AUTOMATIC DENSE TISSUE SEGMENTATION BASED ON FULLY CONVOLUTIONAL NETWORK FOR MAMMOGRAPHY IMAGES

CARLOS SANTIAGO BENITEZ MALAVER

INDUSTRIAL UNIVERSITY OF SANTANDER
FACULTY OF PHYSICAL-MECHANICAL ENGINEERING
SCHOOL OF ELECTRICAL, ELECTRONIC AND TELECOMUNICATIONS ENGINEERING
BUCARAMANGA
2020

AUTOMATIC DENSE TISSUE SEGMENTATION BASED ON FULLY CONVOLUTIONAL NETWORK FOR MAMMOGRAPHY IMAGES

CARLOS SANTIAGO BENITEZ MALAVER

A thesis to get the degree of Electronic Engineer

Director:
SAID DAVID PERTUZ ARROYO
PhD in Computer Science

INDUSTRIAL UNIVERSITY OF SANTANDER
FACULTY OF PHYSICAL-MECHANICAL ENGINEERING
SCHOOL OF ELECTRICAL, ELECTRONIC AND TELECOMUNICATIONS ENGINEERING
BUCARAMANGA
2020

AGRADECIMIENTOS

A Dios agradezco por la fortaleza y sabiduría que me ha brindado. A mi familia por su inmenso esfuerzo para apoyar cada uno de metas y objetivos. A mi director de proyecto quien de la mejor manera mostro el camino para llevar a cabo este trabajo y a cada una de las personas que estuvieron presente para motivarme.

CONTENTS

pag.

LIST OF TABLES

pag.

LIST OF FIGURES

pag.

RESUMEN

TÍTULO: SEGMENTACIÓN AUTOMÁTICA DE TEJIDO DENSO BASADO EN REDES COMPLETAMENTE CONVOLUCIONALES EN IMAGENES DE MAMOGRAFIA*

AUTOR: CARLOS SANTIAGO BENITEZ MALAVER**

PALABRAS CLAVE: DENSIDAD PORCENTUAL DE SENO, TEJIDO FIBROGLANDULAR, MAMOGRAFÍA DIGITAL, REDES COMPLETAMENTE CONVOLUCIONALES

DESCRIPCIÓN: La densidad porcentual de seno (PD) es uno de los factores de riesgo más importantes asociados con el desarrollo del cáncer de seno. Por lo tanto, la estimación precisa de la PD es una tarea importante para la evaluación del riesgo de cáncer de seno basada en el análisis mamográfico. Para evitar variabilidad entre lectores se opta por el desarrollo de algoritmos de estimación de densidad automáticos. Sin embargo, la segmentación automática del tejido fibroglandular (FGT) es una tarea difícil ya que las características morfológicas tanto sutiles como complejas se proyectan en una Mamografía Digital de Campo Completo (FFDM). En este trabajo presentamos un algoritmo híbrido basado en una red completamente convolucional y un algoritmo de agrupamiento basado en intensidad para la estimación de tejido denso. Para fines de validación, utilizamos un conjunto de datos que se usan como referencia y consta de 582 mamografías con tejido denso segmentado manualmente por un radiólogo experto. Como resultado, la segmentación de tejido denso usando la selección de clúster demuestra una mejoría de 8% en la mediana del coeficiente de similitud de Dice (DSC) respecto a la segmentación dada por la red neuronal. las estimaciones de PD obtenidas con el método propuesto no arrojan diferencias estadísticamente significativas con respecto a las estimaciones de PD del radiólogo. Además, el método propuesto arroja unas medianas de DSC y error de PD de 0.795 y 0.077, respectivamente. Al compararse con un algoritmo clínicamente validado del estado del arte el algoritmo propuesto alcanzó un mayor rendimiento.

