 \pm 0.073}$	0.803 ± 0.083
Dic	$\mathbf{0.966 \pm 0.061}$	0.874 ± 0.091
FPR	$\mathbf{0.007 \pm 0.012}$	0.038 ± 0.024
FNR	$\mathbf{0.028 \pm 0.050}$	0.085 ± 0.084

a classical approach. The comparison was made based on the indices already used, and all the results were confirmed through a statistical test. Table IV shows that the Mobilenet-Unet network presented a significantly better result for all metrics in all cases. This fact demonstrates the expected superiority of CNN-based models over classic models. This can be explained as traditional methods tend to use more general features, while deep network models are able to find more specific descriptors.

Finally, Figure 2 illustrates segmentation examples obtained by the five CNN models and the MLP-based traditional model. In the first two lines, the segmentations were obtained by training all models individually in each base. MI+IN means the combinations of both datasets. The segmentations show both the breast tissue and the pectoral muscle. It is observed, in general, that the segmentation models Resnet50-Unet and Mobilenet-Unet obtained better segmentation quality than the others.

V. CONCLUSION

For the improvement of a CAD system, the performance of the segmentation step is essential. Besides, all the processes must be optimized, including the computational cost. Therefore, considering the trade-off between accuracy and training time, in this work, several CNN-based semantic segmentation models and a conventional model based on MLP network were compared for segmenting relevant anatomic regions of a mammography image, namely, breast tissue and pectoral muscle. For this task, four CNN variations of Unet were evaluated: Resnet50-Unet, Mobilenet-Unet, Vgg-Unet, and Unet with Vanilla encoder. Additionally, a Segnet was also assessed. This exploration of different backbones is one of the main contributions of this study.

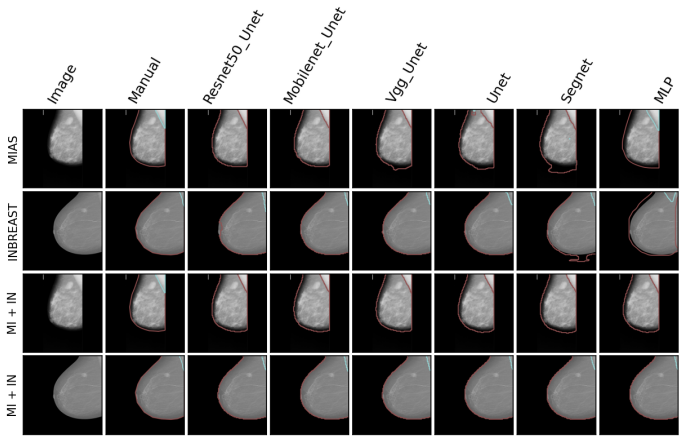

Fig. 2. Segmentation of the breast and pectoral muscle by the five CNN models and the conventional model for each dataset's mammographic image.

The objective was to evaluate the generalization capacity of well-established pre-trained CNN models that have been developed by the computer vision community for the automatic segmentation of mammography images. The models generated in this study are available on request by the authors, which can be tested on new images or improved trough transfer learning on other datasets.

Two public datasets were used combined for evaluating the CNN models: MIAS dataset with 322 images, INbreast with 201 images. For reducing class imbalance problems between classes, only MLO images were used.

According to the experimental results, the model Mobilenet-Unet obtained the best segmentation performance in most scenarios with $Jac > 0.94$ and $Dic > 0.96$. These results prove the greater efficiency of networks with more layers. Besides, Mobilenet-Unet's training time is approximately 2 to 3 times shorter than that of Resnet50-Unet in our computing platform due to the use of depthwise separable convolutions that considerably reduce the number of parameters.

The best Mobilenet-Unet model was compared with a conventional model based on MLP. As expected, the results demonstrate the superiority of the CNN-based model in all three datasets.

Therefore, according to experimental results, Mobilenet-Unet is a potential candidate for the semantic segmentation of anatomic regions in a fully automated end-to-end CAD system.

Our future work will explore the inclusion of more mammographic images, such as those found in the Curated Breast Imaging Subset of DDSM (CBIS-DDSM), the training of more CNN models, such as DeepLab and Mask-RCNN. It will incorporate the classification of breast density and anomalous regions as microcalcifications and masses. The evaluations should involve more tests where the models are trained in a set of images and evaluated in others in order to validate their generalization capabilities.

REFERENCES

[1] https://www.inca.gov.br/, Acessed: 2020-06-03

[2] Carvalho, I. M., Luz, L. M. S., Alvarenga, A. V., Azevedo, C. M., Pereira, W. C. A., Infantosi, A. F. C., (2006) Segmentação do Músculo Peitoral em Mamografias Utilizando Operadores Morfológicos, In: Proceedings of the XX Congresso Brasileiro de Engenharia Biomédica, p. 204–207.

[3] Elter M, Schulz-Wendtland R, Wittenberg T. The prediction of breast cancer biopsy outcomes using two CAD approaches that both emphasize an intelligible decision process. Med Phys. 2007 Nov;34(11):4164-72. doi: 10.1118/1.2786864. PMID: 18072480.

