



OPEN

Pneumonia detection based on RSNA dataset and anchor-free deep learning detector

Linghua Wu¹, Jing Zhang², Yilin Wang¹, Rong Ding², Yueqin Cao², Guijin Liu², Changsheng Liufu³, Baowei Xie², Shanping Kang⁴, Rui Liu⁵, Wenle Li⁶✉ & Furen Guan⁷✉

Pneumonia is a highly lethal disease, and research on its treatment and early screening tools has received extensive attention from researchers. Due to the maturity and cost reduction of chest X-ray technology, and with the development of artificial intelligence technology, pneumonia identification based on deep learning and chest X-ray has attracted attention from all over the world. Although the feature extraction capability of deep learning is strong, existing deep learning object detection frameworks are based on pre-defined anchors, which require a lot of tuning and experience to guarantee their excellent results in the face of new applications or data. To avoid the influence of anchor settings in pneumonia detection, this paper proposes an anchor-free object detection framework and RSNA dataset based on pneumonia detection. First, a data enhancement scheme is used to preprocess the chest X-ray images; second, an anchor-free object detection framework is used for pneumonia detection, which contains a feature pyramid, two-branch detection head, and focal loss. The average precision of 51.5 obtained by Intersection over Union (IoU) calculation shows that the pneumonia detection results obtained in this paper can surpass the existing classical object detection framework, providing an idea for future research and exploration.

Pneumonia, one of the most deadly diseases worldwide, is particularly deadly in children¹. In 2015 alone, nearly 1 million children died from pneumonia and its complications². With the popularity of chest radiography and the decreasing price of the equipment, the use of chest radiography for pneumonia detection has attracted the attention of a large number of researchers and clinicians³. The existing chest X-ray-based pneumonia tests require, firstly, a large amount of clinical data for learning and teaching training. Second, it requires specialized radiologists as well as clinicians to make the diagnosis. The diagnosis is based on markers, clinical manifestations, and the results of physiological tests^{4,5}. Currently, with the rapid development of artificial intelligence, convolutional neural networks are used in a large number of fields such as image and signal processing due to their powerful feature extraction capabilities in Euclidean space. Convolutional neural networks are equally powerful in the field of medical image and signal research^{6–8}. However, based on current research, although deep learning-based object detection frameworks can achieve excellent results, they require a lot of tedious parameter settings and experience to guarantee the effectiveness of object detection frameworks because their frameworks are anchor-based^{9,10}. In other words, anchor-based object detection requires manual setting of the preset bounding box used for scanning in order to achieve good object detection results. For example, we use a large preset bounding box with an approximate aspect ratio of 1:2 for lung detection with a large rectangular object, but if we use this object detection framework for lung nodule detection, we need to use a smaller square preset bounding box with an aspect ratio of 1:1. In the case of more complex when dealing with more complex subjects or setups, more complex parameter settings are required, which also places higher demands on the user¹¹, which is clearly impractical and not user-friendly for clinicians without code and computer engineering experience.

Based on this, an anchor-free pneumonia detection framework based on this paper is proposed. First, the chest X-ray images undergo an image preprocessing step with denoising and image enhancement to expand

¹Internal Medicine Department, Taizhou Fifth People's Hospital, Taizhou, China. ²Respiratory and Critical Care Medicine, Taizhou Fourth People's Hospital, Taizhou, China. ³Department of Gerontology, Dongguan First Hospital Affiliated to Guangdong Medical University, Dongguan, China. ⁴Department of Gerontolog, Ruikang Hospital Affiliated to Guangxi University of Chinese Medicine, Nanning, China. ⁵Respiratory and Critical Care Medicine, Ruikang Hospital Affiliated to Guangxi University of Chinese Medicine, Nanning, China. ⁶State Key Laboratory of Molecular Vaccinology and Molecular Diagnostics and Center for Molecular Imaging and Translational Medicine, School of Public Health, Xiamen University, Xiamen, China. ⁷Emergency Department, Zhuhai Hospital of Integrated Chinese and Western Medicine, Zhuhai, China. ✉email: Drlee0910@163.com; 1078389699@qq.com

the training data set, giving the network framework a better training performance. Second, the preprocessed images are fed into a feature pyramid-based anchor-free detection framework, which does not require a preset bounding box but instead a two-branch detection head, which is used to detect the centroid of the object and its scale, respectively. In addition, in order to solve the object detection. In addition, to address the class imbalance problem in object detection, focal loss¹² is used as the loss function of the object detection framework. Finally, the final results of pneumonia detection are obtained by computing the IoU of the RSNA dataset¹³.