---

*Trabajo de grado

** Facultad de Ingeniería Fisicomecánicas. Escuela de Ingenierías Eléctrica, Electrónica y de Telecomunicaciones. Director: Said David Pertuz Arroyo, PhD en Ciencias de la Computación

ABSTRACT

TITLE: AUTOMATIC DENSE TISSUE SEGMENTATION BASED ON FULLY CONVOLUTIONAL NETWORK FOR MAMMOGRAPHY IMAGES *

AUTHOR: CARLOS SANTIAGO BENITEZ MALAVER**

KEY WORDS: PERCENT DENSITY, FIBROGLANDULAR TISSUE, DIGITAL MAMMOGRAPHY, FULLY CONVOLUTIONAL NETWORK

DESCRIPTION: Percent breast density (PD) is one of the most important risk factors associated with developing breast cancer. Therefore, accurate estimation of PD is an important task for breast cancer risk assessment based on mammographic analysis. To avoid variability between readers, the development of automatic density estimation algorithms is chosen. However, the automatic segmentation of fibroglandular tissue (FGT) is a difficult task as both subtle and complex morphological features are projected on a Full Field Digital Mammography (FFDM). In this work, we present a hybrid algorithm based on a fully convolutional network and an intensity-based clustering algorithm for the estimation of dense tissue. For validation purposes, we used a reference data set consisting of 582 dense tissue mammograms manually segmented by an expert radiologist. As a result, dense tissue segmentation using cluster selection demonstrates an 8% improvement in median Dice similarity coefficient (DSC) regarding to the segmentation given by the neural network. PD estimates obtained with the proposed method do not show statistically significant differences with respect to PD estimates from the radiologist. Furthermore, the proposed method yields medians of DSC and PD error of 0.795 and 0.077, respectively. When compared with a clinically validated state of the art algorithm, the proposed algorithm achieved higher performance.

---

* Bachelor Thesis

** Faculty of Physical-Mechanical Engineering. School of Electrical, Electronic and Telecomunications Engineering. Director: Said David Pertuz Arroyo, PhD in Computer Science

INTRODUCTION

One of the most common types of cancer in women is breast cancer [1]. Studies show that breast cancer screening based on mammography may reduce mortality rates up to 63% [2]. Clinically, it is usual to determine percent density (PD) from mammograms since it is one of the strongest risk factors associated with the development of cancer. However, to date, the estimation of PD remains a difficult challenge.

To determine PD, the ratio between the area corresponding to fibroglandular tissue and the total breast area is calculated [3]. Therefore, an adequate segmentation of region of interest (ROI) in mammographic images is essential to accurately quantify breast PD and, consequently, assess breast cancer risk. Typically, an expert radiologist estimates the PD of a patient by segmenting the breast region and dense tissue either manually or with the help of semi-automated CAD (computer aided detection) systems[4]. Unfortunately, both methods are subjective since the segmentation requires human interaction. Also, the fatigue given by analyzing a lot of images can alter the segmentation, or simply, because the morphological structure of the fibroglandular tissue is difficult to segment with manual annotation tools. To solve these problems, radiologists may use a second opinion from fully automated dense tissue segmentation algorithms [5].

---

[1] WORLD HEALTH ORGANIZATION. [Website]. Cancer. 2018. Available from internet: https://www.who.int/cancer/en/.

[2] TABÁR, László, *et al.* Beyond randomized controlled trials: organized mammographic screening substantially reduces breast carcinoma mortality. *Cancer.* 2014.

[3] GIGER, M; KARSSEMEIJER, M and SCHNABEL, J. Breast Image Analysis for Risk Assessment, Detection, Diagnosis, and Treatment of Cancer. *Annual Review of Biomedical Engineering*. 2013, pp. 327-357.

[4] Ibid.

[5] Ibid.

Recently, fully convolutional neural networks (FCNN) are most often used on semantic segmentation and objects detection because they have a better performance [6]. Existing deep learning methods for dense tissue segmentation have shown high performance outperforming previous existing methods [7]. One of the main potentials of FCNN-based methods are their potential for further refinement and improvement using transfer learning methods [8]. However, to date, existing methods are not publicly available for comparison, validation, or further development. FGT segmentation based on deep learning methods is a new concept, to which little attention has yet been paid. For this reason, we show an algorithm implementation composed of deep learning and unsupervised method for PD estimation.

