

RESEARCH

Open Access



DCNAS-Net: deformation convolution and neural architecture search detection network for bone marrow oedema

Chengyu Song¹, Shan Zhu^{2†}, Yanyan Liu^{3*}, Wei Zhang¹, Zhi Wang², Wangxiao Li¹, Zhenye Sun², Peng Zhao² and Shengzhang Tian²

Abstract

Background Lumbago is a global disease that affects more than 500 million people worldwide. Bone marrow oedema is one of the main causes of the condition and clinical diagnosis is mainly made by radiologists manually reviewing MRI images to determine whether oedema is present. However, the number of patients with Lumbago has risen dramatically in recent years, which has brought a huge workload to radiologists. In order to improve the efficiency of diagnosis, this paper is devoted to developing and evaluating a neural network for detecting bone marrow edema in MRI images.

Related work Inspired by the development of deep learning and image processing techniques, we design a deep learning detection algorithm specifically for the detection of bone marrow oedema from lumbar MRI images. We introduce deformable convolution, feature pyramid networks and neural architecture search modules, and redesign the existing neural networks. We explain in detail the construction of the network and illustrate the setting of the network hyperparameters.

Results and discussion The detection accuracy of our algorithm is excellent. And its accuracy of detecting bone marrow oedema reached up to 90.6%, an improvement of 5.7% compared to the original. The recall of our neural network is 95.1%, and the F1-measure also reaches 92.8%. And our algorithm is fast in detecting it, taking only 0.144 s per image.

Conclusion Extensive experiments have demonstrated that deformable convolution and aggregated feature pyramid structures are conducive for the detection of bone marrow oedema. Our algorithm has better detection accuracy and good detection speed compared to other algorithms.

Keywords Lumbago MRI, Bone marrow oedema, Neural networks, Target detection

Introduction and background

Lumbago is one of the major global health problems, with a prevalence of 11–84% [1]. According to the Global Burden of Disease, lumbago troubles over 500 million people in 2015, 17.3% higher than that in 2005, and it is one of the leading causes of disability in most countries [2]. The clinical assessment of lumbago involves taking a correct history and performing a thorough physical examination and imaging of the low back [3]. There are many causes

[†]Shan Zhu is a Equal contributor

*Correspondence:

Yanyan Liu

lytianjin@nankai.edu.cn

¹Tianjin University, Tianjin, China

²Tianjin Hospital, Tianjin University, Tianjin, China

³Nankai University, Tianjin, China



of lumbago, of which bone marrow oedema is the most common one. Bone marrow oedema in the lumbar region is mainly examined clinically by MRI. Typically, patients have MRI images of the lumbar region taken with medical equipment and these MRI images are described in a radiology report by a radiologist who are responsible for evaluating and tracking the lesion site quantitatively [4]. With the dramatic increase in the number of patients suffering from bone marrow oedema in recent years, the corresponding number of MRI images has also increased dramatically. The enormous workload of reviewing the images has put great pressure on the doctors concerned, making them more prone to misdiagnosis and missed diagnoses, which has caused great disturbance to doctors and patients' recovery. As a result, new techniques are urgently needed to be applied to this field.

In this paper, we design a new neural network to detect bone marrow edema. Different from the general classification network, our network can not only accurately determine whether the MRI image is abnormal, but also accurately give the location of bone marrow edema. The paper is organized as follows. First, we discuss the research and application of deep learning in MRI image processing at present. Then, aiming at the problem of bone marrow edema detection, we introduce deformable convolution module and multilevel feature pyramid structure to improve the detection effect of neural network, and introduce the latest neural structure search module to streamline the network. Then we proved the effectiveness of our improvement through a large number of experiments. Finally, we summarize and look forward to the research work of this paper.