Our contributions can be summarized as follows.

- (1) A series of data pre-processing methods were used to denoise and expand the chest X-ray images, thus enhancing the training of the object detection framework;
- (2) An anchor-free detection framework was used for pneumonia detection, where feature pyramids, two-branch detection head, and focal loss were used;
- (3) The RSNA dataset is used to verify the effectiveness of the proposed framework in this paper, and the IoU experimental results demonstrate that the proposed framework can achieve good object detection results WITHOUT preset bounding box.

The remaining chapters of this paper are organized as follows: Chapter 2 gives an introduction to the methods used in this paper, Chapter 3 gives the experimental data set and the experimental results, and the last chapter presents the conclusions and the related discussion.

Related works

With the advent of deep learning as a model with superb feature extraction capabilities on image and signal structure, the object detection task has also shifted from traditional corner detection and circle detection, for example, to deep learning-based object detection^{14,15}. The initial frameworks are based on sliding windows, and this class of frameworks designs a series of sliding windows of different shapes and sizes, which are subsequently slid from the image to be measured and deep learning is used to classify each of the intercepted images. Representative works of such methods include Fast RCNN¹⁶ and others, also known as two-step detectors^{17,18}. In addition, there is another class of one-step detectors, represented by YOLO¹⁹ and SSD²⁰, among others. This type of framework classifies sliding windows directly, saving the time of interception before classification. In contrast, two-step detectors are generally considered to have higher classification accuracy, but their training and testing speeds are relatively slower. One-step detectors are less accurate, but due to their processing speed, they are also heavily used in real-world scenarios with higher real-time requirements.

Deep learning-based detection frameworks, although achieving excellent results and being used extensively in industrial and medical scenarios, among others^{21,22}, have attracted a lot of attention in recent years. Despite the superior effectiveness of these frameworks and the possibility of choosing one-step or two-step frameworks depending on the actual needs, their framework effectiveness depends heavily on the effectiveness of the anchor^{23,24}. In other words, when a trained model is used in other applications, the anchor settings need to be modified in order to obtain better results. Therefore, the effect of the anchor-based framework largely depends on the user's experience, and the framework has poor generalization performance. In the case of a large domain gap, its framework test results are not good. Based on the above, there are some novel approaches in recent years to design frameworks that discard anchors. For example, the anchor-guide framework²⁵, uses guided anchors to replace the traditional complex design of anchors, thus avoiding the drawbacks associated with preset anchors. Besides, there are some novel anchor-free frameworks, including FCOS²⁶ and CSP²⁷, which use a point-based scheme to find the point with the highest classification accuracy and thus determine the centroid location of the object to be measured. A summary of related works is shown in Table 1.

Method

Overall framework

The overall framework used in this paper is shown in Fig. 1. Firstly, the input images are first subjected to a series of data augmentation, which aims to expand the dataset through this scheme, thus improving the training effect and reducing overfitting. Secondly, the data-expanded input images are fed into the detection framework for training. The detection framework is an anchor-free detector, which does not require a preset anchor to scan the image to be measured, thus making its detection effect independent of the parameters and settings of the

Method	Use anchor?	One/two-step
Fast RCNN ¹⁶	Y	Two
YOLO ¹⁹	Y	One
SSD ²⁰	Y	One
Guide anchor ²⁵	N	Two
FCOS ²⁶	N	One
CSP ²⁷	N	One

Table 1. A summary of related works.

