

Gravity Network for end-to-end small lesion detection

Ciro Russo^a, Alessandro Bria^a and Claudio Marrocco^{a,*}

^aDepartment of Electrical and Information Engineering, University of Cassino and L.M., Via G. Di Biasio 43, 03043 Cassino (FR), Italy

ARTICLE INFO

Keywords:

Small lesion detection
Pixel-based anchor
Convolutional neural networks
Mammograms
Ocular fundus images

ABSTRACT

This paper introduces a novel one-stage end-to-end detector specifically designed to detect small lesions in medical images. Precise localization of small lesions presents challenges due to their appearance and the diverse contextual backgrounds in which they are found. To address this, our approach introduces a new type of pixel-based anchor that dynamically moves towards the targeted lesion for detection. We refer to this new architecture as GravityNet, and the novel anchors as gravity points since they appear to be “attracted” by the lesions. We conducted experiments on two well-established medical problems involving small lesions to evaluate the performance of the proposed approach: microcalcifications detection in digital mammograms and microaneurysms detection in digital fundus images. Our method demonstrates promising results in effectively detecting small lesions in these medical imaging tasks.

1. Introduction

Detection of small lesions in medical images has emerged as a compelling area of research, which holds significant relevance in medicine, especially in fields like Radiology and Oncology, when a timely disease diagnosis is essential [43]. Small lesions are primarily characterized by a limited size and can vary greatly in nature depending on their location and the involved tissue. In numerous real-world scenarios, the identification and classification of small lesions is a challenging and critical diagnostic process. For example, retinal microaneurysms are the earliest sign of diabetic retinopathy and are caused by small local expansion of capillaries in the retina [14]. In ischemic stroke imaging, early identification of small occlusion is crucial to initiate timely treatment [58]. In cancer diagnosis, many forms of cancer originate as small lesions before they grow and spread, such as breast calcifications, which are one of the most important diagnostic markers of breast lesions [46], or pulmonary nodules, which can be the first stage of a primary lung cancer [44]. The ability to early and accurately detect small lesions can make a difference in the treatment and prognosis of patients and have a substantial impact on patient health. Manual interpretation of medical images can be time-consuming and susceptible to human error, especially when the task involves of localization and identification of small lesions within the full image space [13].

There is a long tradition of research on automatic lesion detection methods [55]. Traditional image processing methods, such as thresholding, edge detection, and morphological operations, can be effective for detecting small lesions in images with clear and well-defined structures. However, these methods are limited by the presence of noise and variability in medical images. The use of Machine Learning,

and in particular Deep Learning, helps to enhance reliability, performance, and accuracy of diagnosing systems for specific diseases [54]. Actually, the first lesion detection system based on Convolutional Neural Networks (CNNs) was proposed back in 1995 to detect lung nodules in X-ray images [40]. However, only in the last ten years, CNNs have acquired great popularity thanks to their remarkable performance in computer vision [32], rapidly becoming the preferable solution for automated medical lesion detection [59, 17, 21, 22]. The reason for this success is due to the ability of learning hierarchical representations directly from the images, instead of using handcrafted features based on domain-specific knowledge. CNNs are able to build features with increasing relevance, from texture to higher order features like local and global shape [33].

A typical CNN architecture for medical image analysis is applied to subparts of an image containing candidate lesions or background. This means that the image is divided into patches of equal size and partially overlapping, and each patch is processed individually. The output image is formed by reassembling the individually processed patches [25]. Despite patch-based methods being widely used, they suffer from several problems, especially in the case of small lesion detection [7], where accurate detections requires both local information about the appearance of the lesion and global contextual information about its location. This combination is typically not possible in patch-based learning architectures [37], even with a multi-scale approach where the appearance of a small lesion can be missed. An alternative is to use anchoring object detection methods of computer vision [26], like RetinaNet [36], which can be adapted to be used in lesion detection problems [42]. These methods face difficulties when the objects to be detected are very small, mainly for two reasons: (i) lesions have an extremely small size compared to natural objects; (ii) lesions and non-lesions have similar appearances, making it difficult to detect them effectively [6].

We propose a novel one-stage end-to-end detector based on a new type of anchoring technique customised to small

*Corresponding author:

Phone: +39 0776 2993381, Fax: +39 0776 2994355

✉ ciro.russo@unicas.it (C. Russo); a.bria@unicas.it (A. Bria);

c.marrocco@unicas.it (C. Marrocco)

ORCID(s): 0009-0002-8751-8605 (C. Russo); 0000-0002-2895-6544 (A. Bria); 0000-0003-0840-7350 (C. Marrocco)

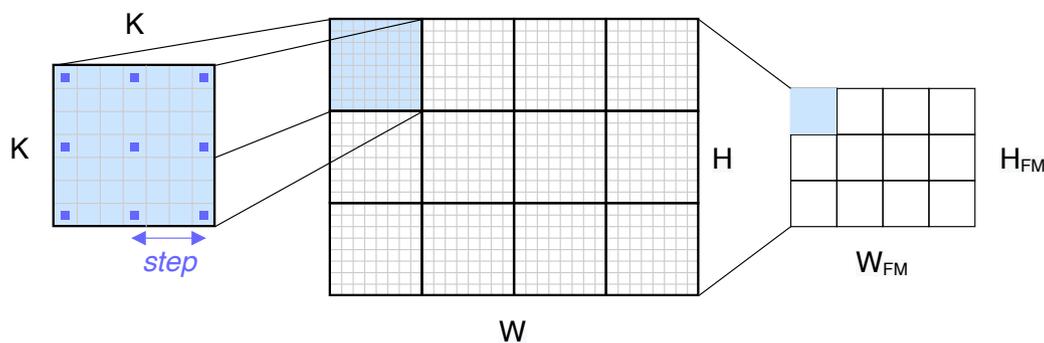


Figure 1: Gravity-points distribution: on the left, the feature grid of size $K \times K$; in the middle, the entire image $H \times W$; on the right, the feature map $H_{FM} \times W_{FM}$.

lesion detection in medical images. Differently from classical anchor methods, which make use of anchor boxes to capture scale and aspect ratio of specific classes of objects, the proposed anchor is pixel-based and moves towards the lesion to be detected. We named *GravityNet* this new architecture and *gravity points* the new anchors, because they are distributed over the whole image and seem to be “attracted” by hypothetical “gravitational masses” located in the centres of the lesions. Such gravitational anchoring reveals to be particularly effective when small lesions have to be detected in the whole image space. To evaluate the performance of the proposed approach, we focused on two small lesions: microcalcifications on digital mammograms and microaneurysms on digital fundus images. In both cases, the lesions occupy only few pixels within an image, resulting in limited features for them to be distinguished from the surrounding tissues. Thus, their accurate localization becomes a main challenge due to their appearance and to the heterogeneity of their contextual backgrounds.

The paper is organized as follows. Section 2 is a brief overview of object detection techniques in medical images and consequently of small lesion detection methods. Section 3 introduces the proposed method. Section 4 reports the experimental analysis, followed by results in Section 5. Finally, Sections 6 and 7 end the paper with discussion and conclusions.

2. Related work

2.1. Object detection in medical images

Object detectors can be divided into two categories: (i) two-stage detector, the most representative is Faster R-CNN [15], (ii) one-stage detector, such as YOLO [49], and SSD [39]. Two-stage detectors are characterized by high localization and object recognition accuracy, whereas the one-stage detectors achieve high inference speed [23, 67]. In a two-stage approach, the first stage is responsible of generating candidates that should contain objects, filtering out most of the negative proposals, whereas the second stage performs the classification into foreground/background classes and regression activities of the proposals from the previous stage.

Recently, the most popular object detection methods in computer vision have been applied to medical imaging [68, 6, 21]. In [12], Faster R-CNN [15] is applied with the VGG-16 [38] network as backbone for pulmonary nodule detection. The YOLO architecture has been modified for lymphocyte detection in immunochemistry [60, 50] and for pneumothorax detection on chest radiographs [47]. In [48], a deep learning algorithm based on the YOLOv5 detection model is proposed for automated detection of intracranial artery stenosis and occlusion in magnetic resonance angiography. Other studies [29, 53, 42] exploited architectures such as RetinaNet and Mask R-CNN for lung nodules and breast masses localization. In [8], Mask R-CNN [19] is used by first assigning bounding boxes for each tumor volume to perform detection and classification of normal and abnormal breast tissue.