Related work

In recent years, with the progress of image processing technology, researchers have developed CAD (Computer Aided Detection) systems, which are able to be partly automated in diagnostic process. At the same time, deep learning shines in the field of image processing. The medical image detection based on deep learning has also made great progress. Different from the traditional methods, the deep learning method is based on the characteristics of images and targets to learn features, which can effectively identify images and targets with different styles. At the same time, the deep learning algorithm is easier to design and deploy. Therefore, researchers actively apply the idea of deep learning to the field of medical images, including a large number of MRI image processing and analysis. For example, Liao et al. [5] proposed a neural network model for automatic detection of lung nodules by improving the U-net network. They added a 3D convolutional module to the U-net network to increase the accuracy of detecting lung nodules

in MRI images. Dezsó Ribli et al. [6] proposed a CAD system based on the Faster-R-CNN, which improved the network by employing a reactive propagation mechanism with weight decay and a stochastic gradient descent mechanism, and it effectively improved the detection accuracy of malignant breast lesions. Huang Tao et al. [7] designed a neural network specifically for grading the severity of pleural effusion. It effectively improved the accuracy of pleural effusion grading by designing three data processing methods and two loss functions. In the field of bone marrow disease detection, Klontzas et al. [8] designed a classification model based on CNN ensemble. They combined three kinds of CNN, VGG-16, Inception-V3 and Inception-ResNet-V2, into a CNN ensemble. In their experiment, the AUC (Area Under Curve) of the model for detecting bone marrow diseases increased from 93.68 to 95.97%. And the detection effect of this CNN ensemble has reached the same or higher performance of a radiologist. Lee et al. [9] develop a method for detecting bone marrow edema by MRI of the sacroiliac joints and a deep-learning network. It is found that the classification network based on ResNet18 has the best performance, which shows that the detection accuracy of bone marrow edema can reach 94.59%. Their research suggests that analysis based on deep learning can be an effective supplementary means for clinicians to diagnose bone marrow edema.

However, in the field of medical image processing, most of the work combined with deep learning focuses on image classification. Although the classification task can quickly identify the abnormal images, it can't accurately point out the abnormal positions on the images. For MRI images of lumbar region, there are many organs and tissues, which are similar to the features of bone marrow edema. Coupled with the diversity in the size and shape of bone marrow oedema, current network models perform poor effectiveness in detecting bone marrow oedema in MRI of the lumbar region. As shown in Fig. 1, bone marrow edema varies greatly on different MRI images. We can easily find out the bone marrow edema in Fig. 1a, while Fig. 1b–d needs more careful observation to find it. Moreover, the background elements in MRI images are very different due to the different slices, which will further reduce the accuracy of bone marrow edema detection. In general, there are three main problems with the detection of bone marrow oedema:

- Variable shape and size of bone marrow edema on MRI of the lumbar region. Conventional convolution networks are not effective in extracting irregular targets, especially bone marrow oedema.
- Lumbar MRI images contain a large amount of organ tissue, and bone marrow oedema has a high degree

