

A Deep Learning Model for Mammography Mass Detection Using Mosaic and Reconstructed Multichannel Images

Nada M. Hassan[1](✉), Safwat Hamad[2], and Khaled Mahar[3]

[1] College of Computing and Information Technology, Arab Academy for Science and Technology, Cairo, Egypt
nadamahmoud@aast.edu

[2] Faculty of Computer and Information Sciences, Ain Shams University, Cairo, Egypt

[3] College of Computing and Information Technology, Arab Academy for Science and Technology, Alexandria, Egypt

Abstract. Breast cancer is the most publicized cancer that hits women around the world. It's considered as the second cause of death among females. Early detection helps a lot in increasing the survival rate, and the probability of recovery from this disease. The mammogram is the main screening modality that is used regularly for breast cancer diagnosis. The accurate interpretation of the mammogram is very important for mass detection and diagnosis. The rapid evolution of deep learning is contributing to introduce more accurate systems that can act as a second opinion for the radiologists, and accordingly, this can help in providing an accurate diagnosis. In this paper, we propose a model for mass detection and classification based on You Look Only Once (YOLO)v4. We designed the experiment to investigate the performance of different augmentation techniques using YOLOv4 including mosaic that was introduced by YOLOv4. Furthermore, in the preprocessing phase, the images were reconstructed to be in a multichannel format which enhanced the detection accuracy by almost $\simeq$ 10%. The model was evaluated with the usage of different combinations of augmentation techniques (mosaic, mix-up, and conventional augmentation). The experiments were conducted on the INbreast and MIAS datasets, the results of INbreast showed that mosaic with YOLOv4 achieved the best results with mAP (mean average precision), precession, and recall of almost $\simeq$ 99.5%, 98%, and 94% respectively for detection. In addition, the proposed model achieved AP of 99.16% and 99.58% for classifying the detected masses as benign and malignant respectively. Furthermore, the best results on MIAS achieved mAP, precession, and recall of 95.28%, 93%, and 90% respectively. Finally, our methodology showed competitive performance compared to other similar studies.

Keywords: Mammogram · Mass detection · CAD system · INbreast · YOLOv4 · Mosaic

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2022
O. Gervasi et al. (Eds.): ICCSA 2022, LNCS 13375, pp. 544–559, 2022.
https://doi.org/10.1007/978-3-031-10522-7_37

1 Introduction

According to the Global Observatory Cancer (GLOBOCAN) estimates for 2020 [1], breast cancer is the most diagnosed cancer with 2.3 million new cases which represents 11.7% of the whole newly diagnosed cancer cases. It's considered as the first cancer type that causes death with 15.5% of the whole deaths from cancer among the females in the world. The clinical studies proved that the early detection of breast cancer can enhance the survival rate [2] and increase the chances of early recovery from this cancer type. Mammogram is one of the screening modalities that is mostly used as a regular screening tool and for early breast cancer diagnosis. It provides to some extent a good quality image that is produced through a low dose of x-rays.

The accurate diagnosis mainly needs a well-experienced radiologist, however, with a large number of cases that need examination this can lead to the wrong diagnosis. Recently, clinical trials proved that the Computer-Aided (Detection/Diagnosis) CAD systems can act as a second opinion for radiologists, and also it can be used to reduce the efforts of the manual annotation of the mammographic images. The importance of the existence of such systems that can perform accurately in detecting the masses, especially the small ones is increasing. The deep learning-based object detection models showed substantial success in achieving high accuracy with good computational power.

YOLO [3] is considered as the state of art model among the deep learning-based object detection models, as it surpassed other models such as the regional-based Convolutional Neural Network (RCNN) and Single Shot multi-box Detector (SSD) with competing performance at both accuracy and speed. There are four versions of YOLO which are YOLO v1 [3], YOLO v2 [4], YOLOv3 [5] and the recent one is YOLOv4 [6] which has many enhancements that affect positively in terms of accuracy. YOLOv4 enhanced the average precision by almost 10% more than YOLOv3 on the benchmark dataset Microsoft Common Objects in Context (MS COCO).

In this work, the model is designed for mass detection and classification based on YOLOv4. The paper aims to investigate the impact of the newly introduced augmentation techniques mosaic and mix-up on detecting the masses in the mammographic images and how these techniques will affect the performance of the model. Furthermore, the effect of converting the images into multichannel images on the performance of the model was investigated.

The main contribution of this paper can be summarized as the following:

1. Enhance the appearance of the breast masses and tissues by reconstructing the mammographic images into multi-channel-colored images.
2. Evaluate the impact of enhancing the appearance of the breast masses and tissues on the performance of the YOLOv4 on mammographic images.
3. Utilizing and evaluating the impact of using Mosaic and Mix-up with conventional augmentations and with different combinations.
4. Evaluate the generalization of the model with unseen images from other datasets that were obtained from different scanners.

The paper is organized as the following; Sect. 2 presents the related work, then Sect. 3 gives an overview of the used methods in this research including the used dataset

in the experiments and the implementation details. Then Sect. 4 presents the experimental results and provides a discussion of these results. Finally, Sect. 5 contains the conclusion and directions for future work.

2 Related Work

In the last few years, efforts increased towards developing more accurate CAD systems that can be used clinically. The leap that occurred in the object detection and the computer vision fields due to the appearance of deep learning and transfer learning contributes hugely to introducing more reliable systems.

The main advantage of deep learning is the automated extraction of the features that the network learned through training. Furthermore, transfer learning helped a lot to overcome the problem of the need for large datasets for deep learning, in addition, it reduces the training time.