[6] LONG, Jonathan; SHELHAMER, Evan and DARRELL, Trevor. Fully Convolutional Networks for Semantic Segmentation. 2014.

[7] LEE, Juhun and NISHIKAWA, Robert M. Automated mammographic breast density estimation using a fully convolutional network. *Medical Physics*. 2018, vol. 45, no. 3, pp. 1178–1190.

[8] SHIN, Hoo-chang, *et al.* Deep Convolutional Neural Networks for Computer-Aided Detection: CNN Architectures, Dataset Characteristics and Transfer Learning. *IEEE Transactions on Medical Imaging*. 2016, vol. 35, no. 5, pp. 1285–1298.

1. METHODS AND MATERIALS

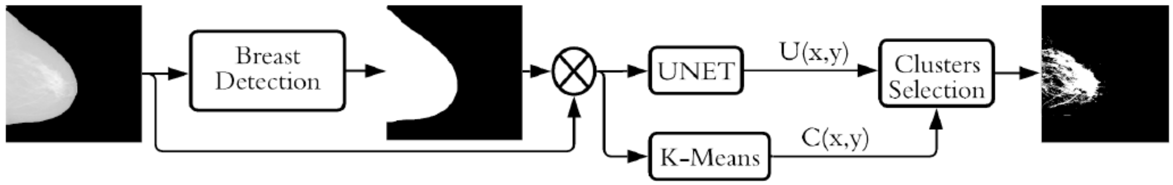
1.1 DATASET

We have access to a mammographic dataset of 582 images corresponding to the craniocaudal view of screening mammograms. 244 mammograms were captured with a Philips MicroDose SI and 338 were captured with a GE Senographe Essential mammography system. Ground-truth segmentation was performed by a radiologist with experience in breast imaging. For training and validation purposes, the dataset was randomly divided into a training, validation, and testing sets with 384 (66%), 99 (17%) and 99 (17%) mammograms, respectively.

1.1 PROPOSED METHOD

In this project, we do not focus on breast area segmentation method. Instead, it is used OpenBreast to breast detection. On the other hand, we proposed a dense tissue segmentation with cluster thresholding method using a CNN segmentation and intensity-based clustering. In fact, dense tissue segmentation is improved compared with the use of only CNN. In the figure 1 is shown the proposed method.

Figure 1. Block diagram of the proposed method. The input is the pre-processed mammogram. Dense tissue segmentation is obtained by performing cluster selection on the output of CNN-based segmentation and an intensity-based clustering using k-Means.



1.1.1 Pre-processing. Mammography images are resized to 512 x 512 pixels in grayscale of 8 bits format. OpenBreast 9 is utilized for the detection of the breast region and pectoral wall. An erosion of 1.32 mm on the breast mask is utilized to discard the breast outer contour. The masked mammogram is used for training purposes and be the input of U-NET and K-MEANS algorithms.

1.1.2 CNN-based segmentation. UNET[10], a network used for medical segmentation was trained on MATLAB®. In table 1 is shown the training parameters. Data augmentation will be done using translations of ±32 pixels and reflections. Default settings are selected for epsilon, gradient and squared gradient decay factors as is shown in Kingma et al.[11]. The network was trained during 8064 iterations achieved accuracy of 93% measured using intersection over union. Besides, learning rate is changed every 10 epochs on a factor of 0.1 yielding a low oscillation near the loss function minimum.

---

[9] PERTUZ, Said, *et al.* Open Framework for Mammography-based Breast Cancer Risk Assessment, IEEE-EMBS International Conference on Biomedical and Health Informatics. 2019.

[10] RONNEBERGER, Olaf; FISCHER, Philipp and BROX, Thomas. U-Net: Convolutional Networks for Biomedical Image Segmentation. 2015.