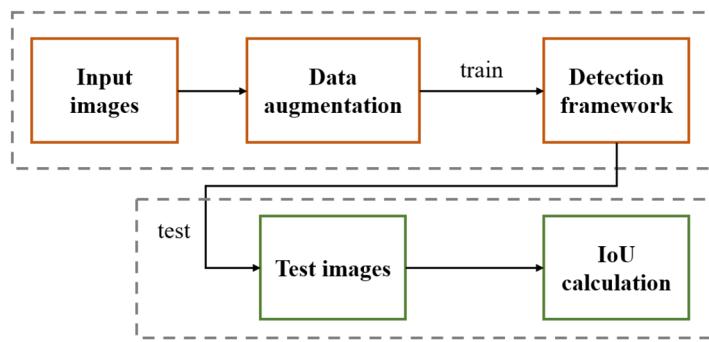


Figure 1. Overall flow chart of the framework.

anchor, etc., and achieving more robust results. After training, the framework detects the image to be measured so that the IoU results are calculated and the final detection results are obtained.

Data augmentation

The data augmentation means include a series of methods. First, the image is flipped horizontally and vertically to expand the original input image by a total of $4\times$. Second, the image is mirrored and then flipped both horizontally and vertically to obtain a total of $8\times$ expansion.

In addition to the basic data augmentation, luminance augmentation and random cropping are also employed. Luminance augmentation refers to randomly changing the luminance parameter of the input image, while the luminance is randomly generated for each training, which can effectively adapt the detection framework to each type of test image and thus obtain better detection results. Random cropping means that the input image is cropped randomly, i.e., only some information about the input image is retained. Like the luminance augmentation, random cropping also makes the training input different for each training session, thus making the overall framework more robust for testing.

Detection framework

One of the most serious problems in object detection is the inconsistent scale of the object. In the past years, only one feature extraction module has been applied in deep learning-based detectors, which is hard to detect all the objects. Specifically, if the well-trained model shows a good detection performance on big-scale objects, the detection results will greatly decrease on tiny-scale ones. Hence, a feature pyramid module is adopted to extract features with different scales, which has been verified by numerous detection frameworks.

The detection framework used in this paper is shown in Fig. 2. First, an image to be measured is input to a five-layer backbone, B1 to B5, and then each layer passes through a 1×1 convolution kernel to form a five-layer feature pyramid, L1 to L5, where the strides of each layer are 8, 16, 32, 64, and 128, respectively. As we all know, object detection tasks include two subtasks: classification and localization. Hence, we also introduce a set of two-branch detection heads for these subtasks. In each detection head, a 3×3 convolutional kernel is passed first, and then two 1×1 convolutional kernels are passed to predict the two branches, respectively.

A series of loss functions are applied in the detection framework. One of them is for CENTER prediction, as it faces a serious class imbalance problem. In other words, since the number of CENTER points in an image is very unbalanced with the number of pixel points in the background, there are usually only a few points or even no points in a graph that belong to the CENTER category. Hence, the traditional cross-entropy loss function

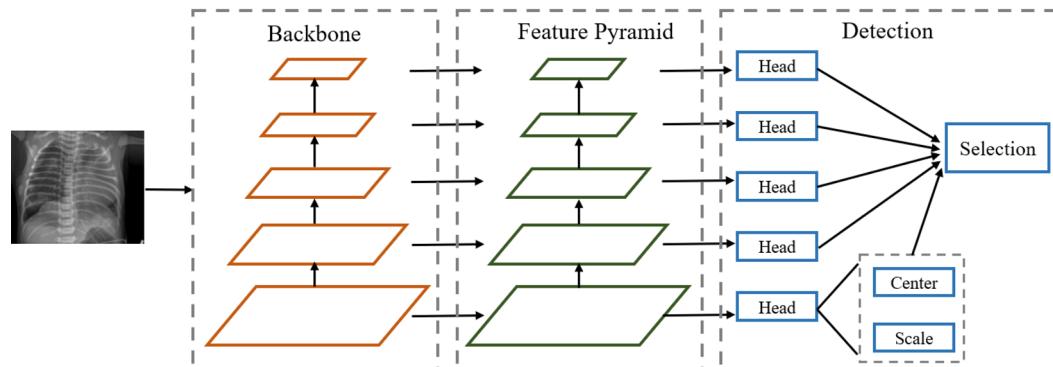


Figure 2. Schematic diagram of the detection framework used in this paper.