Researchers are still exploiting the benefits of deep learning and transfer learning with medical images to introduce reliable CAD systems. In [7], Al-antari et al. introduced a breast cancer CAD system based on YOLO to detect the masses in mammograms. Their proposed system was fully automated in the phases of mass detection, segmentation, and classification. YOLO was used to both detect and localize the masses; Full Resolution Convolutional Network (FRCN) was used to segment the detected masses and finally, these segmented masses were classified into benign and malignant through a pre-trained CNN with AlexNet architecture. Their experiment was conducted on the INbreast dataset and their model achieved an accuracy of 98.96% for detection, 92.97% for segmentation, and 95.64% for classification. Moreover, in [8], they modified the segmentation and classification phases in the same model they introduced in [7]. The model achieved an accuracy of 97.27%, and 92.97% for detection and segmentation respectively. For classification, they used InceptionResNet-V2 with an average overall accuracy of 95.32%.

Ribli et al. [9] introduced a model based on Regional Convolutional Network R-CNN for localizing the masses and classifying them. They investigated the impact of the different resolutions on the performance of the model, their experiments proved that the higher resolution images give better results. The model detected almost 90% of the malignant masses in the INbreast dataset and achieved an AUC of 95% for classification.

In [10], Agarwal et al. proposed a CAD system based on Faster R-CNN for mass detection and classification. They conducted their experiments on three datasets INbreast, OPTIMAM, and private dataset. Their model achieved sensitivity ranging from 95% to 71% and a specificity of 70% for detection. In their work, they used images that were obtained from different scanners.

Peng et al. [11] modified the backbone of the Faster R-CNN model; they used a multiscale feature pyramid network between the output layers of the backbone, this modification was mainly done to enhance the small mass detection in their experiments. Their model achieved a True Positive Rate (TPR) of 0.934 for CBIS-DDSM and o.955 for INbreast datasets.

These reviewed researches utilized various deep learning-based object detection models in their introduced models for mammographic mass detection and classification and these approaches achieved promising results. However, the majority of these

researches didn't focus too much on enhancing the quality of the images themselves before the training. Furthermore, the detection of the small masses and the detection of the masses generally within the dense breast tissues are still considered as open challenges [12]. This paper mainly contributes to addressing these problems by firstly enhancing the visibility of the breast tissues in the mammographic image by reconstructing the images in a multichannel format. Accordingly, this can increase the discriminant features the model can learn. Secondly, utilizing a new augmentation technique named mosaic increased the complexity of the model at the training phase and improved the model's ability at detecting masses at smaller and different scales.

3 Methodology and Background

In this work, YOLOv4 has been used for mass detection and classification in mammogram images. Also, the mosaic and mix-up data augmentation techniques were used to investigate their effects on the detection of the existing masses in the breast.

3.1 Background

You Look Only Once (YOLO). YOLO works through a single neural network trained in an End-to-End fashion which takes an image as input in a single shot and predicts bounding boxes and class labels for each bounding box directly. The YOLO algorithm splits the input image into (s x s) grid cells. The cell is considered to be responsible for detecting an object if the center of this object falls in this cell. Each cell predicts the location of a fixed number of bounding boxes as well as their Confidence Score. The coordinates of the bounding box are represented by 4 values (x, y, w, h) where x and y are representing the center point of the box. On the other hand, the confidence score is calculated as shown in Eq. (1), where the Intersection over Union IoU is representing the overlapped area between the predicted bounding box and the ground truth. If the box contains an object, it predicts a probability of this object belonging to every class Ci, where i = 1,…, K as K represents the number of the classes in the dataset which is 2 in our experiments. The output for each bounding box is represented by a vector composed of the bounding box coordinates and the probability of the class that the object belongs to.

$$Confidence\ score = prob(containing\ an\ object)\ x\ IoU\ (predicted\ \ bb,\ \ ground\ \ truth) \tag{1}$$

The basic building block of the YOLO detector family is composed of three parts; backbone, neck, and head. The backbone extracts the feature maps from the input image and this part can be VGG16, ResNet-50, SpineNet, EffecientNet-B0/B7, CSPResNeXt50, or CSPDarknet53 which is recommended in the paper of YOLOv4 [6] based on many experiments. The neck is responsible for enhancing the feature discriminability and robustness via adding extra layers between the backbone and the head to enrich the information that feeds into the head. Finally, the head which is also named dense prediction is mainly responsible for locating the bounding boxes and classifying the objects inside these boxes.

YOLOv4 utilizes the Cross Stage Partial (CSP) connections with the Darknet-53 as a backbone for feature extraction. While in the neck they used a modified Path Aggregation Network (PAN) [13], a modified Spatial Attention Module (SAM) [14], and Spatial Pyramid Pooling (SPP) [15]. In addition, they introduced in YOLOv4 bag of freebies and a bag of specials that are used for the backbone and the detector to enhance the performance of the network. Figure 1 illustrates the architecture of YOLOv4.

In this work, the new data augmentation technique that was introduced in YOLOv4 which is called Mosaic was used, in addition to Mix-up data augmentation to investigate how these techniques can affect the performance of the model at detecting the masses in the mammograms and classifying them into benign and malignant. Also, the CSPDarknet53 was used as a backbone for extracting the discriminant features from the input mammogram images to detect two types of masses benign and malignant.