[11] KINGMA, Diederik and BA, Jimmy. Adam: A method for stochastic optimization. 2014.

Table 1. Training parameters for UNET.

Parameter	Value
Optimizer	ADAM
Initial Learning Rate	1 E-5
Weight Factor	5 E-4
Mini Batch Size	1
Shuffle	Every-Epoch
Validation Frequency	50

1.1.3 Cluster selection. Only using the pre-processed image as input, clustering is performed with K centroids using K-Means[12] algorithm. For reproducibility, initial centroids are defined evenly spaced. Based on clustered images, the output of the UNET segmentation $U(x,y)$ is used to segment dense tissue via majority voting. since the images were clustered, the i-th cluster $C_i(x,y)$ is used to calculate a partial mask $P_i(x,y)$ in (1).

$$P_i(x,y) = U(x,y) \cap C_i(x,y) \quad (1)$$

$$\lambda_i = \frac{\sum_{(x,y)} P_i(x,y)}{\sum_{(x,y)} C_i(x,y)} \quad (2)$$

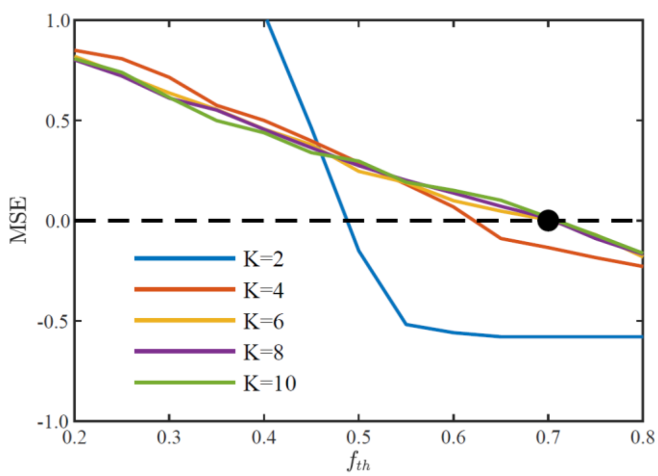
In (2), the ratio $\lambda_i \in [0,1]$ between P_i and C_i is used for quantify the tissue segmented by UNET into the clusters. Dense tissue segmentation is performed by selecting the lambda value such that $\lambda_i > f_{th}$. Where, f_{th} is the voting threshold.

[12] LLOYD, Stuart P. Least squares quantization in PCM*. Information Theory on IEEE Transactions.* 1982, vol. 28, pp. 129-137.

1.2 HYPERPARAMETER TUNING

The proposed method depends on two hyper-parameters: the number of clusters K and the voting threshold f_{th}. In order to find an optimal threshold f_{th_o}, a f_{th} sweep is performed in the range of [0.2,0.8] assessing Mean Signed Error (MSE) between the PD estimated by the proposed method and the ground truth on the validation set. To analyze effects of K, the sweep is performed for K values as is shown in the fig. 2. Since the sweeps cut on zero MSE for majority K, else f_{th_o} is equal to 0.7. Notice that from $K = 6$ the sweeps have similar linear behavior cutting on f_{th_o}. It is considered that K values less than 6 have poor quantity of clusters and for K greater than 6 the PD estimation should not change. By default, in this work we used $K = 6$ to segment dense tissue on FFDM.

Figure 2. Mean signed error of the estimated PD as a function of the selection threshold $\boldsymbol{f_{th}}$ in the validation set.