usually performs poorly when facing such a situation. Based on this, this paper adopts a focal loss¹² as the loss function of the center:

$$L_{center} = -\frac{1}{N} \sum_{i=1}^{W/r} \sum_{j=1}^{H/r} \alpha_{ij} (1 - p_{ij})^\gamma \log(p_{ij}) \quad (1)$$

where p_{ij} is the probability of position (i, j), ranging from 0 to 1. α_{ij} is the weight parameter.

In addition to center, a Smooth L1 loss has also been applied to scale predictions as follows:

$$L_{scale} = \frac{1}{N} \sum_{n=1}^N \text{SmoothL1}(s_n, gt_n) \quad (2)$$

where N is the number of objects to be measured, s is the predicted result, and gt represents the baseline label.

Since the five detection heads train and predict the feature map separately, some uncertainty arises between each detection head. In this paper, we use a selection strategy, i.e., the largest stride in the prediction result is used as the best detection result for retention. Since the feature size is not consistent in each layer, we also set the size of the bounding boxes in each layer. Suppose the maximum distance of the regression of layer i on a certain object is m_i and the regression point is (l, t, r, b) . If $m_i < (l, t, r, b) < m_{i-1}$, then this current regression point is considered as a sample of the object. Hence, we can limit the prediction uncertainty between multiple feature layers by setting the maximum distance of each layer.

Testing and IoU computing

Intersection over Union (IoU) is a common measure of the difference between the tested bounding box and the real box in object detection. In other words, the larger the IoU result, the better the framework is for object detection, and vice versa. The IoU is calculated as follows:

$$IoU = \frac{A \cap B}{A \cup B} = \frac{C}{A + B - C} \quad (3)$$

Evaluation Metrics

Average precision (AP), one of the most commonly used experimental metrics for object detection, is adopted in this paper to evaluate the detection performance, which is defined as follows:

$$\text{Average Precision}(AP) = \int_{r=0}^1 p(r) dr$$

where p(r) represents the Precision-Recall curve calculated via the confusion matrix. In other words, AP is the area under the Precision-Recall curve.

Among them, AP, without special settings and explanations, defaults to the number of detected frames plus one if the test frame and the groundtruth label are considered to have overlapped at $\text{IoU} > 0.5$. In addition, the common AP_S , AP_M and AP_L metrics represent the detected AP values for small, medium and large objects, respectively. Since there is no corresponding small object in the dataset used in this paper, AP_M and AP_L are used as the metrics. In addition, AR_{10} , AR_M and AR_L , which are measures of average Recall, are also used as the experimental metrics in this paper.

Experimental results

Experimental dataset

A dataset for pneumonia detection, RSNA¹³, was adopted as the training and testing dataset for this paper, which was published on the Kaggle platform. In this dataset, there are three different detection classifications, namely normal lung, lung without pneumonia but abnormal lung, and lung images with pneumonia. In total, 6012 different patients had at least one or more pneumonia conditions, which are collected by 18 radiologists from 16 different institutions. The mean value of labeler demographics' experience is 10.6 years. The basic imaging examples of normal and pneumonia are shown in Fig. 3.