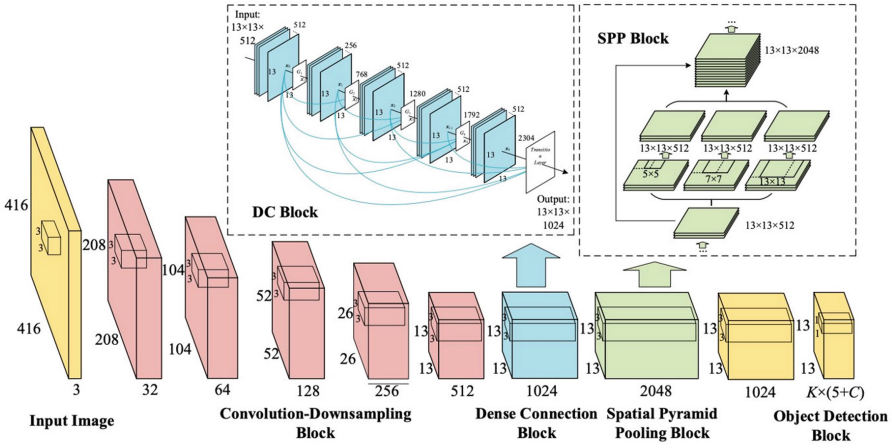

Fig. 1. YOLOv4 architecture [16]

Mosaic Data Augmentation. In this paper, mosaic augmentation was used to increase the size of the training set, in addition to gaining the benefit from this technique in identifying the small masses. Mosaic is one of the newest techniques for data augmentation which was introduced in [6]. The idea behind this algorithm is to combine four of the training images into one image with the original size of the single image, so the objects in each image of the four are appearing on a smaller scale. Accordingly, this augmentation technique can enhance the performance of the model in identifying the objects at small scales. Furthermore, with this technique, there's no need for a large mini-batch size through training. The algorithm mainly depends on locating the four corners of the image (the upper left, upper right, lower left, and lower right) to apply the processing of the data augmentation respectively on the four corners. This approach increases the learning ability of the model four times of learning the original image; and accordingly, that enhances the performance of the model.

Mix-up Data Augmentation. In addition to the Mosaic, we used the Mix-up augmentation technique which was introduced in 2018 [17]. Mix-up blends the features of two images in one image with their corresponding labels from the training set. The mix-up algorithm can be represented in the following formula:

$$\hat{x} = \lambda x_i + (1 - \lambda)x_j \tag{2}$$

$$\hat{y} = \lambda y_i + (1 - \lambda)y_j \tag{3}$$

In Eq. (2) x_i and x_j are the raw input vectors of the images while y_i and y_j in Eq. (3) represent the one hot-encoders for the labels of these images. These pairs (x_i, y_i) and (x_j, y_j) are randomly selected from the training set. Accordingly, from Eqs. (2) and (3), it can be said that Mix-up expands the training distribution by factoring in the prior knowledge that linear interpolations of feature vectors should result in linear interpolations of the targets.

3.2 Implementation

Dataset. The experiments were done using three different datasets INbreast, MIAS, and CBIS-DDSM.

INbreast. The dataset [18] contains 410 mammograms for 115 cases and 90 cases were diagnosed with cancer. The dataset images were acquired using Siemens MammoNovatio scanner which provided Full Field Digital Mammograms (FFDM) that have high quality than the Digital Mammogram (DM), the dataset includes different breast abnormalities. The dataset has 107 mammograms containing masses (benign and malignant) which were used in the experiments. In the experiments that were done in this work, the masses that have BI-RADS scores of 1, 2, and 3 were considered benign, while the ones that have a score of 4, 5, and 6 were considered malignant. The images are provided in DICOM format, also the dataset contains images in both views Medio Lateral Oblique (MLO) and Carnio Caudal (CC) view.

MIAS. MIAS [19] is a mammographic dataset that is composed of 322 mammography images in Portable Gray Map (PGM) format. The dataset contains only images with MLO view only; the images are at 50-micron resolution. The type of the images is Screening Film Mammography (SFM), and the ground truth data are associated with the dataset. Different abnormalities are included in the dataset in addition to normal images; 55 images with masses were used from this dataset in the experiments.

CBIS_DDSM. Some images were used in the experiments from the Curated Breast Images Subset of Digital Database for Screening Mammography (CBIS_DDSM) [20]. The dataset contains 10239 images for 1566 patients with images containing a total of 1696 tumors. However, the dataset is considered as the largest publicly available mammographic dataset, the quality of the images isn't good as these images are in DM format. The dataset includes both views MLO and CC, all the images are in the DICOM format. The ROIs are provided with the dataset in addition to CSV files that describes the pathological information for the patients. The images of this dataset were obtained from four different scanners DBA, LUMISYS, and HOWTEK (A/D).

Data Preprocessing and Augmentation. The images were converted from DICOM to JPEG in INbreast and from PMG to JPEG, then these images were scaled into 8 bits. The images were cropped to extract the breast region and accordingly minimize the area of the background; so, the Gaussian filter and Otsu's thresholding [21, 22]were used. After that, the images were normalized to normalize the pixels intensity distribution of the ROI of the breast region to be in the range of (0–255). The Contrast Limited Adaptive Histogram Equalization (CLAHE) [23]was applied at two different clip limits. CLAHE is an improved version of the Adaptive Histogram Equalization (AHE), and this technique aims to decrease the amplified noise by distributing equally the histogram part that surpass the clip limit.