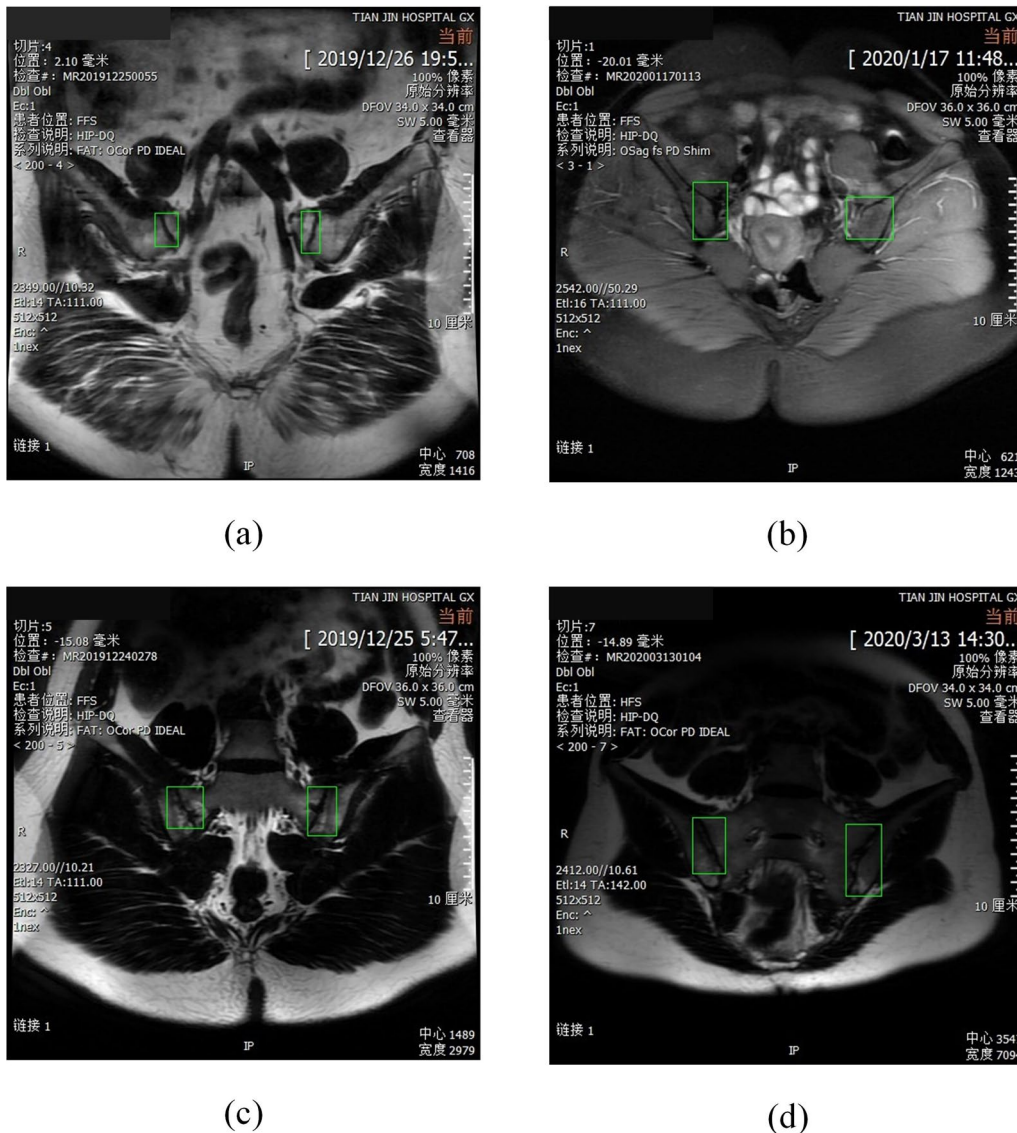


Fig. 1 Different styles of MRI images and different shapes of bone marrow edema. **a** Easily distinguishable bone marrow oedema; **b** Indistinguishable bone marrow oedema; **c** Indistinguishable bone marrow oedema; **d** Indistinguishable bone marrow oedema

of similarity to these. In such images, conventional feature extraction modules are unable to accurately extract the features of weak targets such as bone marrow oedema.

- MRI images taken by different models of equipment can be significantly different. Conventional algorithms are generally only effective for fixed style images and are less effective in processing MRI images from different devices.

In order to solve these problems, we design a new neural network to detect bone marrow oedema in MRI images.

Our work is inspired by the idea of the deformable convolution of DCN [10] networks. We replace the regular convolution and pooling modules of the feature extraction layer with deformable convolution and deformable RoI (Region of Interest) pooling modules, which can improve the detection ability of bone marrow edema targets with different shapes. In addition, on the basis of the FPN (Feature Pyramid Networks) optimization ideas in networks such as FPN [11], PA-Net [12], Libra R-CNN [13], SEPC-Neck [14], BIFPN [15] and OPANAS [16], our research introduces one single path feature aggregation module based on NAS (Neural Architecture Search) [17,

18]. The module automatically searches for the optimal feature extraction module by means of 4 unique feature extraction paths. It improves the ability to extract smaller bone marrow oedema in MRI images. In addition to this, our algorithm inserts 2 special parameter-free information paths, which can reduce the redundancy of the algorithm and the phenomenon of overfitting. Because of the combination of the deformable convolution module and the NAS module, our algorithm is called DCNAS-Net (Deformable Convolution and Neural Architecture Search Networks).

Methods

Structure of DCNAS-Net

The structure of DCNAS-Net is shown in Fig. 2. DCNAS-Net is structured with ResNet-50 as backbone, with the addition of the deformable convolution module in conv3-conv5. Then it is followed by the NAS-FPN module. It constructs the automatic search space by means of 4 unique information paths and 2 parameter-free information paths. 6 paths form different connections between the backbone network and the detection head. It will form a complementary and efficient aggregation module. The output of the NAS-FPN is superimposed on the feature matrix of the backbone. Then there is the deformable RoI pool layer. Finally, a 1024-dimensional fully connected layer is accessed. The inference information is then returned to the original graph by prediction frames and predictive classification.