Experimental setup

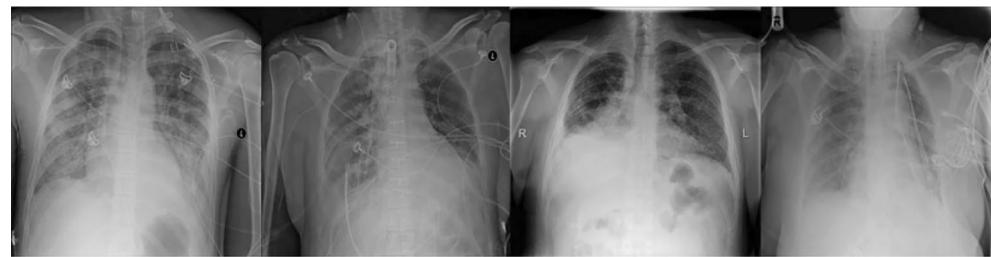
At the software level, ResNet-50²⁸ was used to detect the backbone of the framework. The initial learning rate was set to 0.001, and the learning rate was reduced to 1/10th of the original rate after every 100 cycles with a variable learning rate setting. Anchor-free NMS was set to 0.5. ADAM²⁹ was used to detect the optimizer of the framework. The settings and parameters of small, medium, and large objects follow COCO datasets³⁰. The confidence score threshold for the centroid was set to 0.1. We built using the Python programming language and the Pytorch deep learning framework. The batch size was set to 1. The whole framework was experimented and accelerated on a single RTX 2080ti.

Experimental results

The detection results of the proposed framework are shown in Table 2. Where "Ours" represents the object detection framework using only the automatically selected feature layer proposed in this paper. "Ours + DA" represents the effect of using the detection framework of this paper with data augmentation. "Ours + DA + FL" represents the result of this paper with data enhancement and focal loss function. It can be seen from the table



(a) Normal cases



(b) Pneumonia-related cases

Figure 3. Some samples from RSNA dataset.

Method	AP	AP _M	AP _L	AR ₁₀	AR _M	AR _L
Ours	48.6	22.4	55.1	96.1	89.6	96.3
Ours + DA	50.2	25.4	57.0	96.1	90.1	96.8
Ours + DA + FL	51.5	27.6	58.9	96.6	92.6	98.9

Table 2. Detection effect of the framework in this paper with ablation study.

that both improvements proposed in this paper can effectively improve the experimental metrics. In particular, the AP can be improved from 48.6 in the base framework to 50.2 and 51.5. For AP_M and AP_L, the two boosts proposed in this paper can also improve their decibels by about 5% and 4%. In addition to the accuracy results, the experimental results of 96.6, 90.6 and 98.9 can be effectively achieved for the three recall results in this paper.

To better compare the experimental results of the proposed framework and other papers in this paper, a comparison is shown in Table 3. SSD, full name Single Shot MultiBox Detector²⁰, is an object detection algorithm proposed by Wei Liu at ECCV 2016, which has an obvious speed advantage over Faster R-CNN and an obvious detection performance advantage over YOLO. As a one-stage detector, SSD inherits the idea of converting detection into regression from YOLO and completes object localization and classification at one time. In addition, SSD also proposes a similar Prior box based on the Anchor in Faster R-CNN, and adds a Pyramidal Feature Hierarchy based detection method, which predicts objects on feature maps with different perceptual fields.

After the design experience of R-CNN and Fast R-CNN, Ross B. Girshick proposed the new Faster RCNN in 2016³¹, which structurally has integrated feature extraction, proposal extraction, bounding box regression, and classification all integrated in one network, which makes the comprehensive performance improved. Compared with other two-stage detectors, Faster R-CNN is particularly fast in terms of detection speed. In addition, the detection accuracy is greatly improved compared to one-stage detectors.

Method	AP	AP _M	AP _L	AR ₁₀	AR _M	AR _L
SSD	25.2	10.4	30.6	74.1	52.7	86.4
Faster R-CNN	36.5	19.2	42.7	88.9	65.3	96.2
RetinaNet	45.8	25.5	52.5	97.7	92.6	99.3
FCOS	48.6	24.7	53.3	95.6	90.2	96.7
NYNet	49.1	22.7	57.6	94.2	88.7	96.1
Ours + DA + FL	51.5	27.6	58.9	96.6	92.6	98.9

Table 3. Comparison table.