The mammography images are considered as grayscale images that consist of black, white and shades of gray and this make these images challenging at learning their features, however, the edges and the texture features are affecting hugely the mass identification. Accordingly, to enhance the mass appearance in the mammogram and the contrast of the breast region, a new multi-channel image for the mammographic image was reconstructed by merging three single-channel images which are (normalized, and the other two images after CLAHE). This step enhanced the appearance of the masses and the breast tissues as shown in Fig. 2.

Due to the small size of the INbreast and MIAS datasets, traditional data augmentation techniques were applied on both training and testing sets to increase the size of the dataset by augmenting the original images with processed ones using operations of (blur, random brightness contrast, shift scale rotation, random rotate, transpose, multiplicative noise, and vertical flip) as shown in Fig. 3. The INbreast dataset increased to be 801 mammographic images and MIAS became 382 images after this step.

The mosaic and mix-up augmentation techniques were applied only on the training dataset through enabling them in the network settings to investigate their impact on the performance of the model in detecting the masses. Figure 4 shows samples for the mosaic augmentation on the INbreast mammographic images.

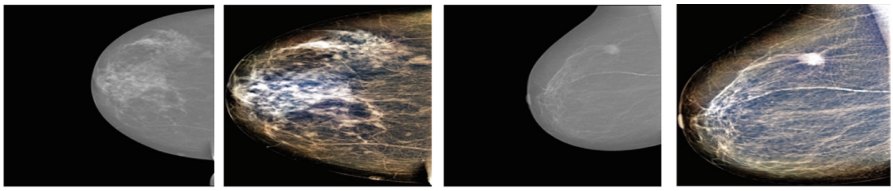

Fig. 2. Images from INbreast before preprocessing (a,c) and after preprocessing (b,d).

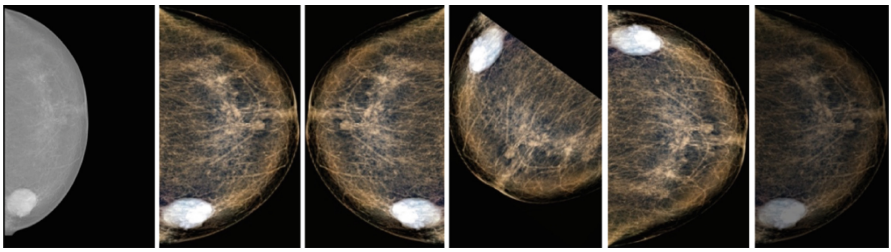

Fig. 3. A mammogram from INbreast before and after preprocessing in addition to samples of its corresponding images after applying conventional data augmentation techniques (flipping, brightness, and rotation).

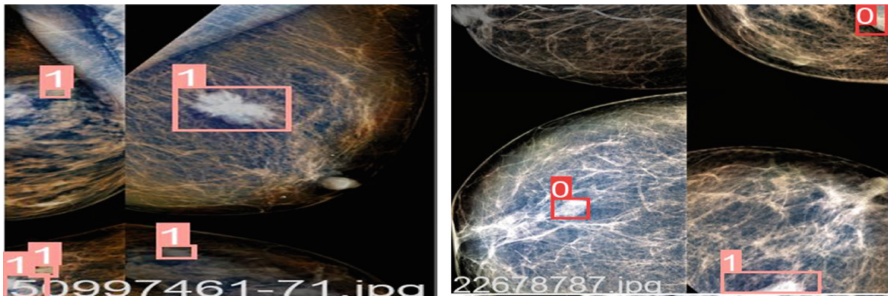

Fig. 4. Images from INbreast with mosaic data augmentation with annotations where 0 refers to benign and 1 to malignant.

Implementation Environment. The experiments were done on a workstation with an Intel® Core i7-11700k processor with 32 GB RAM, 3.200 GHz frequency, and NVIDIA GeForce RTX 3080Ti GPU with 12 vRAM. The proposed CAD system was implemented using C++ and Python 3.8 on the Windows 10 operating system.

Implementation Design. The images were split randomly into a training set and testing set with a ratio of 80%–20% respectively. The experiment was designed to evaluate the effectiveness of the multichannel images on the detection performance. In addition, it evaluated the performance of the mosaic and the mix-up augmentation techniques using the YOLOv4 network to investigate the impact of using such techniques on detecting the masses and classifying them into benign and malignant.

Some parameters have to be settled for the Yolov4 network, the value of these parameters is changed based on the domain and the used dataset in the experiment. For this work these parameters have been set as the following for both datasets INbreast and MIAS:

- The input image size for the network is 416 x 416.
- The number of classes is 2 which represents benign and malignant.
- Learning Rate (LR) is 0.001 based on different experiments that show that this value gives better results.

- The Scales that were used in the experiments to change the value of the LR are 0.1,0.1.
- The steps that represent the number of iterations and act as checkpoints for applying the scales on the learning rate, were set to 3200, and 3600 respectively. This means that the LR will be adjusted by multiplying its value by 0.1 after 3200 batches and multiplying it again by 0.1 after 3600 batches.
- Max_batches (number of iterations) was set to 4000 based on the equation mentioned in [3] (number of classes x 2000).
- The number of filters was set to 21 which is set to (number of classes + 5) × 3

The experiments were done over the following combinations:

1. YOLOv4 + Conventional Machine Learning + Mosaic + Normalized images
2. YOLOv4 + Conventional Machine Learning + Mosaic + Multichannel images
3. YOLOv4 + Conventional Machine Learning + Multichannel images
4. YOLOv4 + Conventional Machine Learning + Mixup + Multichannel images
5. YOLOv4 + Conventional Machine Learning + Mosaic + Mixup + Multichannel images