2.2. Small lesion detection

Although existing object detection models have been very successful with natural images [23, 67], in medical images the high resolution makes the problem particularly challenging to discover small lesions, requiring complex architectures and the use of more than one stage for multi-resolution detection. In [56], three CNN architectures, each at different scale, are applied to lung nodule detection, whereas in [27] a multi-stream CNN is designed to classify skin lesions, where each individual stream worked on a different image resolution. In [64], a context-sensitive deep neural network is developed to take into account both the local image features of a microcalcification and its surrounding tissue background. In [52], a multi-context architecture is proposed, based on the combination of different CNNs with variable depth and individually trained on image patches of different size. In [3], the problem of class imbalance between lesions and background is addressed by proposing a two-stage deep learning approach where the first stage is a cascade [2] of one-level decision trees, and the second is a CNN, trained to focus on the most informative background configurations identified by the first stage.

Recently, in [63] a hierarchical deep learning framework consisting of three models each with a specific task is proposed for bone marrow nucleated differential count. Some

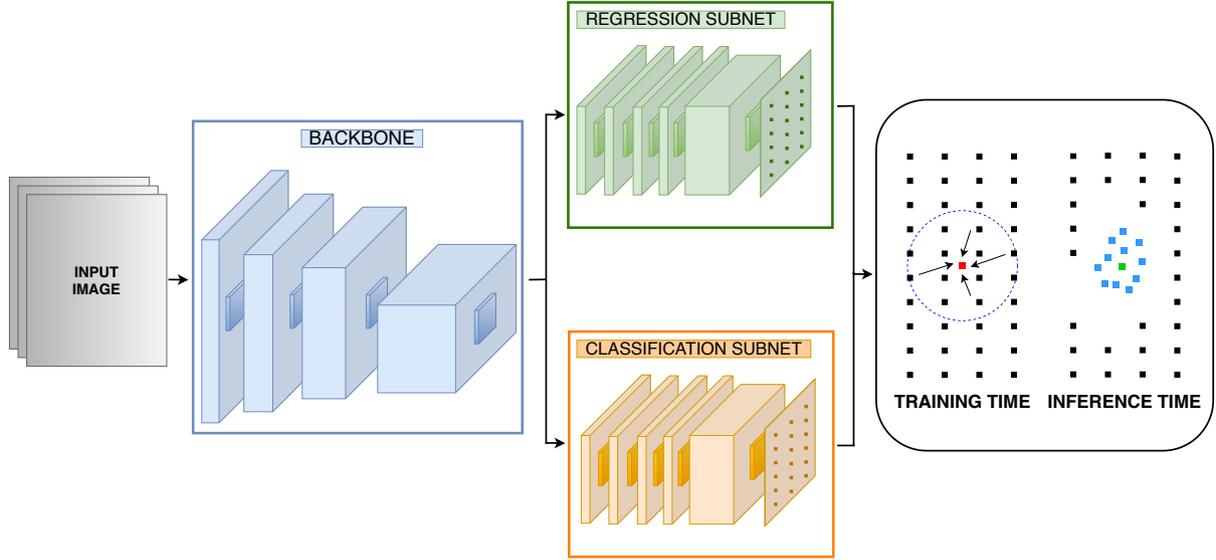


Figure 2: GravityNet architecture is composed of a backbone (blue) and two subnetworks, attached to the backbone output, one for classification task (orange) and one for regression task (green). The output is a representation of the gravity points in the grid pattern at training time and the subsequent attraction behavior towards the lesion at inference time. Gravity points in light blue correspond to positive candidates trained to collapse toward the ground truth in light green

studies [10, 18] combined image processing techniques and deep learning algorithms to evaluate lung tumor and liver tumor detection respectively. In [66], the visibility of microcalcifications in mammographic images is increased by difference filtering using the YOLOv4 model. A three-stage multi-scale framework for the microaneurysms detection is designed in [57], whereas multi-scale approach based on YOLOv5 is proposed for the detection of stroke lesions [5].

3. GravityNet

This section explains the proposed network architecture and the concept of *gravity points*, a new anchoring technique designed for small lesion detection.

The code is available at this link.

3.1. Gravity points

We define a *gravity point* as a pixel-based anchor, which inspects its surroundings to detect lesions. The gravity-points distribution is generated with a grid of points spaced by a user-defined *step* parameter. A base configuration is generated in a squared reference window, named *feature grid*, of size $K \times K$ where K is equal to the upper integer of the ratio between the dimensions of the image and the feature map:

$$K \times K = \left\lceil \frac{H}{H_{FM}} \right\rceil \times \left\lceil \frac{W}{W_{FM}} \right\rceil \quad (1)$$

Assuming that the first gravity point is located in the upper left corner of the feature map, the number of gravity points in a feature grid is equal to:

$$N_{GP}^{FG} = \left(\left\lfloor \frac{K-2}{step} \right\rfloor + 1 \right)^2 \quad (2)$$

where $0 < step \leq K-2$. In cases where $K-2$ is multiple of *step* the distribution will be equispaced in the feature grid.

Since each pixel in the feature map corresponds to a feature grid in the image, the complete configuration is obtained by sliding the base configuration over the whole image. The total number of gravity points N_{GP} in the image is equal to the base configuration times the number of feature grids:

$$N_{GP} = N_{GP}^{FG} \cdot H_{FM} \cdot W_{FM} \quad (3)$$

Fig. 1 shows an example of gravity-points distribution.

3.2. Architecture

GravityNet is a one-stage end-to-end detector composed of a backbone network and two specific subnetworks. The backbone is a convolutional network and plays the role of feature extractor. The first subnet performs convolutional object classification on the backbone output, whereas the second subnet performs convolutional gravity-points regression. Fig. 2 shows the overall architecture.

The backbone is the underlying network architecture of a detection model and provides a feature map containing basic features and representations of input data, which are then processed to perform a specific task. The bottom layers of a backbone net usually extract simple features such as edges and corners, while the top layers learn more complex features like parts of lesions. The feature maps generated by these layers are used as a representation of the input image

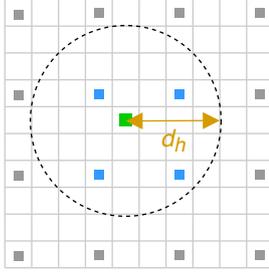


Figure 3: Hooking process where gravity points (light blue) are hooked to a lesion (light green)

and fed into two models for classification and regression tasks.

The classification subnet is a fully convolutional network that outputs the probability of lesion presence at each gravity-point location. The subnetwork applies four 3×3 convolutional layers, each with 256 filters, where the first one maps the number of features output from the backbone, followed by ReLU activations. The last layer applies a filter with $N_{AP} \cdot 2$ outputs and sigmoid activation to obtain the binary predictions for each gravity point.

The regression subnetwork is connected to the output of the backbone with the purpose of regressing the offset from each gravity point to the closest lesion. The design of the regression subnet is the same of the classification subnet. The last layer outputs $N_{AP} \cdot 2$ values, indicating the offsets to move each gravity point towards a lesion. It is worth noting that the classification and regression subnets, though sharing a common structure, use separate parameters.

3.3. Gravity loss

Gravity Loss (GL) is a multi-task loss that contains two terms: one for regression (denoted as GL_{reg}) and the other for classification (denoted as GL_{cls}).

The multi-task loss can be written as:

$$GL = GL_{cls} + \lambda GL_{reg} \quad (4)$$

where λ is an hyperparameter that controls the balance between the two task losses.

3.3.1. Classification loss

Since significant class imbalance between lesion and background is usually present in medical images [3], the classification loss is a variant of Focal Loss [36]. This loss is designed to address the issue of class imbalance in object detection tasks, where the majority of the examples belong to the negative class (e.g., background) and only a few examples belong to the positive class (e.g., lesion).

The classification loss is defined as:

$$GL_{cls} = -\alpha_t \cdot (1 - p_t)^\varphi \cdot \log(p_t) \quad (5)$$

where p_t is the predicted probability of the true class (lesion), φ is a focusing parameter that controls the rate at which the

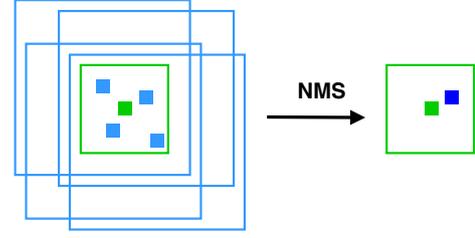


Figure 4: An example of NMS: on the left, gravity points and corresponding boxes (light blue) hooked to a lesion (green); on the right, the final candidate corresponding to the gravity point with the highest score (blue)

modulating factor α_t decreases as the predicted probability p_t increases.

To evaluate p_t with gravity points, we introduce a criterion based on the Euclidean distance between the gravity points and the ground-truth lesions¹. We consider as belonging to the positive class those gravity points with a distance from the closest ground-truth lesion lower than a threshold distance that we named *hooking distance* d_h . All the gravity points within that distance are hooked to the lesion and trained to move towards it. Fig. 3 shows an example of gravity-points hooking process.

3.3.2. Regression loss

Let us indicate as (d_x, d_y) the distance between a gravity point and the relative hooked lesion, and as (o_x, o_y) the output of the regression subnetwork, which represents the offset to move each gravity point towards the hooked lesion.