RetinaNet³² is a unified object detection network consisting of a backbone and two subnetworks. The main role of the backbone is to obtain a feature map of the whole input image through a series of convolution operations. The two subnetworks perform object classification and position regression based on the output feature map, respectively. Compared with the original FPN, the convolution process of RetinaNet uses ResNet, and the upsampling and side-joining are still FPN structures. Through the backbone network, a multi-scale feature pyramid is generated. Then two sub-networks are connected at the back for classification and regression, respectively.

Most of the current advanced object detection models, such as RetinaNet, SSD, YOLOv3, and Faster R-CNN, rely on predefined anchor boxes. In contrast, FCOS²⁶ is an anchor box free framework and is also proposal free, that is, it does not rely on predefined anchor boxes or proposal regions. By removing the pre-defined anchor boxes, FCOS completely avoids the complex operations on anchor boxes, such as calculating the overlap during training, and saves the memory usage during training. More importantly, all hyperparameters related to the anchor frame that are very sensitive to the final detection results are avoided in this paper. Since only non-maximum suppression (NMS) is used for post-processing, the FCOS proposed in this paper has the advantage of being simpler than previous first-order detectors based on anchor frames. In addition, an enhanced FCOS (NYNet), which is proposed by Yan et al.³³, is adopted for comparisons. Specifically, for fair competition, the detection performance using ResNet50 is included in Table 3.

By comparing with the above classical detection frameworks, we can find that the proposed framework can achieve the best AP results. In particular, compared with SSD, Faster R-CNN, RetinaNet, and FCOS, the proposed framework achieves about 26%, 15%, 6%, and 3% higher APs, respectively. Similarly, for AP_M and AP_L , the proposed framework achieves about 17% and 28% improvement compared with SSD, which is the worst of the classical frameworks, respectively. When compared with the best classical framework, FCOS and NYNet, this framework still achieves about 3% and 2% improvement in AP, respectively. As for the AR-related metrics, since the previous classical work has achieved results close to 100, this paper can still achieve results greater than 90. Only in two items, AR_{10} and AR_L , the comparison is about 1% lower than RetinaNet. From the table of comparison results, it can be seen that the proposed framework can effectively achieve the detection of pneumonia and achieve better results than the classical algorithms.

The schematic diagram of the real test on the RSNA dataset is shown in Fig. 4. In the figure, the white bounding box is the groundtruth label, and the red box is the detection result of the proposed framework in this paper. From Fig. 4, it can be seen that on the first and last graphs, the framework of this paper can effectively detect the object of pneumonia, and the red and white boxes have a relatively high overlap. In both the second and third images, the proposed framework in this paper can detect the right side of the lung in place, but the detection effect is poor, probably due to the lighter color of the left image.

Conclusion

In this paper, an anchor-free pneumonia detection framework is proposed. First, a series of data pre-processing methods are used to denoise and expand the chest X-ray images, which in turn expands the training set and improves the training effect of the detection framework. Secondly, an anchor-free detection framework incorporating feature pyramids is proposed, and ambiguity between feature layers is avoided by a selection mechanism. Finally, focal loss is used as the loss function of the framework in this paper to solve the extreme class imbalance problem in pneumonia detection. The experimental results on RSNA show that the proposed framework can effectively improve pneumonia detection. Compared with four classical object detection frameworks, including SSD, Faster R-CNN, RetinaNet, and FCOS, the proposed framework can achieve superior results, thus verifying the effectiveness and advancement of this paper's method.

Discussion

In contrast to traditional object detection frameworks, the proposed framework does not need to be based on the anchor structure, making its effect dependent on fewer parameters to achieve object detection and localization. In addition, for some applications that require fine-tuning of parameters such as bounding box and anchor, the anchor-free framework used in this paper can also avoid these problems. In other words, for pneumonia or other medical-related object detection tasks, when the users are clinicians or medical experts, who often do not have the ability to tune the parameters for AI systems, the anchor-free framework is more friendly as an object detection framework with fewer parameters and higher applicability for such users.