Evaluation Metrics. In this work, the performance of the model in the conducted experiments was evaluated through various metrics. These metrics are Intersection over Union (IoU), mean Average Precision (mAP), F1-score, True Positive (TP), False Positive (FP) precision, and recall for detection. Furthermore, the False Negative Rate (FNR) was calculated for mass detection according to Eq. (4) because of its importance in diagnosing diseases. For classification, the Average Precision (AP), TP, and FP were used.

$$FNR = \frac{False\ Negative}{True\ Positive + False\ Negative} \tag{4}$$

The mAP is calculated to evaluate the model accuracy for a set of object detections compared to its ground truth annotations. It's mainly calculated based on taking the average of the Average Precision (AP) for each class (benign and malignant).

$$\text{mAp} = \frac{1}{N}\sum_{i=1}^{N}(AP)_i,\ where\ N\ is\ the\ number\ of\ classes. \tag{5}$$

IoU is describing to what extent the ground truth is overlapping with the predicted bounding box, and it can be calculated as the following:

$$\frac{A \cap B}{A \cup B},\ where\ A\ is\ the\ predicted\ bounding\ box\ and\ B\ is\ the\ ground\ truth\ box. \tag{6}$$

For each prediction, the IoU is calculated and then thresholded to a value between (0.5 – 0.9). In our experiments, we used the value of 0.5 as a threshold where the detected masses are considered as a correct detection (TP) only if the IoU >= 0.5. In addition, the precision, sensitivity, and F1-score are calculated according to the following Eqs. (7), (8), and (9) respectively.

The precision here is describing the percentage of the correct predictions, in other words, the percentage of the true positives regarding the total detected positive cases. While the recall determines the percentage of the ability of the model at detecting the positive cases regarding the total cancer cases.

Average precision was used to describe the area under the curve (AUC) for the precision and recall curve. The AP is calculated to evaluate the accuracy of the model predictions for each class.

$$Precision = \frac{True\ Positive}{True\ Positive + False\ Positive} \tag{7}$$

$$Recall\ (senstivity) = \frac{True\ Positive}{True\ Positive + False\ Negative} \tag{8}$$

$$F1 - score = 2 \times \frac{precision \times recall}{precision + recall} \tag{9}$$

4 Experimental Results and Discussion

This section illustrates and discusses the results of the conducted experiments. Two experiments are conducted to assess the performance of the model before and after the step of converting the images into multichannel images. In addition, another four experiments were performed to evaluate the impact of the different augmentation techniques on the model performance, these techniques were applied to the training set only. The first trial was applied with the mosaic, the second one with the mix-up, while in the third trial both mosaic and mix-up were applied, and finally, the last trial was done with just only the traditional augmentation techniques that were applied to the whole dataset as mentioned in Sect. 3.2.

Table 1 shows that the multichannel images enhanced significantly the mAP of the detection by almost $\simeq 10\%$ and the sensitivity by $\simeq 7\%$, moreover, it improved the classification accuracy for each class. It can be noticed from Table 1 that the number of TP increased while the number of FP and FN decreased which means that this step enhanced the detection of the truly existing masses. Based on these results, it can be demonstrated that preprocessing can overcome the problem of small datasets. Deep learning models need large datasets to provide higher accuracy, however, the experiments showed that enhancing the images can provide more discriminant features and this can ease the learning process of the model.

Considering the results from Table 2 and 3 mosaic augmentation achieved the best mAP among the other techniques; furthermore, the mosaic improved the number of TP and FN of mass detection, whether for INbreast or MIAS, as mosaic allowed the model to learn the masses at different scales and different contexts. The mix-up and mosaic together showed competitive performance; They achieved almost the same sensitivity (recall) score as mosaic however it showed lower mAP than mosaic only. Possibly this happens because of mix-up, as it changes the pixel values of the mammogram image when mixing up two images; This perhaps affects the appearance of the mass shape and the surrounding tissues and, accordingly, the detection task.

Figure 5, 6, Table 4, and 5 showed the results of testing unseen images from different datasets that were digitized with different scanners using the model that was trained on INbreast only; these images are from CBIS-DDSM which provides mammographic images in DM format and MIAS that composed of (SFM) images. In the experiments, 21 images with masses from MIAS were used as unseen images for testing the model that was trained on INbreast with mosaic.

FFDM has better accuracy in detecting breast cancer than SFM [24] and this obviously can be demonstrated from the results in Table 2 and 3. In addition, this can explain the results that showed in Table 4; The model which was trained with conventional + mosaic combination using INbreast dataset only succeeded in testing some unseen images from MIAS dataset for detecting masses in them. The testing results of this experiment showed mAP of 71.93%. It is important to note that MIAS dataset is composed of SFM images while INbreast is composed of FFDM images only and accordingly this can clarify this result as the model was trained on FFDM images and tested on SFM images which have lower quality. However, the model with conventional + mosaic combination when trained and tested on MIAS showed good results as shown in Table 3.

The training with the mosaic combination and the conventional augmentation allowed for better generalization rather than the other combinations, and also it increased the learning ability of the model. The main advantage of mosaic especially in these experiments over the other evaluated techniques is mainly in producing new images at different and smaller scales which allowed a variation of the different mass sizes that may exist in the mammogram and this increased the ability of the model at detecting the masses with variant sizes that range from very small to very large. The variation of the mass sizes and shapes in the medical images specifically 2D images makes the detection and localization task challenging.