Deformable convolution and deformable RoI pooling layers

One of the factors why bone marrow edema is difficult to be detected in MRI images is its variable size and shape. The conventional convolution module is limited

to a fixed ensemble structure and its convolution unit samples the input feature map at a fixed location. In addition, the pooling layer reduces the spatial resolution by a fixed ratio. Obviously, it lacks the internal mechanism for geometric transformations. Therefore, it performs poorly when detecting variable shaped targets, especially in cases such as bone marrow oedema. In general, expanding the data samples could solve this problem to some extent. However, due to the specific nature of medical data, it is impossible to collect a large amount of data. To address this problem we add the deformable convolution and deformable RoI pooling layer module to the backbone network. The two modules can improve the ability of the backbone network to extract irregularly shaped features. The working mechanism of the deformable convolution is shown as Fig. 3. We add a 2D offset to the standard convolution module, which allows sampling grids to be freely deformed and enables the neural network to adapt to the different shapes of the bone marrow oedema. The density of deformable convolution and its deformation mode depend on the input feature layer. Our research incorporates an additional layer of deformable convolution so that the offset of the features can be learned from the upper feature layer. The green dotted box in Fig. 3 shows the offset (black arrow) of the deformable convolution based on the standard convolution. And the blue dotted box in Fig. 3 indicates the feature extraction process in the upsampling process. In the training process, the feature extraction layer is activated by a single activation unit located in the target area. The offset of features will be determined according to the location and shape of bone marrow edema. In Fig. 3, the feature offset will be more inclined to the upper left and lower right in the initial position of the activation unit. After several

