UNIVERSIDAD SAN FRANCISCO DE QUITO USFQ

Colegio de Ciencias e Ingenierías

Towards breast mass detection in mammography images using Deep learning

Carlos Mauricio Reátegui Tapia

Ingeniería en Ciencias de la Computación

Trabajo de fin de carrera presentado como requisito
para la obtención del título de
Ingeniero en ciencias de la computación

Quito, 19 de Diciembre de 2021

UNIVERSIDAD SAN FRANCISCO DE QUITO USFQ

Colegio de Ciencias e Ingenierías

HOJA DE CALIFICACIÓN
DE TRABAJO DE FIN DE CARRERA

Towards breast mass detection in mammography using deep learning

Carlos Mauricio Reátegui Tapia

Nombre del profesor, Título académico **Noel Pérez, Ph.D.**

Quito, 19 de Diciembre de 2021

© DERECHOS DE AUTOR

Nombres y apellidos: Carlos Mauricio Reátegui Tapia

Código: 00201852

Cédula de identidad: 1725234403

Lugar y fecha: Quito, 19 de Diciembre de 2021

RESUMEN

En este proyecto, un modelo de aprendizaje profundo fue entrenado para localizar lesiones (específicamente masas) en mamografías. El modelo sigue una arquitectura de YOLOv4 darknet y usa imágenes de la base de datos de acceso público DDSM para entrenamiento, validación y prueba. Una técnica de 5-fold cross validation fue usada para asegurar que los resultados fueran independientes de las particiones de sets de entrenamiento y prueba, alcanzando una mean average precision del 72%. Con modificaciones futuras a la arquitectura descrita en este documento, los resultados podrían ser incluso mejores, creando una poderosa y útil herramienta para radiólogos que buscan masas en mamografías.

Palabras clave: YOLOv4, darknet, Mammography, Mass, DDSM, Localization, Deep Learning, Preprocessing.

ABSTRACT

In this project, a deep learning model was trained to locate lesions (masses specifically) on mammography. The model follows a YOLOv4 darknet architecture and uses images from the public DDSM database to train, validate and test it. A 5-fold cross-validation technique was used to ensure the resultsare independent of training and test sets partition, reaching a mean average precision of 72%. With future modifications to the architecture described in this paper, the results could be even better, resulting in a powerful and helpful tool for radiologists looking for masses in mammograms.

Key words: YOLOv4, darknet, Mammography, Mass, DDSM, Localization, Deep Learning, Preprocessing.

TABLE OF CONTENTS

INDEX OF FIGURES

INTRODUCTION

Breast cancer is one of the most common for women around the world. According to the World Health Organization (WHO), in 2020, 2.3 million women were diagnosed with breast cancer, and 685.000 died [1]. It is the world's most prevalent cancer, with 7.8 million women living with it for the past five years. It can occur in women after puberty, but the rate increases as they become older. Early detection of this disease is essential to reduce the risk of death, and one of the most popular mechanics to diagnose is by using the mammogram's screening. However, detecting breast cancer through screening can be difficult due to human errors introduced by exhaustion, workload, and even the implicit difficulty of the lesions and their surrounding tissues, e.g., breast density, lesion location variation, shape, and texture, leading to an inappropriate diagnosis by radiologists. Therefore, automatic detection tools are beneficial to aid the experts when analyzing mammograms, giving them a second opinion about the presence of possible lesions.

Many different models have been used to solve this problem. A study using a model based on YOLOv3 using the database INBreast obtained an 89% precision [2]. Another study used a vanilla YOLOv4 model using the DDSM database and achieved an 85% precision finding masses [3]. In this study, using a fusion-model approach, a precision of 95% was reached on the DDSM database, 98% on the INBreast Database and 98% on a private database [3]. While using a YOLO-based CAD system, a 99% location accuracy was achieved using the DDSM database [4]. A model based on Retinanet achieved a precision of 95% [5]. All the results

mentioned above are for finding masses.

In this work, we propose to use the standard YOLOv4 deep architecture to detect breast mass lesions in mammography images. The model must accurately localize the masses inside the mammograms and highlight the region of interest (area containing the lesions). We focused this research only on mass lesions because they are difficult to detect based on their similarity with the surrounding tissue.