We evaluate the regression loss as:

$$GL_{reg} = \sum_{\forall \text{ hooked GP } i \in \{x,y\}} \text{smooth}_{L1}(d_i - o_i) \quad (6)$$

where $\text{smooth}_{L1}(t)$ is the Smooth L1 loss [15], defined as:

$$\text{smooth}_{L1}(t) = \begin{cases} 0.5t^2, & \text{if } |t| < 1 \\ |t| - 0.5, & \text{otherwise} \end{cases} \quad (7)$$

3.4. Inference time

The model produces two output predictions for each gravity point for each subnetwork. *Non-Maxima-Suppression* (NMS) is applied to reduce the number of false candidates (see Fig. 4): (i) an $L \times L$ box is built for each hooked gravity points, where L is chosen equal to the average size of the lesions to be detected; (ii) all boxes with an Intersection over Union (IoU) greater than 0.5 are merged; and (iii) for each merger, the gravity point with the highest score is considered as final candidate. After NMS, we determine the lesion class with a threshold γ on the classification score: all predictions with scores above γ belong to the positive class (lesion), the remaining ones to the negative class (no lesion).

¹Without loss of generality, we consider as ground truth the center of the smallest bounding box containing the lesion.

4. Experiments

We proved the effectiveness of *GravityNet* on two detection problems in medical image analysis: (i) microcalcifications on full field digital mammograms and (ii) microaneurysms on digital ocular fundus images.

Microcalcifications (MCs) are calcium deposits and are considered as robust markers of breast cancer [41]. MCs appear as fine, white specks, similar to grains of salt, with size between 0.1 mm and 1 mm. Due their small dimensions and the inhomogeneity of the surrounding breast tissue, identifying MCs is a very challenging task. Moreover, mammograms contain a variety of linear structures (such as vessels, ducts, etc.) that are very similar to MCs in size and shape, making detection even more complex.

Microaneurysms (MAs) are the earliest visible manifestation of Diabetic Retinopathy, one of the leading causes of vision loss globally [34]. MAs are described as isolated small red dots of 10-100 μm of diameter sparse in retinal fundus images, but sometimes they appear in combination with vessels. Retinal vessels, together with dot-hemorrhages and other objects like the small and round spots resulting from the crossing of thin blood vessels, make MAs hard to distinguish.

4.1. Dataset

4.1.1. Microcalcifications dataset

We used the publicly available INBreast database [45], acquired at the Breast Centre in Centro Hospitalar de S. João (CHSJ) in Porto, Portugal. The acquisition equipment was the MammoNovation Siemens FFDM, with a solid-state detector of amorphous selenium, pixel size of 70 μm (microns) and 14-bit contrast resolution. The image matrix was $4,084 \times 3,328$ (243 images) or $3,328 \times 2,560$ (167 images), depending on the compression plate used in the acquisition and according to the breast size of the patient.

The database has a total of 410 images, amounting to 115 cases, from which 90 cases are from women with both breasts, and 25 are from mastectomy patients. Calcifications can be found in 313 images for a total of 7,142 individual calcifications. In this work, only calcifications with a radius of less than 7 pixels were considered for testing, for a total of 5,657 microcalcifications identified in 296 images.

Mammograms have been cropped to the size $3,328 \times 2,560$ to have all images in the dataset with equal size. We ensured that no MC was missed after cropping.

4.1.2. Microaneurysms dataset

We used the publicly available E-optha database [11], designed for scientific research in Diabetic Retinopathy. The acquired images have dimensions ranging from $960 \times 1,440$ to $1,696 \times 2,544$ with a 45° field of view (FOV) and a pixel size of 7-15 μm . The database has a total of 381 images: 148 images from unhealthy patients containing 1,306 microaneurysms, and 233 images from healthy patients.

The original retinal fundus images are RGB, but in this work, green channel has been extracted due to its rich information and high contrast in comparison with the other

Table 1
Data overview

INbreast	Images		Unhealthy		MCs	
	1-fold	2-fold	1-fold	2-fold	1-fold	2-fold
Train	143	143	108	117	2,408	2,051
Validation	62	62	39	45	516	724
Test	205	205	154	142	2,756	2,901

E-optha-MA	Images		Unhealthy		MAs	
	1-fold	2-fold	1-fold	2-fold	1-fold	2-fold
Train	154	151	60	60	542	552
Validation	38	38	14	14	105	107
Test	189	192	74	74	659	647

two color channels [62]. We also evaluated the average dimensions of all the retinas in the dataset and resized all the images to an average dimensions of $1,216 \times 1,408$.

4.2. Data preparation

For all experiments we applied 2-fold image-based cross-validation. The dataset is divided into two equal-sized folds, where one fold is used as training set and the other as test set and vice versa. A subset of the training set fold is used as validation set for parameter optimisation. See Tab. 1 for more details.

In both datasets, data augmentation techniques are used to address class imbalance and enhance the model robustness and accuracy [31]. Three more samples for each image are generated by using horizontal and vertical flipping. All data are normalized with min-max transformation.

4.3. Architecture parameters

GravityNet uses ResNet [20] as its backbone, in order to solve the well-known vanishing/exploding gradient problems [1] by using residual connections. ResNet is composed of 5 max-pooling layers, each halving the dimensions of the feature map. According to the input dimensions, the feature map size is 104×80 for mammograms and 38×44 for retina fundus images. As a consequence, according to Eq. 1, we obtain $K = 32$ and a feature grid of 32×32 . We generate gravity-points configurations with *step* multiple of $K - 2$ to ensure equi-spatiality. To take into account the computational cost, we chose to use configurations that did not exceed 300,000 gravity points. Fig. 5 shows some examples of the initial configurations used in this work. To ensure that at least one gravity point hooks a lesion, the *hooking distance* d_h was always chosen equal to the *step*. At inference time, we use NMS with $L = 7$ for MCs and $L = 3$ for MAs, which correspond to the average size of the lesions to be detected.

We train ResNet in transfer learning by using a model pretrained on natural images [65]. Both subnetworks are initialised with Xavier technique [16]. During training we

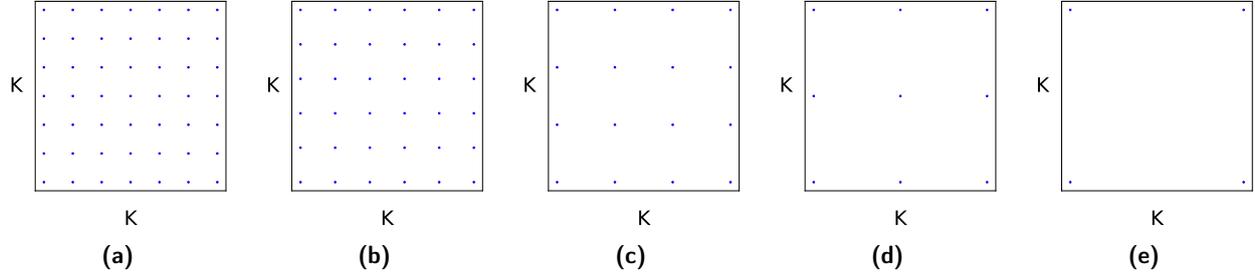


Figure 5: Examples of initial gravity-points configurations represented in a reference window $K \times K$, where: (a) $step = 5$ (b) $step = 6$ (c) $step = 10$ (d) $step = 15$ (e) $step = 30$

apply Adam optimization algorithm [30]. The learning rate was set to an initial value of 10^{-4} and decreased with a factor of 0.1 with *patience* equal to 3. The balance between the two task losses (see Eq. 4) is handled by λ equal to 10. The batch size is by default equal to 8. All training parameters were optimized on the validation set. The training was stopped after 50 epochs. Experiments were conducted on a GPU NVIDIA A100 80GB.

4.4. FROC analysis

The detection quality was evaluated in terms of lesion-based Free-Response Operating Characteristic (FROC) curve by plotting True Positive Rate (TPR) against the average number of False Positives per Image (FPpI) for a series of thresholds γ on the classification score associated to each sample.

A prediction with a value higher than θ is considered as True Positive (TP) when its distance from the center of a lesion is no larger than the largest side of the bounding box containing the ground-truth lesion. Otherwise, it is considered as False Positive. Notably, (i) if multiple predictions are associated to the same lesion, only the one with the highest classification score is selected as TP, and (ii) all predictions for gravity points outside the tissue were ignored.

To analyze and compare FROC curves, we chose the non-parametric approach suggested in [4]. The figure-of-merit is the Partial Area under the FROC curve ($AUFC_\gamma$) to the left of $FPpI = \gamma$ calculated by trapezoidal integration. We normalized $AUFC_\gamma$ by dividing with γ to obtain an index in the range [0, 1]. In particular, for both MCs and MAs detection, we selected $\gamma = 10$, a commonly used value in the literature of the respective fields [9, 52]. All results are presented in percentage values.