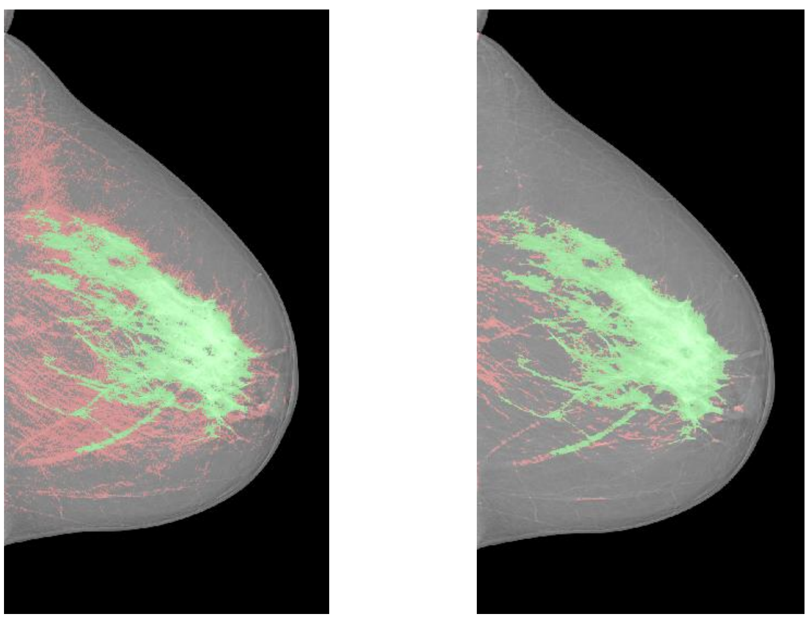


2. RESULTS

2.1 EFFECT OF CLUSTER SELECTION

For dense tissue segmentation, Dice Similarity Coefficient (DSC) assessment is performed using U-NET and the proposed method, showing 0.72 and 0.80 median values, respectively. In the fig. 3, a sample of both segmentation methods are shown compared with ground truth.

Figure 3. Effect of cluster selection on segmentation. (a) Without cluster selection. (b) With cluster selection. Correctly and wrongly segmented are shown in green a red pixels, respectively.



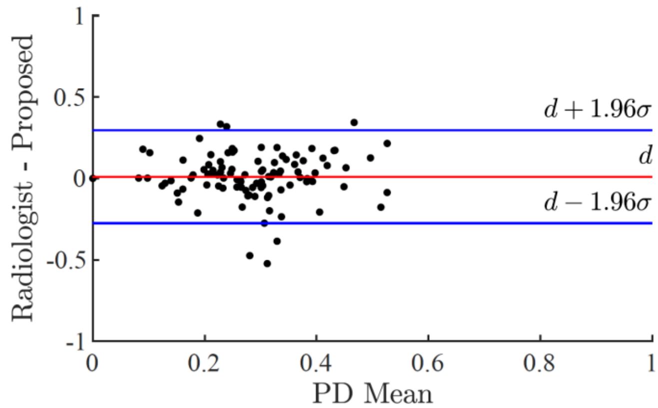
2.2 PERFORMANCE ANALYSIS

In the figure 4, Bland-Altman plot is performed to compare PD measures from radiologist and the proposed method. Mean difference between methods denoted as d in red color line is approximately zero, this indicates nonsystematic error due to tuning performed in 3.3 section. The limits of agreement (LOA) represented by blue color line are established on ±0.25 approximately. Comparing between majority scatter points and wide LOA results an ambiguous analysis. Hence, Wilcoxon rank sum test is performed for PD estimation by the proposed method and radiologist, resulting no significant statistical differences with P-value = 0.557. In the table 2 are assigned PD estimation by category of American College of Radiology Breast Imaging Reporting and Data System (BI-RADS) four edition. Based on the proposed method, in the 1^{th} category were 24 PD estimation without category change regarding ground truth and 37 correctly estimated in the 1^{nd} category. In total, 61 PD estimations did not have category changes. To analyze PD estimation differences from mammography system, testing set is divided on 40 mammograms acquired by Philips and 58 by General Electric Medical Systems. Where, there are statistical differences based on Wilcoxon test with p-value = 1.567 E-3.

Table 2. PD estimated in each category by radiologist and proposed method. *PD estimations by our method without category changes regarding to ground truth.