MATERIALS AND METHODS

Database

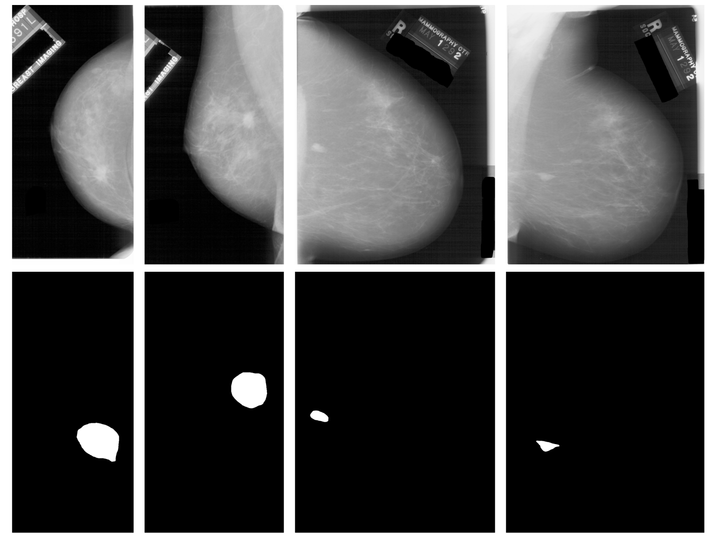

Figure 1: Some image samples with different breast views of the DDSM database (top row) and their lesion ground truth (bottom row)

The database used for the training, validation, and testing of this model is the DDSM-Database. This database was originally organized in volumes and cases. Each volume contained a collection of cases, and each case contained up to four mammograms and an overlay file that shows the results of the analysis made to the images in the case [6]. Some examples of image with different breast views is shown in Fig. 1.

There were starting coordinates and chain-coded borders of the abnormalities found in the mammography in these overlays, dividing them into masses and calcifications. This information was used to create masks for each lesion in the images, which were then transformed to YOLO annotations (class, x-center, y-center, width, height) that served as ground truth. 600 images containing masses were corrupted, so they had to be deleted. The database was organized to obtain the images containing masses, resulting in 1843 mammograms with a total of 1950 masses.

Deep learning models

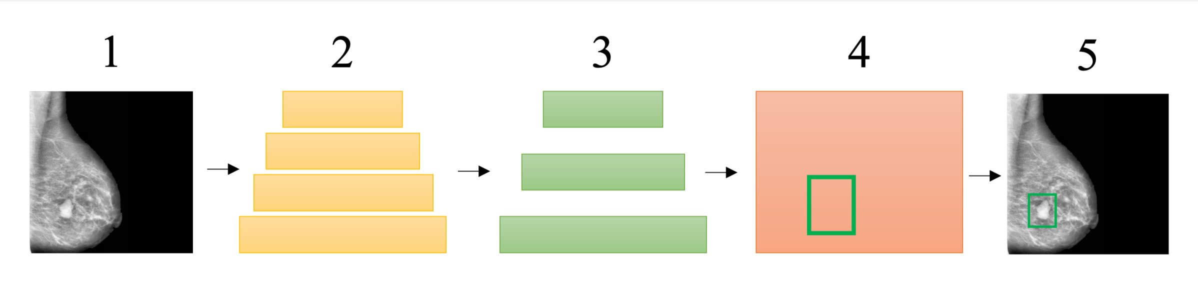

Figure 2: An overview of the YOLOv4 architecture. From left to right, input (1), backbone (2), neck (3), head (4), and the output (5).

Deep learning is a machine learning technique that, in simple words, teaches computers to learn by example. Deep learning can be used for many things, including computer vision. With it, we can localize different objects on an image, classify and localize them [7].

The proposed method was YOLOv4, released in 2020. Yolo stands for ”you only look once” and is the most stable version of YOLO at the moment. It is a one-stage object detection model that makes improvements on YOLOv3. Commonly, two-stage models are used when the background is independent of the foreground. Nevertheless, in the case of mammograms, lesions can never be independent of the breast region because they are always inside, which means a one-stage model works better [8].