5. Results

5.1. Model analysis

To verify the effectiveness of the model, for both small lesion detection problems, experiments were conducted using different gravity-points configurations for all different depths of ResNet². Results are reported in Tab. 2 for MCs,

²It is worth noting that, due to memory constraints, for MCs detection, we use in training a batch size equal to 4 for *ResNet-50* and 2 for *ResNet-101* and *ResNet-152*

and in Tab. 3 for MAs together with the parameters of the gravity-points configurations. The best result for each backbone is shown in bold, whereas the best of all in italic. FROC curves of the best ResNet configurations are shown in Fig. 6.

For MCs, the best result is a $AUFC_\gamma$ equal to 72.25% by using *ResNet-34* and $step = 10$. Configuration with $step = 10$ turns to be the best for all backbones, except *ResNet-50*, which achieves a $AUFC_\gamma$ equal to 71.25% with $step = 6$. Dense configurations present better results with shallower backbones, e.g. with a *ResNet-18* we obtain a $AUFC_\gamma$ equal to 70.89% and 71.47% respectively with $step = 6$ and 10 as opposed to 65.58% and 55.90% respectively with $step = 15$ and 30.

For MAs, the highest $AUFC_\gamma$ (71.53%) is obtained with a *ResNet-50* and $step = 6$. Configuration with $step = 6$ turns to be the best for all backbones, except *ResNet-18*, which achieves a $AUFC_\gamma$ equal to 65.36% with $step = 10$. Sparse configurations decrease the performance, even with deeper backbones, e.g. with a *ResNet-152* we obtain a $AUFC_\gamma$ of 67.51% and 54.18% respectively with $step = 15$ and 30 as opposed to 65.81% and 69.86% respectively with $step = 5$ and 6.

Through result analysis, it becomes evident that we need to find the appropriate density configuration for addressing the detection problem at hand. A sparse configuration might fail to identify all lesions, particularly in the case of small ones, whereas a dense configuration could potentially generate a high number of lesion candidates.

5.2. Comparison with the literature

We compare our best models, that are *ResNet-34* with $step = 10$ for MCs detection and *ResNet-50* with $step = 6$ for MAs detection, with methods proposed in the scientific literature for the detection problems at hand:

- Context-Sensitive CNN (CSNet) [64]: the architecture comprises two convolutional subnetworks: one for processing the large image context with a window of size 96×96 pixels and another for processing the small microcalcification texture with a window of size 9×9 pixels. The features extracted from both subnetworks are subsequently merged and fed into a fully connected network.

Table 2
Results for MCs detection in terms of % of $AUFC_\gamma$

Backbone	Configuration			
	$step = 6$ $d_h = 6$	$step = 10$ $d_h = 10$	$step = 15$ $d_h = 15$	$step = 30$ $d_h = 30$
	$N_{GP}=299,520$	$N_{GP}=133,120$	$N_{GP}=74,880$	$N_{GP}=33,280$
ResNet-18	70.89	71.47	65.58	55.90
ResNet-34	65.08	72.25	67.44	56.17
ResNet-50	71.25	67.73	69.31	57.12
ResNet-101	58.85	64.90	41.69	53.05
ResNet-152	60.60	64.86	62.98	53.86

Table 3
Results for MAs detection in terms of % of $AUFC_\gamma$

Backbone	Configuration				
	$step = 5$ $d_h = 5$	$step = 6$ $d_h = 6$	$step = 10$ $d_h = 10$	$step = 15$ $d_h = 15$	$step = 30$ $d_h = 30$
	$N_{GP}=81,928$	$N_{GP}=60,192$	$N_{GP}=26,752$	$N_{GP}=15,048$	$N_{GP}=6,688$
ResNet-18	60.95	61.42	65.36	63.17	53.88
ResNet-34	65.01	68.57	68.38	64.80	58.76
ResNet-50	68.89	71.53	64.57	68.25	54.33
ResNet-101	66.07	69.77	69.13	65.59	54.97
ResNet-152	65.81	69.86	66.84	67.51	54.18

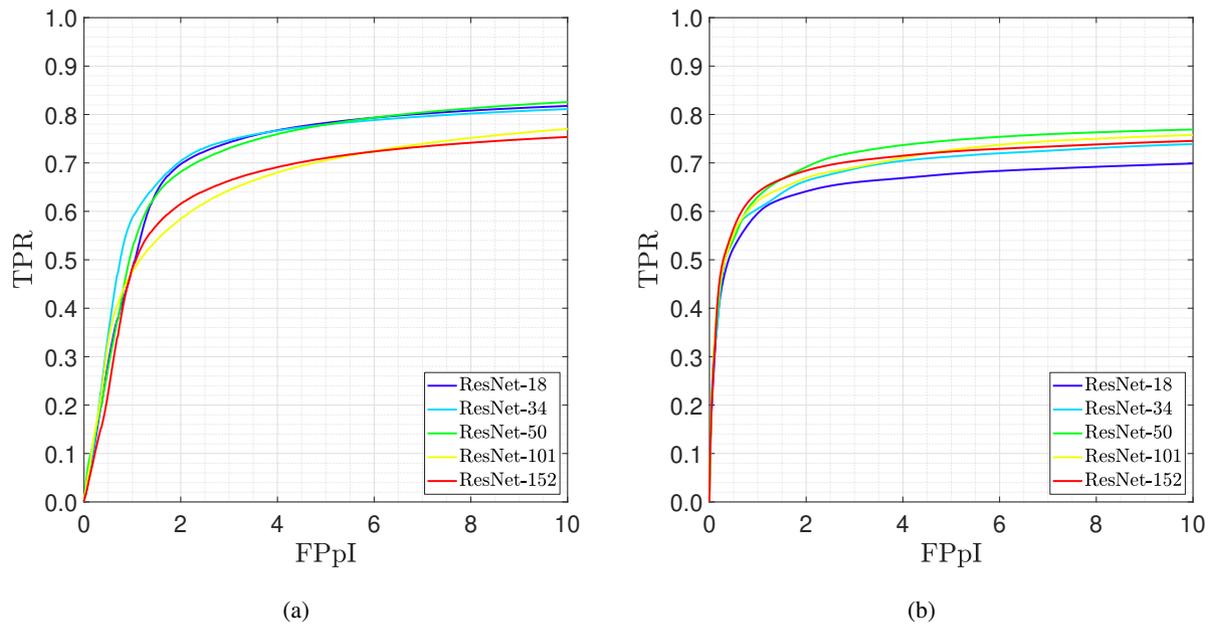


Figure 6: FROC results with the best gravity-points configurations for each ResNet backbone on INbreast (a) and E-optha-MA (b)

- Deep Cascade (DC) [2]: a cascade of decision stumps able to learn effectively from heavily class-unbalanced datasets. It builds on Haar features computed in a small detection window of 12×12 pixels, which can contain diagnostically relevant lesions, while limiting the exponential growth of the number of features that are extracted during training.
- MCNet with DC hard mining (DC-MCNet) [3]: a two-stage patch-based deep learning framework, which comprises a DC for hard mining the background samples, followed by a second stage represented by a CNN that discriminates between lesions and the more challenging background configurations.
- Multicontext Ensemble of MCNets (ME-MCNet) [52]: a multi-context ensemble of CNNs aiming to learn different levels of image spatial context by training multiple-depth networks on image patches of different dimensions (12×12 , 24×24 , 48×48 , and 96×96).

To evaluate the behavior of the proposed anchoring technique, we also compared with RetinaNet [36], a well-known one-stage object detector based on anchoring technique. We slightly modified the original anchors configuration by using an anchor box size ranging from 8^2 to 128^2 in order to be more suitable for small lesion detection. We applied it to the whole image without any kind of rescale or patching.

We applied a statistical comparison by means of bootstrap method [51] to test the significance of observed performances. Cases were sampled with replacement 1,000 times, with each bootstrap containing the same number of cases as the original set. At each bootstrapping iteration, FROC curves were recalculated, and differences in figures-of-merit $\Delta AUFC_\gamma$ between *GravityNet* and each of the compared methods were evaluated. Finally, the obtained FROC curves were averaged along the TPR axis, and p -values were computed as the fraction of $\Delta AUFC_\gamma$ populations that were negative or zero. The statistical significance level was chosen as $\alpha = 0.05$. Average FROC curve are shown in Fig. 7.

The statistical comparison results for MCs and MAs detection are shown in Tab. 4 where significant performances are indicated in bold. Results of the proposed architecture were statistically significantly better than all the other considered approaches. The highest improvement in terms of $AUFC_\gamma$ is of +50.04% with RetinaNet for MAs and of +42.25% with CSNet for MCs. Compared to patch-based methods such as DC, DC-MCNet and ME-MCNet the improvement is respectively +41.00%, +19.52%, +11.90% for MCs and +15.72%, +10.15%, +5.90% for MAs.