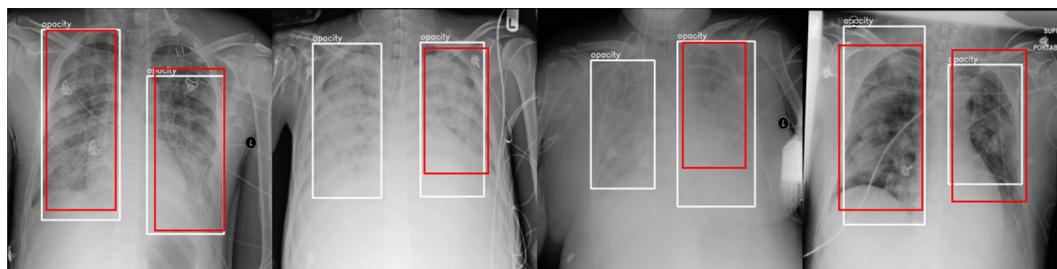


Figure 4. Schematic diagram of the RSNA dataset measured in practice.

Although the framework has been validated on a dataset for pneumonia detection, the effectiveness of improved schemes such as data augmentation and loss functions has also been verified. The main advantage of this anchor-free framework is that it addresses the problem that traditional detection frameworks rely too much on the parameter design of the anchor and thus design a framework that can achieve good object detection results without fine-tuning the anchor. However, the effectiveness of this anchor-free framework needs more data sets and more applications to be verified, such as pedestrian detection in autonomous driving, vehicle and ship detection in remote sensing images, and underwater object detection. In addition, the data augmentation and loss function improvement tools in this paper focus on the widespread problems in medical image object detection, including the high noise level in medical images and the small number of objects to be measured. However, whether these improvements can be applied to other medical image detection, such as tumor detection, spontaneous brain hemorrhage detection, and blood vessel detection, needs further experimental validation in future research.

Data availability

The dataset for this study can be requested from the RSNA Pneumonia Detection Challenge, (<https://www.rsna.org/education/ai-resources-and-training/ai-image-challenge/rsna-pneumonia-detection-challenge-2018>).