This detector has four areas connected in the work pipeline: input, backbone, neck, and head. Images are used as input for the model. The backbone consists of CSPDarknet53. The neck is formed by a path aggregation network and spatial pyramid pooling. Finally, YOLOv3 is used for the head [8]. Figure 2 shows the architecture of the model.

The model only has one class: masses. The output of the model will be an image (the mammography used as input) with a bounding box surrounding the mass (in case a mass was found). The mass will have a number on top of it, showing the confidence that the model has that the prediction is correct. This number will range from 0 to 1.

Experimental setup

All the experiments required for the training and testing of this model were executed on an NVIDIA Tesla V100-DGXS- 32GB GPU and AlexeyAB darknet implementation, available on github [9].

Data processing

In order to enhance the images for training, many preprocessing techniques had to be used. The first change was to crop the images in all directions. This had to be done because many mammograms had white borders that do not provide any helpful information and can affect the training of the model by creating an arbitrary edge as a feature in the images that the model could learn [10]. 1% of the image was cropped from the left and right side, and 4% was cropped from the top and bottom sides.

After that, all useless artifacts for training the model were removed. These artifacts can be annotations or noise outside the breast. In order to do this, four steps were created.

The first step is binarizing the image. A threshold of 40 and a max value of 255 were used to find the contours in the image. This results in a new image with black and white pixels called a mask. This brings us to the second step: edit the mask. In this step, we make the contours of the different white shapes rougher and get rid of small contours that look like noise. After editing the mask, it has well-defined contours of objects in the image. The next step is to find the biggest object in the image, which will always be the breast. After we find it, we create a new mask that only contains this object, which results in a mask that only contains the breast and no extra noise or artifacts. The fourth and final step is to apply this mask to the image, which outputs an image that contains only the breast and a black background free of noise or other artifacts.

The next step consisted of making an optional flip. This flip is only made to images of breasts on the right, facing left. Having all images face the same direction will make it easier to reach the padding step because all the padding will be done to the right. After that, we applied CLAHE enhancement to the image, with a clip of 4 and a tile of 8x8. This was done to highlight lesions in the breast, because in the original image, breast density makes it hard to spot any abnormalities.

Finally, we padded the image to the right to make them have a squared shape and resized them to have a dimension of 512x512. We can afford to resize the images to this size because masses are not as small as calcifications and can still be identified with smaller image sizes.

At the training moment, image augmentation with mosaic and horizontal flip was used for all images.

Training, validation and test sets

The training set of masses was chosen to be 90% of all the images that contain masses, leaving the test set to receive the remaining 10%. This resulted in 1658 images and YOLO annotations on the training set and 185 images and YOLO annotations on the test set. 10% of the training set (165 images) was chosen to be the validation set.

Model configuration

A batch size of 16 is used, with 64 subdivisions. The size of the input is 512x512 pixels. We use a momentum of 0.949, a decay of 0.0005, and a learning rate of 0.0001. All these parameters are configured in the cfg file yolov4-obj.cfg that we will use for the YOLO model. After making tests with different CLAHE clip enhancements, a clip of 4 was found to be the most optimal for image processing. A confidence threshold of 0.25 and an intersection over the union (IOU) threshold of 0.15 were used. All these values were found to be optimal empirically.

Assesment metrics

The assessment metrics to be used for this model are precision, recall, f1-score, and mean average precision.

RESULTS AND DISCUSSIONS

The model was tested to make predictions on the test set defined previously. Normally, the data on the test set do not have a ground truth, but in this case, we had both the masks and the YOLO annotations to make comparisons.

A sample of predictions is shown in figure 3, containing comparisons between the original image used as input, its corresponding mask showing the location of the mass, and the output of the model with a bounding box surrounding the mass found. Most predictions are accurate. However, we can see some errors like a false negative on the third column and an extra bounding box of a true positive on the last column. These errors could be attributed to high breast density and the particular shape of the masses.