Category	Ground Truth	Proposed	*Correctly
Q1	39	39	24
Q2	55	54	37
Q3	5	6	0
Q4	0	0	0
Total	99	99	61

Figure 4. Bland-Altman plot comparing obtained density estimates and ground-truth by the radiologist.



4.3 COMPARATIVE ANALYSIS

The proposed method is compared with LIBRA. Assessing DSC and PD absolute error, then, was performed Kolmogorov-Smirnov test. None of the measures showed normal distributions. In the table 3, is appreciated DSC and PD error median given by proposed and LIBRA methods. P-value indicates that our method did not present statistical differences on both measures compared to LIBRA. Statistical tests were performed with CI = 95%.

Table 3. Performance comparison of methods. *Wilcoxon rank sum test between methods.

Measure	Proposed	LIBRA	*p-value
DSC	0.795	0.767	0.654
PD-err	0.077	0.092	0.093

3. DISCUSSION AND CONCLUSSIONS

Mammography images are characterized for high contrast variations that difficult the segmentation tasks of FGT using only CNN. A sample of UNET segmentation in fig. 3,a presented spurious segmentation, the same occurs for all FGT segmentation. Nevertheless, spurious should not be attenuated using morphological operators due to each segmented pixel is significant. Through clusters selection is considered all UNET segmented pixels and via majority voting is segmented FGT tissue.
The proposed algorithm shown statistical differences on dense tissue segmentation from distinct mammographic system. Like LIBRA, the performance may change depending on the system. However, U-net was trained with very few mammograms. Based on p-value in table 3, our method has similar performance comparing to clinically validated software, LIBRA. From the proposed method, in the table 2, 38 out of 99 cases PD estimation change of category. According to fig. 4, highest PD values tends to be underestimated.

A combination of deep learning and unsupervised machine learning techniques is presented, resulting an improvement of 8% DSC median compared to U-net segmentation. We leave an open method that may be used as basis of algorithms that improve the PD estimation only changing the CNN architecture or clustering algorithm.

BIBLIOGRAPHY

GIGER, Maryellen; KARSSEMEIJER, Nico and SCHNABEL, Julia. Breast Image Analysis for Risk Assessment, Detection, Diagnosis, and Treatment of Cancer. *Annual Review of Biomedical Engineering*. 2013, pp. 327-357.

KINGMA, Diederik and BA, Jimmy. Adam: A method for stochastic optimization. *International Conference on Learning Representations.* 2014.

LLOYD, Stuart P. Least squares quantization in PCM*. Information Theory on IEEE Transactions.* 1982, vol. 28, pp. 129-137.

LEE, Juhun and NISHIKAWA, Robert M. Automated mammographic breast density estimation using a fully convolutional network. *Medical Physics*. 2018, vol. 45, no. 3, pp. 1178–1190.

LONG, Jonathan; SHELHAMER, Evan and DARRELL, Trevor. Fully Convolutional Networks for Semantic Segmentation. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 2014, pp. 3431-3440.

PERTUZ, SAID, *et al.* Open Framework for Mammography-based Breast Cancer Risk Assessment, IEEE-EMBS International Conference on Biomedical and Health Informatics. 2019.

RONNEBERGER, Olaf; FISCHER, Philipp and BROX, Thomas. U-Net: Convolutional Networks for Biomedical Image Segmentation. *LNCS*. 2015, vol. 9351, pp. 234–241.

SHIN, Hoo-chang *et al.* Deep Convolutional Neural Networks for Computer-Aided Detection: CNN Architectures, Dataset Characteristics and Transfer Learning. *IEEE Transactions on Medical Imaging*. 2016, vol. 35, no. 5, pp. 1285–1298.

TABÁR, László, *et al.* Beyond randomized controlled trials: organized mammographic screening substantially reduces breast carcinoma mortality. *Cancer.* 2014.

WORLD HEALTH ORGANIZATION. [Website]. Cancer. 2018. Available from internet: https://www.who.int/cancer/en/.