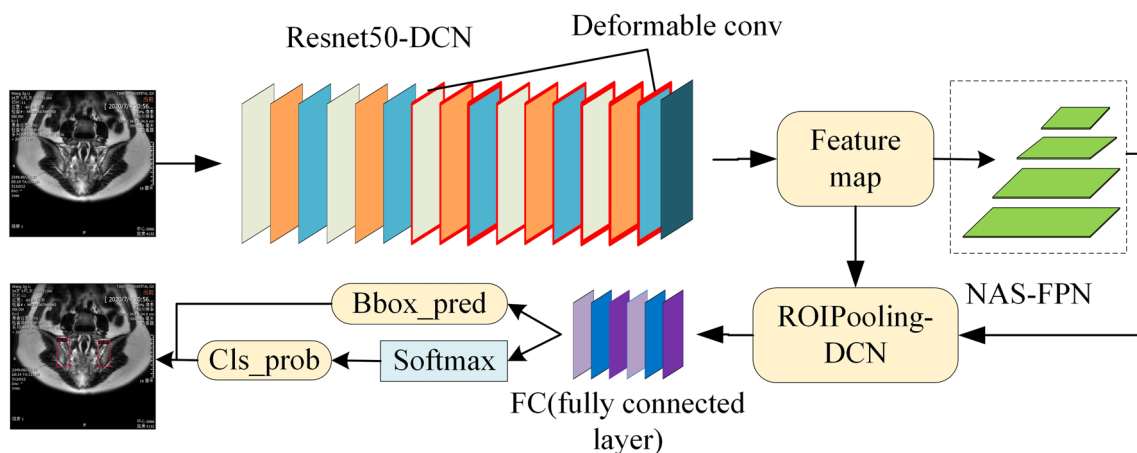


Fig. 2 Structure of the DCNAS-Net

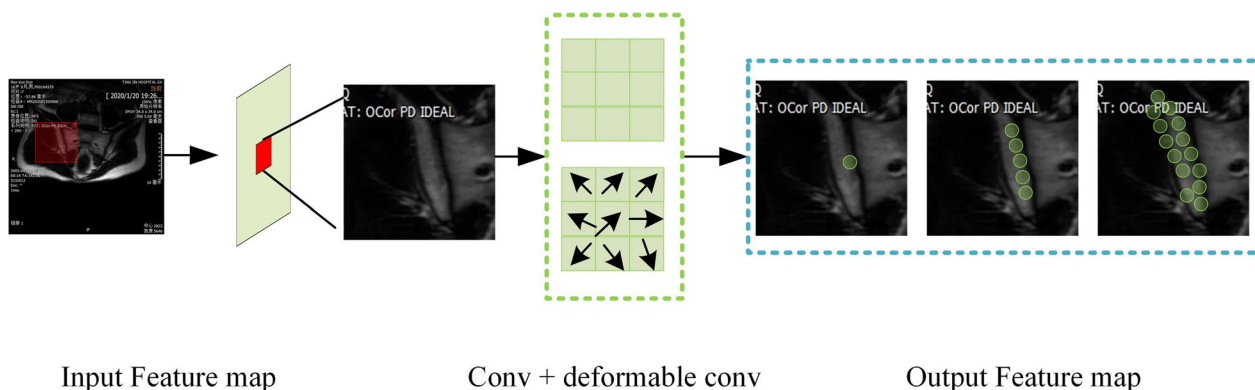


Fig. 3 Mechanism of deformable convolution. The cropped image is realized by python script, and it contains the bone marrow edema area

rounds of training, the output feature layer will be able to better learn the features of the bone marrow edema area.

And then it comes to the deformable RoI pooling layer [19, 20]. This module adds an offset after the RoI pooling layer, and the structure is useful for improving the accuracy of local positioning of targets. Both of these modules are lightweight modules. They can be inserted into a regular network by adding a small number of parameters and computational effort. The construction of the deformable convolution module and the deformable RoI pooling layer are described below. A conventional convolution is built in two steps:

- Sampling the input feature mapping x using a regular grid R ;
- Weighting and summing the sampled values.

Defining R as the size of the perceptual field, and taking a 3×3 convolution as an example, with a convolution hole size of 1, we have

$$R = \{(-1, -1), (-1, 0), \dots, (0, 1), (1, 1)\} \quad (1)$$

For p_0 , each position of the input feature map y , we have

$$y(p_0) = \sum_{p_0 \in R} w(p_0) \times x(p_0 + p_n) \quad (2)$$

Where p_n enumerates all the positions in R . In the deformable convolution, an offset Δp_n ($|n = 1, 2, \dots, N$) is added to the regular sampling grid R , and Eq. (2) becomes

$$y(p_0) = \sum_{p_0 \in R} w(p_0) \times x(p_0 + p_n + \Delta p_n) \quad (3)$$

The sampling grid samples at offset position $p_n + \Delta p_n$. In most cases, the offset Δp_n is generally not an integer.

Then Eq. (3) can be implemented by bilinear difference as follows

$$x(p) = \sum_q G(q, p) \times x(q) \quad (4)$$

Where p denotes any position in R , q is an enumeration of all integer positions in x . $G(q, p)$ is a 2D bilinear interpolation kernel. It can be decomposed as the product of two 1D kernels as

$$G(q, p) = g(q_x, p_x) \times g(q_y, p_y) \quad (5)$$

Where $g(a, b) = \max(0, 1 - |a - b|)$, it follows that $G(q, p)$ is not equal to 0 for only a few q , so it does not consume many computational resources.

Deformable RoI pooling layer

The role of the RoI pooling layer [21] is to unify the feature matrices corresponding to RoI of different sizes into a fixed size bin for output. For a conventional RoI pooling layer, assuming that the input feature matrix is x , its size is $w \times h$, p_0 is its coordinate, and $y(i, j)$ denotes the bin at (i, j) , we have

$$y(i, j) = \sum_{p \in \text{bin}(i, j)} x(p_0 + p) / n_{ij} \quad (6)$$

Where n_{ij} denotes the number of pixels in the bin at position (i, j) . For the deformable pooling layer, we have

$$y(i, j) = \sum_{p \in \text{bin}(i, j)} x(p_0 + p + \Delta p_{ij}) / n_{ij} \quad (7)$$

Typically, Δp is also not an integer, so Eq. (7) can still be implemented with the bilinear difference of Eq. (4). In fact, the offsets in the deformable convolution and deformable

RoI pooling layers can also be considered as parts of the network. It is obtained by adding an offset to the standard convolution, which in turn can also be learned end-to-end by back-passing of gradients. The size and position of the convolution kernels can be dynamically adjusted to the current target shape after adding the offset. The intuitive effect is that the sampling points of the convolutional kernels at different locations will change adaptively according to the shape of the bone marrow oedema, which can improve the detection accuracy.