Received: 11 July 2023; Accepted: 15 January 2024

Published online: 22 January 2024

References

- Mandell, L. A. & Niederman, M. S. Aspiration pneumonia. *N. Engl. J. Med.* **380**(7), 651–663 (2019).
- Papazian, L., Klompas, M. & Luyt, C.-E. Ventilator-associated pneumonia in adults: a narrative review. *Intensive Care Med.* **46**(5), 888–906 (2020).
- Hammoudi, K. *et al.* Deep learning on chest X-ray images to detect and evaluate pneumonia cases at the era of COVID-19. *J. Med. Syst.* **45**, 75 (2021).
- Yang, F. *et al.* Pneumoconiosis computer aided diagnosis system based on X-rays and deep learning. *BMC Med Imaging* **21**, 1–7 (2021).
- Khatri, A., *et al.* Pneumonia identification in chest X-ray images using EMD. In *Trends in Communication, Cloud, and Big Data: Proceedings of 3rd National Conference on CCB, 2018* (Springer Singapore, 2020).
- Sarvamangala, D. R. & Raghavendra, V. K. Convolutional neural networks in medical image understanding: a survey. *Evol. Intell.* **15**(1), 1–22 (2022).
- Xie, Y., *et al.* Cotr: Efficiently bridging cnn and transformer for 3d medical image segmentation. In *Medical Image Computing and Computer Assisted Intervention–MICCAI 2021: 24th International Conference, Strasbourg, France, September 27–October 1, 2021, Proceedings, Part III* **24** (Springer International Publishing, 2021).
- Wu, E. Q., *et al.* Detecting Alzheimer's dementia degree. *IEEE Trans. Cogn. Dev. Syst.* **14**(1), 116–125 (2020).
- Zou, Z., *et al.* Object detection in 20 years: A survey. In: *Proceedings of the IEEE* (2023).
- Tang, Z.-R. *et al.* Multi-expert learning for fusion of pedestrian detection bounding box. *Knowl. Based Syst.* **241**, 108254 (2022).
- Guo, Y., *et al.* Improved YOLOv4-CSP algorithm for detection of bamboo surface sliver defects with extreme aspect ratio. *IEEE Access* **10**, 29810–29820 (2022).
- Tran, G. S., *et al.* Improving accuracy of lung nodule classification using deep learning with focal loss. *J. Healthc. Eng.*, 2019 (2019).
- RSNA Pneumonia Detection Challenge, <https://www.rsna.org/education/ai-resources-and-training/ai-image-challenge/rsna-pneumonia-detection-challenge-2018>
- Glumov, N. I., Kolomiyetz, E. I. & Sergeyev, V. V. Detection of objects on the image using a sliding window mode. *Opt. Laser Technol.* **27**(4), 241–249 (1995).
- Zhang, L., Zhao, J. & Li, W. Online and unsupervised anomaly detection for streaming data using an array of sliding windows and PDDs. *IEEE Trans. Cybern.* **51**(4), 2284–2289 (2019).
- Girshick, R. Fast r-cnn. In *Proceedings of the IEEE International Conference on Computer Vision* (2015).
- Wang, X., Abhinav, S., Abhinav, G. A-fast-rcnn: Hard positive generation via adversary for object detection. In: *Proceedings of the IEEE conference on computer vision and pattern recognition* (2017).
- Sun, X., Pengcheng, W. & Hoi, S. C. H. Face detection using deep learning: An improved faster RCNN approach. *Neurocomputing* **299**, 42–50 (2018).
- Redmon, J., *et al.* You only look once: Unified, real-time object detection. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (2016).
- Liu, W., *et al.* Ssd: Single shot multibox detector. In *Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11–14, 2016, Proceedings, Part I* **14** (Springer International Publishing, 2016).
- McLeavy, C. M., *et al.* The future of CT: Deep learning reconstruction. *Clin. Radiol.* **76**(6), 407–415 (2021).
- Chilamkurthy, S., *et al.* Deep learning algorithms for detection of critical findings in head CT scans: A retrospective study. *Lancet* **392**(10162), 2388–2396 (2018).
- Zhong, Y., *et al.* Anchor box optimization for object detection. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision* (2020).
- Kong, T., *et al.* Foveabox: Beyond anchor-based object detection. *IEEE Trans. Image Process.* **29**, 7389–7398 (2020).
- Wang, J., *et al.* Region proposal by guided anchoring. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (2019).
- Tian, Z., *et al.* Fcos: A simple and strong anchor-free object detector. *IEEE Trans. Patt. Anal. Mach. Intelli.* **44**(4), 1922–1933 (2020).
- Liu, W., *et al.* High-level semantic feature detection: A new perspective for pedestrian detection. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (2019).
- He, K., *et al.* Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (2016).
- Kingma, D. P., Jimmy, B. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980* (2014).
- Lin, T. Y., *et al.* Microsoft coco: Common objects in context. In *Computer Vision–ECCV 2014: 13th European Conference, Zurich, Switzerland, September 6–12, 2014, Proceedings, Part V* **13** (Springer International Publishing, 2014).
- Ren, S., *et al.* Faster r-cnn: Towards real-time object detection with region proposal networks. *Adv. Neural Inf. Process. Syst.*, 28 (2015).
- Lin, T.-Y., *et al.* Focal loss for dense object detection. In *Proceedings of the IEEE international conference on computer vision* (2017).
- Yan, N. & Tao, Y. Pneumonia X-ray detection with anchor-free detection framework and data augmentation. *Int. J. Imaging Syst. Technol.* <https://doi.org/10.1002/ima.22860> (2023).

Author contributions

G.F.R. and W.L.L. completed the study design. L.H.W. performed the study, collected and analyzed the data. L.H.W., J.Z. and Y.L.W. drafted the manuscript. R.D., Y.Q.C., G.Q.L. and C.S.L.F. provided the expert consultations and suggestions. B.W., S.Q.K. and R.L. conceived of the study, participated in its design and coordination, and helped to embellish language. All authors reviewed the final version of the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to W.L. or F.G.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2024