6. Discussion

6.1. Gravity points configuration

The gravity points configuration depends directly on the size of the input image and is managed by the *step* parameter. This implies a higher number of gravity points for images with larger dimensions. In the cases studied, mammograms

have a larger size than retina images and consequently have a higher N_{GP} , so requiring much more computational efforts.

Depending on the chosen configuration, gravity-points will behave differently. We chose to train all the configurations with a d_h equal to the *step* to measure the capacity of gravity-points to move towards ground-truth lesions. A small d_h will have less impact on the movement of gravity-points, compared to a large d_h that let them move more widely, always within the specified distance value. For MCs detection, the best configurations are those with *step* 10 and thus d_h 10 because these values are more representative of the size and distribution of MCs in mammographies. On the other hand, for MAs detection, where lesions are usually isolated, configurations with a higher density, such as *step* 5 and *step* 6, are needed. Fig. 8 shows two detection outputs of the best GravityNet models for MCs with *step* 10 and *ResNet-34* and for MAs with *step* 6 and *ResNet-50*. We can see the gravitational behaviour towards the centres of the lesions in Fig. 8b for MCs and Fig. 8d for MAs. Hooked gravity-points that are, at inference time, within the radius of the lesion to be detected are shown in light blue and are defined as predictions of possible TP. The NMS, whose output can be seen in the right panel of the same figures, merges all hooked gravity-points in a single detection so as to obtain a single prediction (in blue) for each small lesion (in green).

6.2. Comparison with anchoring methods

We compared the proposed one-stage detector with a widespread exponent of one-stage object detection methods, i.e. *RetinaNet*, which has also been usefully applied to medical detection problems [24, 42]. Small lesions such as MCs and MAs are often less than 10 pixels in diameter and, in this case, anchoring methods face two main obstacles: (i) the number and size of anchor boxes, and (ii) the pyramidal approach for multi-scale resolution.

Regarding the first issue, we tried to train *RetinaNet* with the original range of anchor boxes size (from 32^2 to 512^2 according to the *Feature Pyramid Network* (FPN) level), but due to the small size of the lesions, the train failed; thus, we reduced the size in the range 8^2 to 128^2 . The proposed anchoring technique is based on pixel-shaped gravity points, which only require an initial configuration setting without specifying a box size. This is advantageous, especially in the case of small lesions with variable sizes, as demonstrated in the MA results. In addition, considering all FPN resolution levels, *RetinaNet* generates a number of anchor boxes more than 10 times the number of gravity points. This is a considerable advantage in computational and temporal terms (see Section 6.4).

As to the second issue, *RetinaNet* adopts a multi-scale architecture. However, this approach proves to be ineffective because positive anchors (those containing a lesion) only belong to the first level of FPN, which corresponds to the highest resolution level. In *GravityNet*, we decided to not use a multi-scale approach given the shape of the lesions to be detected. For the sake of comparison, we tried to use

Table 4
Results comparison in terms % of $AUFC_\gamma$

	Method	$AUFC_\gamma$	Compared to	$\Delta AUFC_\gamma$	p-Value
MCs detection	RetinaNet	66.47			
	CSNet	30.00			
	DC	31.25			
	DC-MCNet	52.73			
	ME-MCNet	60.35			
	GravityNet	72.25	RetinaNet	+5.78	= 0.037
			CSNet	+42.25	< 0.001
			DC	+41	< 0.001
			DC-MCNet	+19.52	< 0.001
			ME-MCNet	+11.9	< 0.001
MAs detection	RetinaNet	21.48			
	CSNet	40.03			
	DC	55.80			
	DC-MCNet	61.38			
	ME-MCNet	65.63			
	GravityNet	71.53	RetinaNet	+50.04	< 0.001
			CSNet	+31.49	< 0.001
			DC	+15.72	< 0.001
			DC-MCNet	+10.15	< 0.001
			ME-MCNet	+5.9	< 0.001

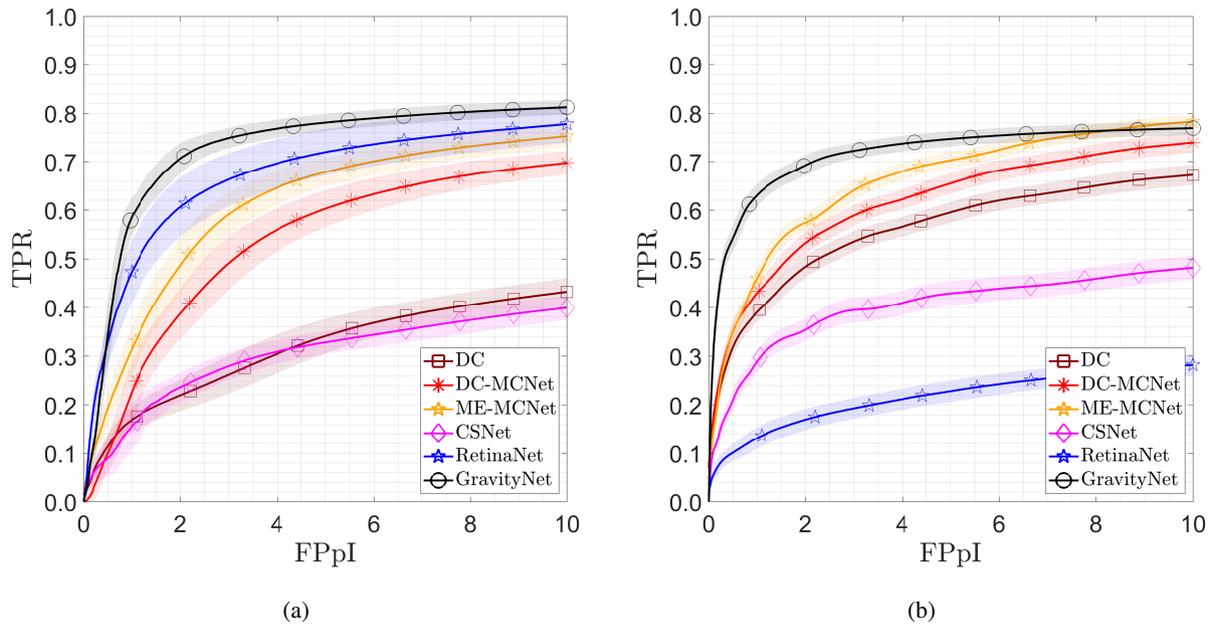


Figure 7: Average FROC curves for INbreast (a) and E-optha-MA (b) obtained from 1,000 bootstrap iterations. Confidence bands (semi-transparent) indicate 95% confidence intervals along the TPR axis.