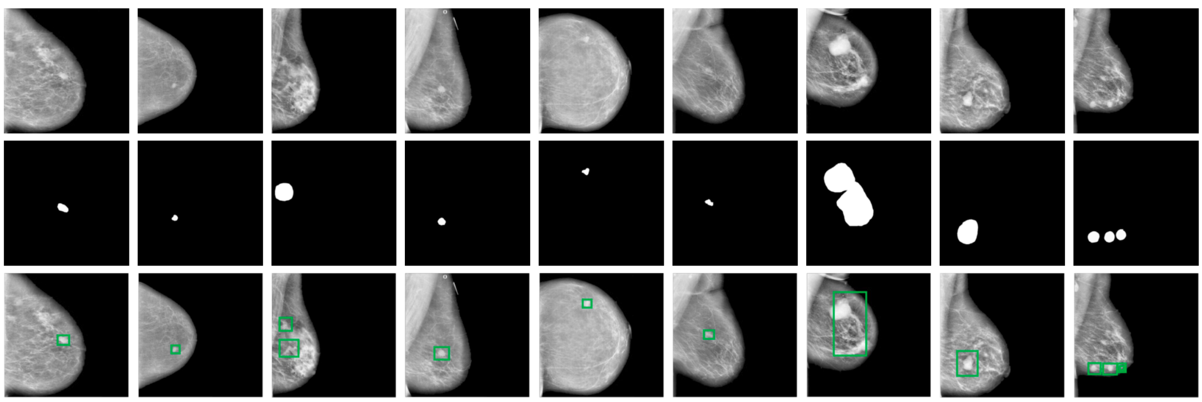

Figure 3: Mass localization predictions on mammograms. Top row: original image, middle row: ground truth, bottom row: predictions

Of the state-of-the-art models mentioned in the introduction, the most relevant to make a comparison to is the study that used a vanilla YOLOv4 model [3] because they use the same database (DDSM) and base model. Regarding the detection of masses, they achieved a precision of 85%, compared to our 72%. This contrast in precision can be attributed to the difference in the number of images. Their study used 2907 mammograms (including both masses and calcifications), in comparison, we used 1843. 600 images containing masses had to be removed due to image corruption, which affected the performance of the model. Another issue is the database itself. DDSM does not have images of such quality as CBIS-DDSM or INBreast. Therefore experiments using these databases have better results.

With 5-fold cross-validation, we achieved a precision of 0.70, a recall of 0.65, an F1-score of 0.67, and a mean average precision of 0.72. These calculations were done with the best weights obtained during training.

Figure 4 shows a plot of training and test loss throughout the iterations of the model. The behavior of both losses tells us there is no overfitting in the model.