The mix-up with conventional augmentation combination showed better results than the conventional augmentation only due to generating new images with linearly mixing up features of two mammographic images; this combination showed more stability and reduces the bad predictions rather than using the conventional augmentation only. Conventional augmentation techniques may resolve the problem of overfitting however, it doesn't improve the generalization too much compared to Mosaic and Mix-up. Furthermore, the Mosaic and Mix-up outperformed the conventional augmentation in terms of detection and distinction between benign and malignant masses as shown in Table 2 and 3.

It can be deduced from these results that the proposed model with combination of mosaic and multichannel images has a good level of generalization and robustness. Furthermore, the model was able to differentiate between the benign masses and malignant masses with AP of 99.16% and 99.58% respectively on INbreast; 91.87% and 98.69% for MIAS. The model showed competitive performance compared to other recent similar studies in terms of accuracy and inference time as shown in Table 6.

Table 1. INbreast detection results of the experimental trials for the model with the normalized images and mosaic vs. multichannel images with mosaic in terms of mAP, F1-score, number of TP, number of FP, Recall, Precision, and FNR. In addition, the results for distinguishing Benign (B) and Malignant (M) masses in terms of AP, TP, and FP.

#	Technique	mAP(%)@IoU = 0.5	F1-score (%)	TP	FP	FN	Recall (%)	Precision (%)	FNR	B-AP (%)	M-AP (%)	TP-FP (B)	TP-FP (M)
Trial 1	Normalized images + mosaic	88.89%	90%	74	5	11	87%	94%	0.129	81.85%	95.94%	21–0	53–5
Trial 2	Multichannel images + mosaic	99.37%	96%	80	2	5	94%	98%	0.058	99.61%	99.58%	24–0	56–2

Table 2. INbreast detection results of the four experimental trials in terms of mAP, F1-score, TP, FP, Recall, Precision, and FNR. In addition, the results for distinguishing the Benign and Malignant masses in terms of AP, TP, and FP.

#	Augmentation technique	mAP(%)@IoU = 0.5	F1-score (%)	TP	FP	FN	Recall (%)	Precision (%)	FNR	B-AP (%)	M-AP (%)	TP-FP (B)	TP-FP (M)
Trial 1	Conventional + Mosaic	99.37%	96%	80	2	5	94%	98%	0.058	99.16%	99.58%	24–0	56–2
Trial 2	Conventional + Mix-up	83.37%	84%	71	14	14	84%	84%	0.164	75.86%	90.87%	18–5	53–9
Trial 3	Conventional + Mosaic + Mix-up	95.14%	96%	81	3	4	95%	96%	0.047	91.19%	99.09%	24–1	57–2
Trial 4	Conventional	75.25%	79%	61	8	24	72%	88%	0.282	66.11%	84.39%	13–2	48–6

Table 3. MIAS detection results of the four experimental trials in terms of mAP, F1-score, TP, FP, Recall, Precision, and FNR. In addition, the results for distinguishing the Benign and Malignant masses in terms of AP, TP, and FP.

#	Augmentation technique	mAP(%)@IoU = 0.5	F1-score (%)	TP	FP	FN	Recall (%)	Precision (%)	FNR	B-AP (%)	M-AP (%)	TP-FP (B)	TP-FP (M)
Trial 1	Conventional + Mosaic	95.28%	92%	84	6	9	90%	93%	0.096	91.87%	98.69%	51–4	33–2
Trial 2	Conventional + Mix-up	92.34%	84%	79	17	14	85%	82%	0.150	90.11%	94.57%	46–1	33–16
Trial 3	Conventional + Mosaic + Mix-up	93.20.65%	93%	83	2	10	89%	98%	0.107	89.33%	97.06%	51–2	32–0
Trial 4	Conventional	65.79%	64%	56	26	37	60%	68%	0.397	50.74%	80.83%	26–10	30–16

Table 4. Results of testing on 21 unseen images from MIAS (B and M refer to Benign and Malignant respectively) on the model that trained on INbreast with Mosaic.

mAP(%)@IoU = 0.5	F1-score (%)	TP	FP	FN	Recall (%)	Precision (%)	FNR	B-AP (%)	M-AP (%)	TP-FP (B)	TP-FP (M)
71.93%	70%	15	7	6	71%	68%	0.285	68.75%	75.11%	6–1	9–6

Table 5. Results of testing unseen images from CBIS_DDSM and MIAS dataset on the model that trained on INbreast with Mosaic.

#	Dataset	Image-id	Detection
a	CBIS-DDSM	P_00118_RIGHT_CC	1 mass (Malignant with $\simeq$ 97%)
b		P_00332_LEFT_MLO	2 masses (Benign with $\simeq$ 37%and94%)
c		P_00265_RIGHT_CC	1 mass (Malignant with $\simeq$ 99%)
d		P_00273_LEFT_CC	1 mass (Benign with $\simeq$ 96%)
e	MIAS	Mdb023	1 mass (Malignant with $\simeq$ 79%)
f		Mdb142	1 mass (Benign with $\simeq$ 92%)
g		Mdb202	1 mass (Malignant $\simeq$ 100%)
h		Mdb010	1 mass (Benign $\simeq$ 92%)

Table 6. Comparison between some of the recent works based on YOLO and the proposed model