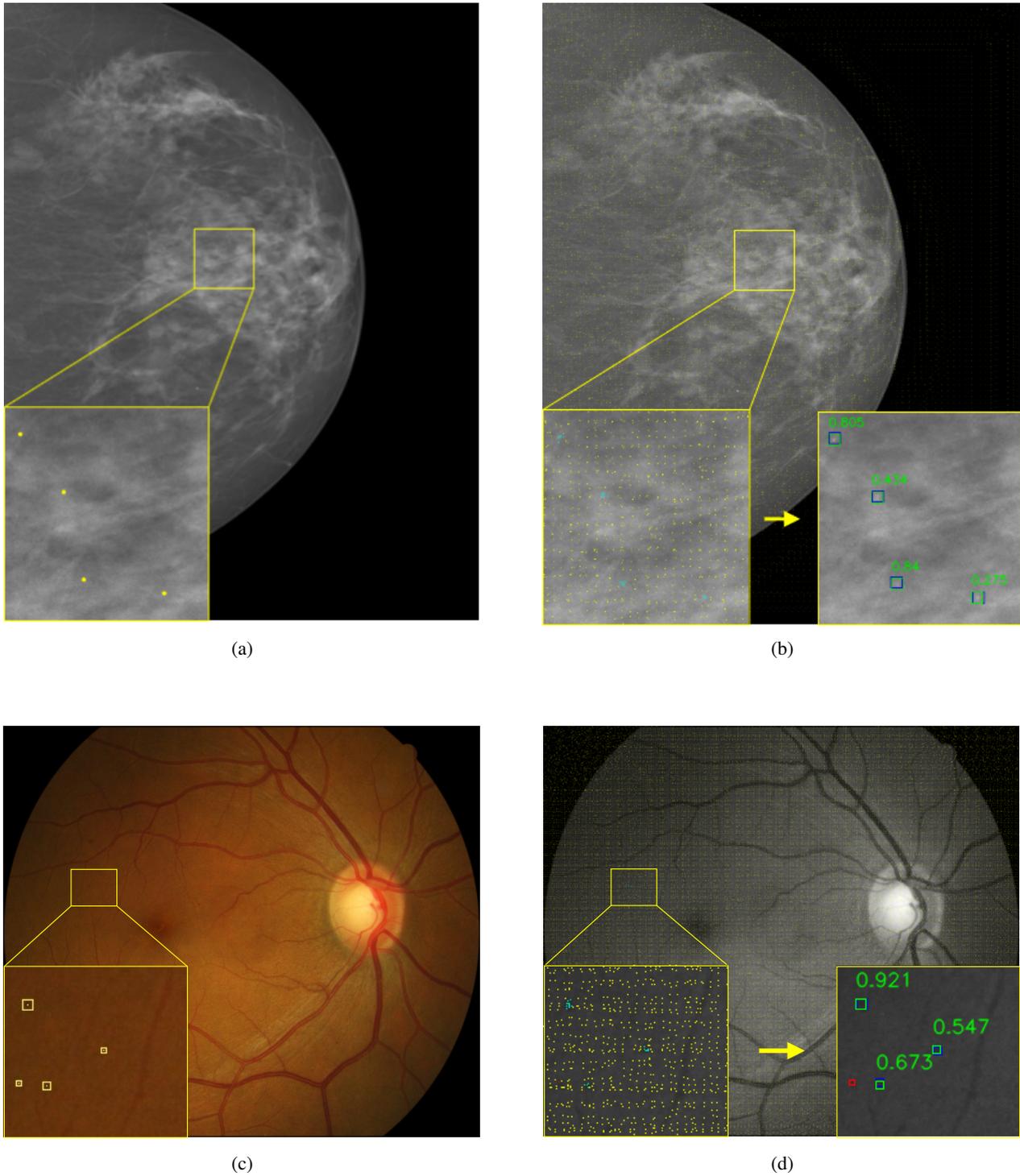


Figure 8: Examples of MCs and MAs detections. (a) and (c): ground-truth annotations; (b) and (d): GravityNet outputs

RetinaNet without FPN, considering only the outputs of the first level, but this did not improve the performance.

6.3. Comparison with patch-based methods

Patch-based methods have the computational disadvantage of assembling all the individual results to obtain the final one, as opposed to end-to-end systems like *GravityNet* that obtain the final result directly.

The class imbalance between lesions and background is another issue that affects small lesion detection. We compared our approach with two existing methods, *DC* and *DC-MCNet*, which are designed to manage this problem. *DC* discards the majority of easily detectable background samples early in the process, while *DC-MCNet* utilizes a CNN on the output of *DC* to enhance detection performance. In this work, we propose the *Gravity Loss*, a variant of *Focal Loss* typically applied in deep learning methods to address class imbalance issues.

Since small lesions do not have a clear appearance and are similar to the surrounding background, *CSNet* and *ME-MCNet* propose two context-sensitive patch-based approaches, where the model is trained with patches of different sizes and then combined. In contrast, our proposal works with the full image without patches and is able to identify small lesions thanks to the new anchoring technique and the regression subnet, which focus more on the distance to the lesion rather than its appearance.

6.4. Computational and inference time

Computational and inference time are an important aspect in medical imaging systems, improving interactivity and the time taken to formulate a diagnosis. We evaluated the computational time for all the compared methods by measuring the average *Time per Epoch* (TpE) in training, and the *Time per Image* (TpI) and the Throughput³ in test. Tab. 5 shows the results. We can see how patch-based methods are computationally time-consuming, whereas our proposal has a very high Throughput and an average TpI below one second.

6.5. Limitations

Although our method achieves excellent results in the detection of small lesions, there are some limitations to be considered:

- Clinical applicability: we require a dataset with individually annotated lesions for the training phase, and this can be difficult to meet in a real clinical scenario. In addition, further post-processing (e.g. benign vs. malignant lesion classification) is needed to build a full CAD system.
- Configuration limit: by employing an equispaced grid configuration, the distribution of gravity points becomes uniform, even in areas of the image where there is no tissue. In training this might not be advantageous.

³Throughput is defined as the maximum number of input instances that the method can process in one second

Table 5

Computational times compared in terms of *Time per Epoch* (TpE) in training, and *Time per Image* (TpI) and *Throughput* in test

	Method	TpE (s)	TpI (s)	Throughput
MCs detection	RetinaNet	1254	0.121	14.70
	CSNet	6959	822	1.2×10^{-3}
	DC	n.a.	1.1	0.9
	DC-MCNet	32	1.2	0.8
	ME-MCNet	5360	386	3.8×10^{-3}
	GravityNet	607	0.061	19.25
MAs detection	RetinaNet	137	0.057	34.04
	CSNet	1260	266	3.8×10^{-3}
	DC	n.a.	1.3	0.7
	DC-MCNet	6	1.4	0.7
	ME-MCNet	1564	203	4.9×10^{-3}
	GravityNet	184	0.045	37.49

Different approaches to generate the initial configuration can be investigated.

- Computational requirements: the number of gravity points directly increases with the size of the image. In case of large images, *GravityNet* can require considerable computational resources. A solution can be to limit the number of gravity points by using sparse initial configuration, but this can affect the detection performance of the method.
- Memory constraints: the use of a backbone in the proposed model necessitates remarkable resource requirements. As the backbone architecture becomes more complex and deeper, it requires a larger memory allocation, which can be a significant limitation for training the model.

7. Conclusions and future work

In this work, we introduced *GravityNet*, a new one-stage end-to-end detector specifically designed to detect small lesions in medical images. The accurate localization of small lesions, given their appearance and diverse contextual backgrounds, is a challenge in several medical applications. To address this point, our approach employed a novel pixel-based anchor that dynamically moves towards the targeted lesion during detection. Through a comparative evaluation with state-of-the-art anchoring and patch-based methods, our proposed approach demonstrated promising results in effectively detecting small lesions.

Our primary future direction will involve testing *GravityNet* in various detection problems, particularly those where the target object is point-like, such as nuclei localization in whole-slide images [22]. We will also explore the possibility of extending the proposed architecture to address other tasks or image dimensionality involving small lesions, such as segmentation [35] or three-dimensional images [28, 61].