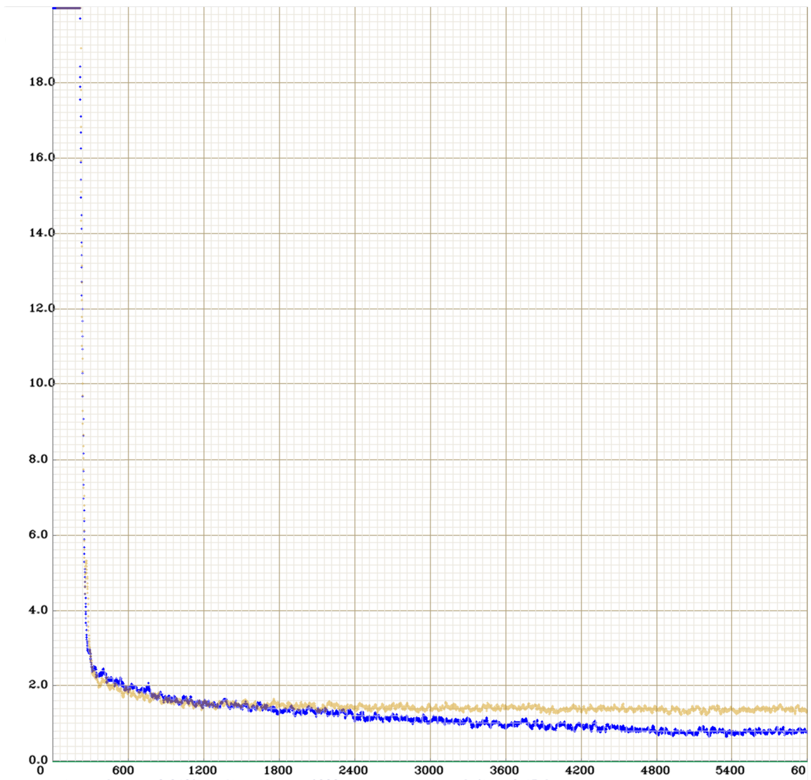

Figure 4: Training Loss (yellow) and Testing Loss (blue) through iterations (x-axis)

Figure 5 shows how the mean average precision evolves through each iteration. The peak value of this metric is first obtained at the 5400th iteration, reaching a value of 72%.

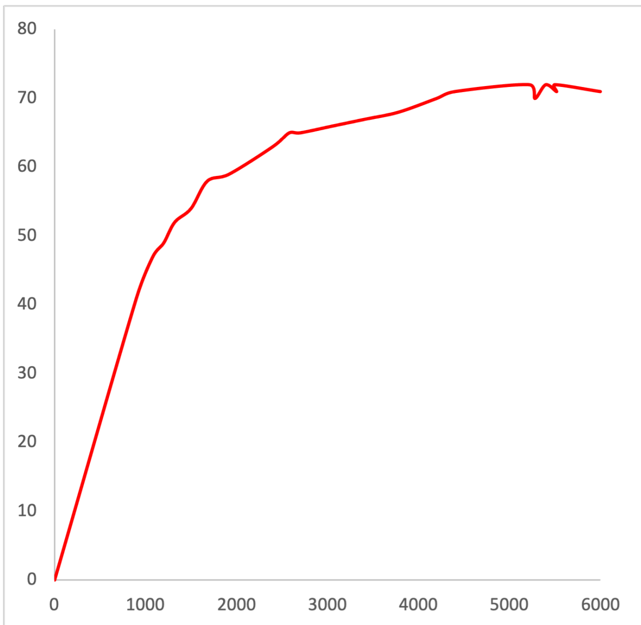

Figure 5: Mean Average Precision (y-axis) through iterations (x-axis)

Acknowledgement

Authors thank to the Applied Signal Processing and Machine Learning Research Group of USFQ for providing the computing infrastructure (NVidia DGX workstation) to implement and execute the developed source code.

Special thanks to Noel Pérez and Diego Benítez for their guidance in the creation of this project, and to my fellow partners that also worked in similar subjects regarding computer vision.

CONCLUSIONS

Using a 5-fold cross validation technique, we reached a 72% mean average precision on our model. These results were achieved on a YOLOv4 architecture with modified hyperparameters for fine tuning. A CLAHE enhancement clip of four was found empirically to be the best choice for preprocessing, and data augmentation proved to be useful at obtaining better results for the model.

There are a few ways it can be improved. We could use images of bigger sizes, which would make the model better at finding small masses like the last one in the results table. A bigger batch size could be used as well as fewer subdivisions. The use of training images that have no lesions (empty YOLO annotations txt files) could improve the performance of the model [9]. This could be done with more memory. Also, having access to the uncorrupted 600 images containing masses that had to be deleted would improve the performance of the model in all metrics.

Another injury that is taken into account when diagnosing breast cancer, is calcification. Attempts at making a model that can accurately detect calcifications could be a problem to tackle in the future.

BIBLIOGRAPHIC REFERENCES

[1] WHO, “Breast cancer,” 2021. [Online]. Available: https://www.who.int/ news- room/fact-sheets/detail/breast- cancer

[2] E.-S. S. T. M. Aly GH, Marey M, “Yolo based breast masses detection and classification in full-field digital mammograms,” 2020. [Online]. Available: https://pubmed.ncbi.nlm.nih.gov/33190942/

[3] A. Baccouche, B. Zapirain, A. Elmaghraby, and C. Castillo, “Breast lesions detection and classification via yolo-based fusion models,” Cmc -Tech Science Press-, vol. 69, pp. 1407–1425, 06 2021.