Reference	Method	Dataset	Detection Accuracy	Inference time/image (sec)
Aly et al.[25]	YOLOv3	INbreast, DDSM	89.5%	0.009
Al-antari et al.[7]	YOLO	INbreast	98.96%	3
Baccouche et al.[26]	YOLO based Fusion models	INbreast, CBIS-DDSM	98.1%	0.58, 0.55
Platania et al.[27]	YOLO based ROI Classifier	CBIS-DDSM	90%	–
Al-antari et al.[28]	YOLO	INbreast	97.27%	0.025
Proposed Model	YOLOv4 + Mosaic + Multichannel images	INbreast	99.37%	0.0112
		MIAS	95.28%	0.0111

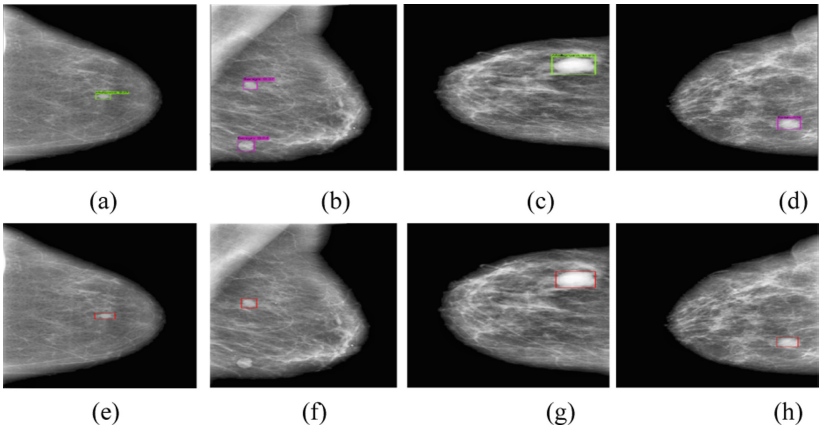

Fig. 5. Results of the model on unseen images (a), (b), (c), and (d) images from CBIS-DDSM and their corresponding ground-truth (e), (f), (g), and (h).

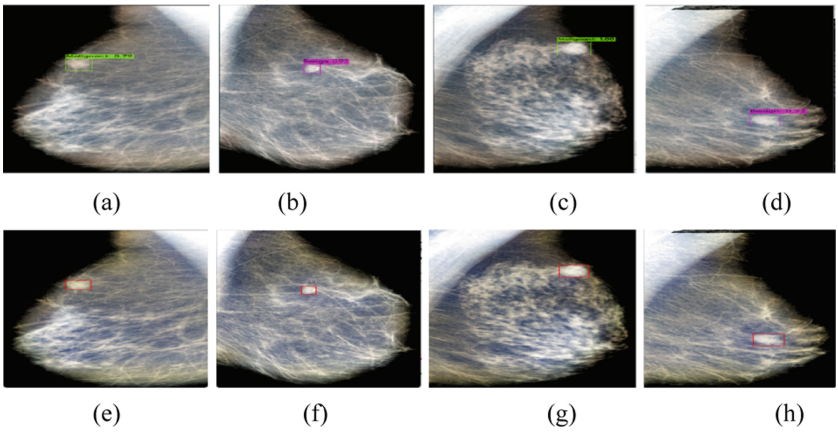

Fig. 6. Results of the model on unseen images (a), (b), (c), and (d) images from MIAS and their corresponding ground-truth (e), (f), (g), and (h)

5 Conclusion

In this paper, we proposed a CAD system for breast mass detection and classification, this model is based on YOLOv4. The model utilized the mosaic data augmentation which was applied only to the training set after applying conventional augmentation techniques to the whole dataset (training and testing). Furthermore, the images were converted to multichannel images which improved the visibility of the breast tissues and the masses; this accordingly enhanced the mAP by almost 10% however the model was working on 416 x 416 input size. The proposed model achieved mAP of 99.37% with a False Negative Rate (FNR) of 0.06; moreover, the model was able to detect masses with different sizes when tested on unseen images from other datasets which means that

the model can be generalized. In addition, the model enhanced the classification results for each mass type (benign and malignant). In future work, we are intended to extend this work to detect other abnormalities in the mammograms and to evaluate this work on other datasets.