References

- [1] Bengio, Y., Simard, P., Frasconi, P., 1994. Learning long-term dependencies with gradient descent is difficult. *IEEE Transactions on Neural Networks* 5, 157–166. doi:10.1109/72.279181.
- [2] Bria, A., Marrocco, C., Karssemeijer, N., Molinara, M., Tortorella, F., 2016. Deep cascade classifiers to detect clusters of microcalcifications, in: Tingberg, A., Lång, K., Timberg, P. (Eds.), *Breast Imaging*, Springer International Publishing, p. 415–422. doi:10.1007/978-3-319-41546-8_52.
- [3] Bria, A., Marrocco, C., Tortorella, F., 2020. Addressing class imbalance in deep learning for small lesion detection on medical images. *Computers in Biology and Medicine* 120, 103735. doi:10.1016/j.combiomed.2020.103735.
- [4] Chakraborty, D.P., 2008. Validation and statistical power comparison of methods for analyzing free-response observer performance studies. *Academic Radiology* 15, 1554–1566. doi:https://doi.org/10.1016/j.acra.2008.07.018.
- [5] Chen, S., Duan, J., Wang, H., Wang, R., Li, J., Qi, M., Duan, Y., Qi, S., 2022a. Automatic detection of stroke lesion from diffusion-weighted imaging via the improved yolov5. *Computers in Biology and Medicine* 150, 106120. doi:https://doi.org/10.1016/j.combiomed.2022.106120.
- [6] Chen, X., Wang, X., Zhang, K., Fung, K.M., Thai, T.C., Moore, K., Mannel, R.S., Liu, H., Zheng, B., Qiu, Y., 2022b. Recent advances and clinical applications of deep learning in medical image analysis. *Medical Image Analysis* 79, 102444. doi:10.1016/j.media.2022.102444.
- [7] Ciga, O., Xu, T., Nofech-Mozes, S., Noy, S., Lu, F.I., Martel, A.L., 2021. Overcoming the limitations of patch-based learning to detect cancer in whole slide images. *Scientific Reports* 11, 8894. doi:10.1038/s41598-021-88494-z.
- [8] Civilibal, S., Cevik, K.K., Bozkurt, A., 2023. A deep learning approach for automatic detection, segmentation and classification of breast lesions from thermal images. *Expert Systems with Applications* 212, 118774. doi:10.1016/j.eswa.2022.118774.
- [9] Dashtbozorg, B., Zhang, J., Huang, F., ter Haar Romeny, B.M., 2018. Retinal microaneurysms detection using local convergence index features. *IEEE Transactions on Image Processing* 27, 3300–3315. doi:10.1109/TIP.2018.2815345.
- [10] Dass, J.M.A., Kumar, S.M., 2022. A novel approach for small object detection in medical images through deep ensemble convolution neural network. *International Journal of Advanced Computer Science and Applications (IJACSA)* 13. doi:10.14569/IJACSA.2022.0130380.
- [11] Decencière, E., Cazuguel, G., Zhang, X., Thibault, G., Klein, J.C., Meyer, F., Marcotegui, B., Quellec, G., Lamard, M., Danno, R., Elie, D., Massin, P., Viktor, Z., Erginay, A., Laÿ, B., Chabouis, A., 2013. Teleophtha: Machine learning and image processing methods for teleophthalmology. *IRBM* 34, 196–203. doi:10.1016/j.irbm.2013.01.010.
- [12] Ding, J., Li, A., Hu, Z., Wang, L., 2017. Accurate pulmonary nodule detection in computed tomography images using deep convolutional neural networks, in: Descoteaux, M., Maier-Hein, L., Franz, A., Jannin, P., Collins, D.L., Duchesne, S. (Eds.), *Medical Image Computing and Computer Assisted Intervention MICCAI 2017*, Springer International Publishing, Cham, p. 559–567. doi:10.1007/978-3-319-66179-7_64.
- [13] Eadie, L.H., Taylor, P., Gibson, A.P., 2012. A systematic review of computer-assisted diagnosis in diagnostic cancer imaging. *European Journal of Radiology* 81, e70–e76. doi:10.1016/j.ejrad.2011.01.098.
- [14] Ezra, E., Keinan, E., Mandel, Y., Boulton, M.E., Nahmias, Y., 2013. Non-dimensional analysis of retinal microaneurysms: critical threshold for treatment. *Integrative biology: quantitative biosciences from nano to macro* 5, 474–480. doi:10.1039/c3ib20259c.
- [15] Girshick, R., 2015. Fast r-cnn, in: 2015 IEEE International Conference on Computer Vision (ICCV), pp. 1440–1448. doi:10.1109/ICCV.2015.169.
- [16] Glorot, X., Bengio, Y., 2010. Understanding the difficulty of training deep feedforward neural networks, in: *Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics, JMLR Workshop and Conference Proceedings*, p. 249–256.
- [17] Gu, F., Wu, X., Wu, W., Wang, Z., Yang, X., Chen, Z., Wang, Z., Chen, G., 2022. Performance of deep learning in the detection of intracranial aneurysms: A systematic review and meta-analysis. *European Journal of Radiology* 155, 110457. doi:10.1016/j.ejrad.2022.110457.
- [18] Han, R., Liu, X., Chen, T., 2022. Yolo-sg: Saliency-guided detection of small objects in medical images, in: 2022 IEEE International Conference on Image Processing (ICIP), pp. 4218–4222. doi:10.1109/ICIP46576.2022.9898077.
- [19] He, K., Gkioxari, G., Dollár, P., Girshick, R., 2017. Mask r-cnn, in: 2017 IEEE International Conference on Computer Vision (ICCV), pp. 2980–2988. doi:10.1109/ICCV.2017.322.
- [20] He, K., Zhang, X., Ren, S., Sun, J., 2016. Deep residual learning for image recognition, in: 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 770–778. doi:10.1109/CVPR.2016.90.
- [21] Jiang, H., Diao, Z., Shi, T., Zhou, Y., Wang, F., Hu, W., Zhu, X., Luo, S., Tong, G., Yao, Y.D., 2023a. A review of deep learning-based multiple-lesion recognition from medical images: classification, detection and segmentation. *Computers in Biology and Medicine* 157, 106726. doi:10.1016/j.combiomed.2023.106726.
- [22] Jiang, H., Zhou, Y., Lin, Y., Chan, R.C.K., Liu, J., Chen, H., 2023b. Deep learning for computational cytology: A survey. *Medical Image Analysis* 84, 102691. doi:10.1016/j.media.2022.102691.
- [23] Jiao, L., Zhang, F., Liu, F., Yang, S., Li, L., Feng, Z., Qu, R., 2019. A survey of deep learning-based object detection. *IEEE Access* 7, 128837–128868. doi:10.1109/ACCESS.2019.2939201.
- [24] Jung, H., Kim, B., Lee, I., Yoo, M., Lee, J., Ham, S., Woo, O., Kang, J., 2018. Detection of masses in mammograms using a one-stage object detector based on a deep convolutional neural network. *PLOS ONE* 13, e0203355. doi:10.1371/journal.pone.0203355.
- [25] Karimi, D., Ward, R.K., 2016. Patch-based models and algorithms for image processing: a review of the basic principles and methods, and their application in computed tomography. *International Journal of Computer Assisted Radiology and Surgery* 11, 1765–1777. doi:10.1007/s11548-016-1434-z.
- [26] Kaur, R., Singh, S., 2022. A comprehensive review of object detection with deep learning. *Digital Signal Processing* 132, 103812. doi:10.1016/j.dsp.2022.103812.
- [27] Kawahara, J., Hamarneh, G., 2016. Multi-resolution-tract cnn with hybrid pretrained and skin-lesion trained layers, in: *Machine Learning in Medical Imaging*, Springer International Publishing, p. 164–171. doi:10.1007/978-3-319-47157-0_20.
- [28] Kern, D., Mastmeyer, A., 2021. 3d bounding box detection in volumetric medical image data: A systematic literature review, in: 2021 IEEE 8th International Conference on Industrial Engineering and Applications (ICIEA), pp. 509–516. doi:10.1109/ICIEA52957.2021.9436733.
- [29] Kim, Y.G., Lee, S.M., Lee, K.H., Jang, R., Seo, J.B., Kim, N., 2020. Optimal matrix size of chest radiographs for computer-aided detection on lung nodule or mass with deep learning. *European Radiology* 30, 4943–4951. doi:10.1007/s00330-020-06892-9.
- [30] Kingma, D.P., Ba, J., 2017. Adam: A method for stochastic optimization. *International Conference on Learning Representations (ICLR)*.
- [31] Kisantal, M., Wojna, Z., Murawski, J., Naruniec, J., Cho, K., 2019. Augmentation for small object detection, in: 9th International Conference on Advances in Computing and Information Technology (ACITY 2019), Aircc Publishing Corporation, p. 119–133. doi:10.5121/csit.2019.91713.
- [32] LeCun, Y., Bengio, Y., Hinton, G., 2015. Deep learning. *Nature* 521, 436–444. doi:10.1038/nature14539.
- [33] LeCun, Y., Bottou, L., Bengio, Y., Haffner, P., 1998. Gradient-based learning applied to document recognition. *Proceedings of the IEEE* 86, 2278–2324. doi:10.1109/5.726791.
- [34] Lee, R., Wong, T.Y., Sabanayagam, C., 2015. Epidemiology of diabetic retinopathy, diabetic macular edema and related vision loss.