[4] P.J.G.G.K.T.R.P.V.E.C.M.H.S.K.T.Al-MasniMA, Al-Antari MA, “Simultaneous detection and classification of breast masses in digital mammograms via a deep learning yolo-based cad system. comput methods programs biomed,” 2018. [Online]. Available: https://pubmed.ncbi.nlm.nih.gov/29477437/

[5] I. L. M. Y. J. L. S. H. O. W. J. K. Hwejin Jung, Bumsoo Kim, “Detection of masses in mammograms using a one-stage object detector based on a deep convolutional neural network,” 2018. [Online]. Available: https: //journals.plos.org/plosone/article?id=10.1371/journal.pone.0203355

[6] U. of South Florida, “Ddsm: Digital database for screening mammography,” No date. [Online]. Available: http://www.eng.usf. edu/cvprg/mammography/database.html

[7] A. Murphy, “Mammography views,” 2021. [Online]. Available: https://radiopaedia.org/articles/mammography- views

[8] IBM, “Deep learning,” 2020. [Online]. Available: https://www.ibm. com/cloud/learn/deep- learning

[9] D. Cochard, “Yolov4 : A machine learning model to detect the position and type of an object,” 2020. [Online]. Avail- able: https://medium.com/axinc- ai/yolov4- a-machine- learning- model- to- detect- the- position- and- type- of- an- object-4f108ed0507b

[10] A. Bochkovskiy, “Yolov4: Optimal speed and accuracy of object detec- tion,” 2020. [Online]. Available: https://arxiv.org/pdf/2004.10934v1.pdf

[11] AlexeyAB, “darknet,” 2020. [Online]. Available: https://github.com/alexeyab/darknet#requirements

[12] C. Wong, “Segmenting abnormalities in mammograms (part 2 of 3),” 2021. [Online]. Available: https://towardsdatascience.com/can-you-find- the- breast- tumours- part- 2-of- 3- 1d43840707fc

[13] M. de Salud Pu ́blica, “Cifras de ecuador – ca ́ncer de mama,” 2021. [Online]. Available: https://www.salud.gob.ec/cifras- de- ecuador- cancer-de- mama/

[14] BootCampAI, “Introduccio ́n a las redes convolucionales,” 2019.[Online]. Available: https://bootcampai.medium.com/redes-neuronales-convolucionales- 5e0ce960caf8

[15] Arc, “Convolutional neural network,” 2018. [Online]. Available: https://towardsdatascience.com/convolutional- neural- network- 17fb77e76c05

[16] H. A. U. Z. e. a. Ribli, D., “Detecting and classifying lesions in mammograms with deep learning,” 2018. [Online]. Available: https://doi.org/10.1038/s41598- 018- 22437- z

[17] A. C. Society, “What does the doctor look for on a mammogram,” 2021. [Online]. Available: https://www.cancer.org/cancer/breast- cancer/screening- tests- and- early-detection/mammograms/what- does- the- doctor- look- for- on- a- mammogram.html

[18] R. S. S. C. E. V. Kerlikowske K, Grady D, “Efficacy of screening mammography: A meta-analysis,” 1995. [Online]. Available: doi:10.1001/jama.1995.03520260071035

[19] G. Pichardo, “Breast cancer and mammograms,” 2020. [Online]. Available: https://www.webmd.com/breast- cancer/mammograms

[20] J. Eetvelt, “End-to-end breast cancer detection in python — part 1,” 2020. [Online]. Available: https://towardsdatascience.com/end- to- end- breast- cancer- detection- in-python- part- 1- 13a1695d455

[21] Techzizou, “Yolov4-darknet installation and usage on your system (windows linux),” 2021. [Online]. Avail- able: https://medium.com/geekculture/yolov4- darknet-installation- and- usage- on- your- system- windows- linux- 8dec2cea6e81

[22] mathieuorhan, “Yolo-v3 and yolo-v2 for windows and linux,” 2018. [Online]. Available: https://github.com/mathieuorhan/darknet/ blob/master/README.md