References

1. Global Cancer Statistics: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. Enhanced Reader (2020). Accessed 28 Feb 2022
2. Hadjiiski, L., Sahiner, B., Chan, H.P.: Advances in CAD for diagnosis of breast cancer. Curr. Opin. Obstet. Gynecol. **18**, 64 (2006). https://doi.org/10.1097/01.GCO.0000192965.29449.DA
3. Redmon, J., Divvala, S., Girshick, R., Farhadi, A.: You Only Look Once: Unified, Real-Time Object Detection
4. Redmon, J., Farhadi, A.: YOLO9000: Better, faster, stronger. In: Proceedings of the 30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017, pp. 6517–6525 (2016). https://doi.org/10.1109/CVPR.2017.690
5. Redmon, J., Farhadi, A.: YOLOv3: An Incremental Improvement (2018)
6. Bochkovskiy, A., Wang, C.-Y., Liao, H.-Y.M.: YOLOv4: Optimal Speed and Accuracy of Object Detection (2020)
7. Al-antari, M.A., Al-masni, M.A., Choi, M.T., Han, S.M., Kim, T.S.: A fully integrated computer-aided diagnosis system for digital X-ray mammograms via deep learning detection, segmentation, and classification. Int. J. Med. Inform. **117**, 44–54 (2018). https://doi.org/10.1016/J.IJMEDINF.2018.06.003
8. Al-antari, M.A., Al-masni, M.A., Kim, T.-S.: Deep learning computer-aided diagnosis for breast lesion in digital mammogram. In: Lee, G., Fujita, H. (eds.) Deep Learning in Medical Image Analysis. AEMB, vol. 1213, pp. 59–72. Springer, Cham (2020). https://doi.org/10.1007/978-3-030-33128-3_4
9. Ribli, D., Horváth, A., Unger, Z., Pollner, P., Csabai, I.: Detecting and classifying lesions in mammograms with deep learning. Sci. Rep. **8**(1), 1–7 (2018). https://doi.org/10.1038/s41598-018-22437-z
10. Agarwal, R., Díaz, O., Yap, M.H., Lladó, X., Martí, R.: Deep learning for mass detection in full field digital mammograms. Comput. Biol. Med. **121**, 103774 (2020). https://doi.org/10.1016/J.COMPBIOMED.2020.103774
11. Peng, J., Bao, C., Hu, C., Wang, X., Jian, W., Liu, W.: Automated mammographic mass detection using deformable convolution and multiscale features. Med. Biol. Eng. Comput. **58**, 1405–1417 (2020). https://doi.org/10.1007/S11517-020-02170-4/FIGURES/7
12. Hassan, N.M., Hamad, S., Mahar, K., Tools, M.: Mammogram breast cancer CAD systems for mass detection and classification: A review. Multimedia Tools Appl. **81**, 1–33 (2022). https://doi.org/10.1007/S11042-022-12332-1
13. Liu, S., Qi, L., Qin, H., Shi, J., Jia, J.: Path aggregation network for instance segmentation. In: Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pp. 8759–8768 (2018). https://doi.org/10.1109/CVPR.2018.00913
14. Woo, S., Park, J., Lee, J.-Y., Kweon, I.S.: CBAM: Convolutional block attention module. In: Ferrari, V., Hebert, M., Sminchisescu, C., Weiss, Y. (eds.) ECCV 2018. LNCS, vol. 11211, pp. 3–19. Springer, Cham (2018). https://doi.org/10.1007/978-3-030-01234-2_1
15. He, K., Zhang, X., Ren, S., Sun, J.: Spatial pyramid pooling in deep convolutional networks for visual recognition. In: Fleet, D., Pajdla, T., Schiele, B., Tuytelaars, T. (eds.) ECCV 2014. LNCS, vol. 8691, pp. 346–361. Springer, Cham (2014). https://doi.org/10.1007/978-3-319-10578-9_23

16. Huang, Z., Wang, J.: DC-SPP-YOLO: Dense Connection and Spatial Pyramid Pooling Based YOLO for Object Detection
17. Zhang, H., Cisse, M., Dauphin, Y.N., Lopez-Paz, D.: Mixup: Beyond Empirical Risk Minimization
18. 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**, 236–248 (2012). https://doi.org/10.1016/J.ACRA.2011.09.014
19. Mammographic Image Analysis Society (MIAS) database v1.21. https://www.repository.cam.ac.uk/handle/1810/250394. Accessed 26 Mar 2022
20. Lee, R.S., Gimenez, F., Hoogi, A., Miyake, K.K., Gorovoy, M., Rubin, D.L.: A curated mammography data set for use in computer-aided detection and diagnosis research. Sci. Data **4**(1), 1–9 (2017). https://doi.org/10.1038/sdata.2017.177
21. D'Haeyer, J.P.F.: Gaussian filtering of images: A regularization approach. Signal Process. **18**, 169–181 (1989). https://doi.org/10.1016/0165-1684(89)90048-0
22. Otsu, N.: Threshold selection method from gray-level histograms. IEEE Trans. Syst. Man Cybern. SMC **9**, 62–66 (1979). https://doi.org/10.1109/TSMC.1979.4310076
23. Pizer, S.M., et al.: Adaptive histogram equalization and its variations. Comput. Vis. Graph. Image Process. **39**, 355–368 (1987). https://doi.org/10.1016/S0734-189X(87)80186-X
24. Song, S.Y., Park, B., Hong, S., Kim, M.J., Lee, E.H., Jun, J.K.: Comparison of digital and screen-film mammography for breast-cancer screening: A systematic review and meta analysis. J. Breast Cancer **22**, 311–325 (2019). https://doi.org/10.4048/jbc.2019.22.e24
25. Aly, G.H., Marey, M., El-Sayed, S.A., Tolba, M.F.: YOLO based breast masses detection and classification in full-field digital mammograms. Comput. Methods Programs Biomed. **200**, 105823 (2021). https://doi.org/10.1016/J.CMPB.2020.105823
26. Baccouche, A., Garcia-Zapirain, B., Olea, C.C., Elmaghraby, A.S.: Breast lesions detection and classification via YOLO-based fusion models. Comput. Mater. Continua **69**, 1407–1425 (2021). https://doi.org/10.32604/CMC.2021.018461
27. Platania, R., Shams, S., Yang, S., Zhang, J., Lee, K., Park, S.-J.: Automated breast cancer diagnosis using deep learning and region of interest detection (BC-DROID). In: Proceedings of the 8th ACM International Conference on Bioinformatics, Computational Biology, and Health Informatics (2017). https://doi.org/10.1145/3107411
28. Al-antari, M.A., Han, S.M., Kim, T.S.: Evaluation of deep learning detection and classification towards computer-aided diagnosis of breast lesions in digital X-ray mammograms. Comput. Methods Programs Biomed. **196**, 105584 (2020). https://doi.org/10.1016/J.CMPB.2020.105584