- Eye and Vision 2, 17. doi:10.1186/s40662-015-0026-2.
- [35] Li, L., Wei, M., Liu, B., Atchaneeyasakul, K., Zhou, F., Pan, Z., Kumar, S.A., Zhang, J.Y., Pu, Y., Liebeskind, D.S., Scalzo, F., 2021. Deep learning for hemorrhagic lesion detection and segmentation on brain ct images. *IEEE Journal of Biomedical and Health Informatics* 25, 1646–1659. doi:10.1109/JBHI.2020.3028243.
- [36] Lin, T.Y., Goyal, P., Girshick, R., He, K., Dollár, P., 2017. Focal loss for dense object detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 42, 318–327. doi:10.1109/TPAMI.2018.2858826.
- [37] Litjens, G., Kooi, T., Bejnordi, B.E., Setio, A.A.A., Ciompi, F., Ghafoorian, M., van der Laak, J.A.W.M., van Ginneken, B., Sánchez, C.I., 2017. A survey on deep learning in medical image analysis. *Medical Image Analysis* 42, 60–88. doi:10.1016/j.media.2017.07.005.
- [38] Liu, S., Deng, W., 2015. Very deep convolutional neural network based image classification using small training sample size, in: 2015 3rd IAPR Asian Conference on Pattern Recognition (ACPR), pp. 730–734. doi:10.1109/ACPR.2015.7486599.
- [39] Liu, W., Anguelov, D., Erhan, D., Szegedy, C., Reed, S., Fu, C.Y., Berg, A.C., 2016. Ssd: Single shot multibox detector, in: Leibe, B., Matas, J., Sebe, N., Welling, M. (Eds.), *Computer Vision – ECCV 2016*, Springer International Publishing, p. 21–37. doi:10.1007/978-3-319-46448-0_2.
- [40] Lo, S.C., Lou, S.L., Lin, J.S., Freedman, M., Chien, M., Mun, S., 1995. Artificial convolution neural network techniques and applications for lung nodule detection. *IEEE Transactions on Medical Imaging* 14, 711–718. doi:10.1109/42.476112.
- [41] Logullo, A.F., Prigenzi, K.C.K., Nimir, C.C.B.A., Franco, A.F.V., Campos, M.S.D.A., 2022. Breast microcalcifications: Past, present and future (review). *Molecular and Clinical Oncology* 16. doi:10.3892/mco.2022.2514.
- [42] Lotter, W., Diab, A.R., Haslam, B., Kim, J.G., Grisot, G., Wu, E., Wu, K., Onieva, J.O., Boyer, Y., Boxerman, J.L., Wang, M., Bandler, M., Vijayaraghavan, G.R., Gregory Sorensen, A., 2021. Robust breast cancer detection in mammography and digital breast tomosynthesis using an annotation-efficient deep learning approach. *Nature Medicine* 27, 244–249. doi:10.1038/s41591-020-01174-9.
- [43] Loud, J., Murphy, J., 2017. Cancer screening and early detection in the 21st century. *Seminars in oncology nursing* 33, 121–128. doi:10.1016/j.soncn.2017.02.002.
- [44] Loverdos, K., Fotiadis, A., Kontogianni, C., Iliopoulou, M., Gaga, M., 2019. Lung nodules: A comprehensive review on current approach and management. *Annals of Thoracic Medicine* 14, 226–238. doi:10.4103/atm.ATM_110_19.
- [45] Moreira, I.C., Amaral, I., Domingues, I., Cardoso, A., Cardoso, M.J., Cardoso, J.S., 2012. Inbreast. *Academic Radiology* 19, 236–248. doi:10.1016/j.acra.2011.09.014.
- [46] Morgan, M., Cooke, M., Mccarthy, G., 2005. Microcalcifications associated with breast cancer: An epiphenomenon or biologically significant feature of selected tumors? *Journal of mammary gland biology and neoplasia* 10, 181–7. doi:10.1007/s10911-005-5400-6.
- [47] Park, S., Lee, S.M., Kim, N., Choe, J., Cho, Y., Do, K.H., Seo, J.B., 2019. Application of deep learning–based computer-aided detection system: detecting pneumothorax on chest radiograph after biopsy. *European Radiology* 29, 5341–5348. doi:10.1007/s00330-019-06130-x.
- [48] Qiu, J., Tan, G., Lin, Y., Guo, J., Dai, Z., Wang, F., Zhuang, C., Wilman, A.H., Huang, H., Cao, Z., Tang, Y., Jia, Y., Li, Y., Zhou, T., Wu, R., 2022. Automated detection of intracranial artery stenosis and occlusion in magnetic resonance angiography: A preliminary study based on deep learning. *Magnetic Resonance Imaging* 94, 105–111. doi:10.1016/j.mri.2022.09.006.
- [49] Redmon, J., Divvala, S., Girshick, R., Farhadi, A., 2016. You only look once: Unified, real-time object detection, in: 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 779–788. doi:10.1109/CVPR.2016.91.
- [50] Rijthoven, M.v., Swiderska-Chadaj, Z., Seeliger, K., Laak, J.v.d., Ciompi, F., 2018. You only look on lymphocytes once. *Medical Imaging with Deep Learning* URL: <https://openreview.net/forum?id=S10Ifw2oz>.
- [51] Samuelson, F., Petrick, N., 2006. Comparing image detection algorithms using resampling, in: 3rd IEEE International Symposium on Biomedical Imaging: Nano to Macro, 2006., pp. 1312–1315. doi:10.1109/ISBI.2006.1625167.
- [52] Savelli, B., Bria, A., Molinaro, M., Marrocco, C., Tortorella, F., 2020. A multi-context cnn ensemble for small lesion detection. *Artificial Intelligence in Medicine* 103, 101749. doi:10.1016/j.artmed.2019.101749.
- [53] Schultheiss, M., Schober, S.A., Lodde, M., Bodden, J., Aichele, J., Müller-Leisse, C., Renger, B., Pfeiffer, F., Pfeiffer, D., 2020. A robust convolutional neural network for lung nodule detection in the presence of foreign bodies. *Scientific Reports* 10, 12987. doi:10.1038/s41598-020-69789-z.
- [54] Shehab, M., Abualigah, L., Shambour, Q., Abu-Hashem, M.A., Shambour, M.K.Y., Alsalibi, A.I., Gandomi, A.H., 2022. Machine learning in medical applications: A review of state-of-the-art methods. *Computers in Biology and Medicine* 145, 105458. doi:https://doi.org/10.1016/j.combiomed.2022.105458.
- [55] Shen, D., Wu, G., Suk, H.I., 2017. Deep learning in medical image analysis. *Annual Review of Biomedical Engineering* 19, 221–248. doi:10.1146/annurev-bioeng-071516-044442.
- [56] Shen, W., Zhou, M., Yang, F., Yang, C., Tian, J., 2015. Multi-scale convolutional neural networks for lung nodule classification, in: Ourselin, S., Alexander, D.C., Westin, C.F., Cardoso, M.J. (Eds.), *Information Processing in Medical Imaging*, Springer International Publishing, p. 588–599. doi:10.1007/978-3-319-19992-4_46.
- [57] Soares, I., Castelo-Branco, M., Pinheiro, A., 2023. Microaneurysms detection in retinal images using a multi-scale approach. *Biomedical Signal Processing and Control* 79, 104184. doi:https://doi.org/10.1016/j.bspc.2022.104184.
- [58] Soun, J., Chow, D., Nagamine, M., Takhtawala, R., Filippi, C., Yu, W., Chang, P., 2021. Artificial intelligence and acute stroke imaging. *AJNR: American Journal of Neuroradiology* 42, 2–11. doi:10.3174/ajnr.A6883.
- [59] Suzuki, K., 2017. Overview of deep learning in medical imaging. *Radiological Physics and Technology* 10, 257–273. doi:10.1007/s12194-017-0406-5.
- [60] Swiderska-Chadaj, Z., Pinckaers, H., Rijthoven, M.v., Balkenhol, M., Melnikova, M., Geessink, O., Manson, Q., Sherman, M., Polonia, A., Parry, J., Abubakar, M., Litjens, G., Laak, J.v.d., Ciompi, F., 2019. Learning to detect lymphocytes in immunohistochemistry with deep learning. *Medical Image Analysis* 58, 101547. doi:https://doi.org/10.1016/j.media.2019.101547.
- [61] Toosi, A., Harsini, S., Ahamed, S., Yousefirizi, F., Bénard, F., Uribe, C., Rahmim, A., 2023. State-of-the-art object detection algorithms for small lesion detection in psma pet: use of rotational maximum intensity projection (mip) images, in: Colliot, O., Išgum, I. (Eds.), *Medical Imaging 2023: Image Processing*, SPIE, p. 124643E. doi:10.1117/12.2654527.
- [62] Tsiakakis, N., Theodoropoulos, D., Manikis, G., Ktistakis, E., Boutsora, O., Berto, A., Scarpa, F., Scarpa, A., Fotiadis, D.I., Marias, K., 2021. Deep learning for diabetic retinopathy detection and classification based on fundus images: A review. *Computers in Biology and Medicine* 135, 104599. doi:10.1016/j.combiomed.2021.104599.
- [63] Wang, C.W., Huang, S.C., Lee, Y.C., Shen, Y.J., Meng, S.I., Gaol, J.L., 2022. Deep learning for bone marrow cell detection and classification on whole-slide images. *Medical Image Analysis* 75, 102270. doi:10.1016/j.media.2021.102270.
- [64] Wang, J., Yang, Y., 2018. A context-sensitive deep learning approach for microcalcification detection in mammograms. *Pattern Recognition* 78, 12–22. doi:10.1016/j.patcog.2018.01.009.
- [65] Yu, X., Wang, J., Hong, Q.Q., Teku, R., Wang, S.H., Zhang, Y.D., 2022. Transfer learning for medical images analyses: A survey. *Neurocomputing* 489, 230–254. doi:10.1016/j.neucom.2021.08.159.
- [66] Yurdusev, A.A., Adem, K., Hekim, M., 2023. Detection and classification of microcalcifications in mammograms images using difference filter and yolov4 deep learning model. *Biomedical Signal Processing and Control* 80, 104360. doi:10.1016/j.bspc.2022.104360.

- [67] Zhao, Z.Q., Zheng, P., Xu, S.T., Wu, X., 2019. Object detection with deep learning: A review. *IEEE Transactions on Neural Networks and Learning Systems* 30, 3212–3232. doi:10.1109/TNNLS.2018.2876865.
- [68] Çalli, E., Sogancioglu, E., van Ginneken, B., van Leeuwen, K.G., Murphy, K., 2021. Deep learning for chest x-ray analysis: A survey. *Medical Image Analysis* 72, 102125. doi:10.1016/j.media.2021.102125.