[4] Wilfrido Gómez-Flores and Wagner Coelho de Albuquerque Pereira, A comparative study of pre-trained convolutional neural networks for semantic segmentation of breast tumors in ultrasound, 2020, Computers in Biology and Medicine, vol. 126, pp. 104036. https://doi.org/10.1016/j.compbiomed.2020.104036.

[5] Sukassini, M & T, Velmurugan. (2016). Noise removal using Morphology and Median filter Methods in Mammogram Images.

[6] Wang K., Khan N., Chan A., Dunne J., Highnam R. (2019) Deep Learning for Breast Region and Pectoral Muscle Segmentation in Digital Mammography. In: Lee C., Su Z., Sugimoto A. (eds) Image and Video Technology. PSIVT 2019. Lecture Notes in Computer Science, vol 11854. Springer, Cham.

[7] de Oliveira H.N., de Avelar C.S., Machado A.M.C., de Albuquerque Araujo A., dos Santos J.A. (2019) Exploring Deep-Based Approaches for Semantic Segmentation of Mammographic Images. In: Vera-Rodriguez R., Fierrez J., Morales A. (eds) Progress in Pattern Recognition, Image Analysis, Computer Vision, and Applications. CIARP 2018. Lecture Notes in Computer Science, vol 11401. Springer, Cham.

[8] Luis Eduardo Aguilar-Camacho, Wilfrido Gómez-Flores and Juan Humberto Sossa-Azuela, A Comparative Study of Neural Computing Approaches for Semantic Segmentation of Breast Tumors on Ultrasound Images, 2021, XXVII Brazilian Congress on Biomedical Engineering, IFMBE Proceedings 83, https://doi.org/10.1007/978-3-030-70601-2_241.

[9] Stephan Trenn, Multilayer perceptrons: Approximation order and necessary number of hidden units, 2008, IEEE Trans Neural Networks 19:836–844.

[10] Jhang, Jieun. Localization of Microcalcification on the Mammogram Using Deep Convolutional Neural Network. (2018).

[11] O'Mahony Niall, Campbell Sean, Carvalho Anderson, et al. Deep Learning vs. Traditional Computer Vision in Advances in Computer Vision (Arai Kohei, Kapoor Supriya. , eds.)(Cham):128–144Springer International Publishing 2020.

[12] Andrik Rampun, Karen López-Linares, Philip J. Morrow, Bryan W. Scotney, Hui Wang, Inmaculada Garcia Ocaña, Grégory Maclair, Reyer Zwiggelaar, Miguel A. González Ballester, Iván Macía, (2019) Breast pectoral muscle segmentation in mammograms using a modified holistically-nested edge detection network. Medical Image Analysis, Volume 57, Pages 1-17.

[13] Suckling, J., et al.: The mammographic image analysis society digital mammogram database. In: Exerpta Medica. International Congress Series, vol. 1069, pp. 375–378 (1994).

[14] Moreira, I.C., Amaral, I., Domingues, I., Cardoso, A., Cardoso, M.J., Cardoso, J.S.: INbreast: toward a full-field digital mammographic database. Acad. Radiol. 19(2), 236–248 (2012).

[15] Ronneberger, O., Fischer, P., Brox, T.: U-Net: convolutional networks for biomedical image segmentation. In: Navab, N., Hornegger, J., Wells, W.M., Frangi, A.F. (eds.) MICCAI 2015. LNCS, vol. 9351, pp. 234–241. Springer, Cham (2015). https://doi.org/10.1007/978-3-319-24574-4 28.

[16] K. He, X. Zhang, S. Ren, J. Sun, Deep Residual Learning for Image Recognition, 2015 arXiv:1512.03385.

[17] Andrew G. Howard and Menglong Zhu and Bo Chen and Dmitry Kalenichenko and Weijun Wang and Tobias Weyand and Marco Andreetto and Hartwig Adam, MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications, 2017 arXiv:1704.04861.

[18] Karen Simonyan and Andrew Zisserman, Very Deep Convolutional Networks for Large-Scale Image Recognition, 2015 arXiv:1409.1556.

[19] V. Badrinarayanan, A. Kendall, R. Cipolla, Segnet: a deep convolutional encoder-decoder architecture for image segmentation, IEEE Trans. Pattern Anal. Mach. Intell. 39 (12) (2017) 2481–2495.

[20] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, L. Fei-Fei, Imagenet: a large-scale hierarchical image database, 2009 IEEE Conference on Computer Vision and Pattern Recognition, IEEE, 2009, pp. 248–255.

[21] D. Gupta, Image Segmentation Keras : Implementation of Segnet, FCN, UNet, PSPNet and other models in Keras, 2021, GitHub, GitHub repository, https://github.com/divamgupta/image-segmentation-keras.

[22] J. Kaufmann, A. G. Schering, Analysis of variance ANOVA (John Wiley & Sons Inc., 2014).