NAS based feature aggregation module

MRI images of the lumbar region are generally dominated by a grey background. Bone marrow oedema and some organ tissues, especially fatty tissues, are bright colors, which can make it difficult to identify bone marrow oedema. Even experienced radiologists take a few minutes to distinguish edema from other organs or tissues. For conventional neural networks, as the convolution layer deepens, the features extracted by the shallow convolution take up very little weight in the subsequent feature layers. By the time the convolution reaches the final layer, it is no longer possible to read the semantic information in the shallow background. That makes it difficult for conventional networks to detect small targets. Several studies in recent years have confirmed that the introduction of new information paths in the FPN structure can build more effective detection heads. Among them, PAFPN introduces bottom-up information paths on top of FPN; Libra R-CNN uses Non-Local path modules and builds multi-level FPN structures with U-shaped structures to construct more effective detection modules; SPEC stacks four scale-equilibrium modules behind a classical FPN to enhance multi-scale correlation. Meanwhile, research on NAS being used to automatically search for detectors in FPN architectures is also underway. NAS-FCOS and Spine-Net [22] have achieved excellent results after incorporating NAS. These prove that NAS architectures are effective. Inspired by the above studies our algorithm introduces four unique information paths and two non-reference paths to construct a multi-level FPN structure. The optimal FPN combination module is automatically searched by the latest efficient NAS module. This effectively improves the accuracy of bone marrow oedema detection.

Methods for constructing 6 information paths

- Top-down information path The top-down information path is similar to the classical FPN. Its structure is shown in Fig. 4a. We denote the output of the feature pyramid as $F_2^t, F_3^t, F_4^t, F_5^t$, and then construct

them in a top-down sequence. Each feature mapping F_i^t is constructed by summing the input pyramid feature mapping P_i at the same level and the output features F_{i+1}^t at a higher level, so that we have

$$F_i^t = W_i^t \otimes (U(F_{i+1}^t) + P_i) \tag{8}$$

Where $U(\cdot)$ denotes the upsampling function and $W_i^t (i = 2, 3, 4, 5)$ are 3×3 convolution module. However, note that for the higher-level features ($i = 3, 4, 5$), W_i^t is a conformable convolution module, which can be better adapted to different pyramidal accesses, while W_2^t is a regular convolution module.

- Bottom-up information path The structure of the bottom-up information path is shown as Fig. 4b, which is constructed in a similar way to the previous path. The outputs of the feature pyramid ($F_2^b, F_3^b, F_4^b, F_5^b$) are constructed sequentially in a bottom-up manner. We have

$$F_i^b = W_i^b \otimes (D(F_{i-1}^b) + P_i) \tag{9}$$

Where $D(\cdot)$ denotes the downsampling function. The convolution module of W_i^b is the same as the top-down information path.

- Scale-balanced information path The scale-balanced information path is driven by SEPC, which superimposes a pyramidal convolution after the classical FPN structure to obtain the correlation among scales. Its structure is shown as Fig. 4c, where each feature matrix F_i^s is obtained by superimposing the input feature matrices that are expected to be adjacent, so that we have

$$F_i^s = U(W_1^s \otimes P_{i+1}) + W_0^s \otimes P_i + W_{-1}^s \otimes P_{i-1} \tag{10}$$

Where $D(\cdot)$ denotes the downsampling function. The convolution module of W_i^b is the same as the top-down information path.

- Two-step fusion split information path The two-step fusion split information path is shown as Fig. 4d. Firstly, the higher and lower levels of the feature pyramid information are each combined into α_s and α_t . We have

$$\alpha_s = P_4 + U(P_5), \quad \alpha_t = D(P_2) + P_3 \tag{11}$$

After obtaining the combined features and fusing them by overlay stitching, we have

$$\beta_s = W_s^f \otimes CON(\alpha_s, D(\alpha_t)), \quad \beta_t = W_t^f \otimes CON(U(\alpha_s), \alpha_t) \tag{12}$$

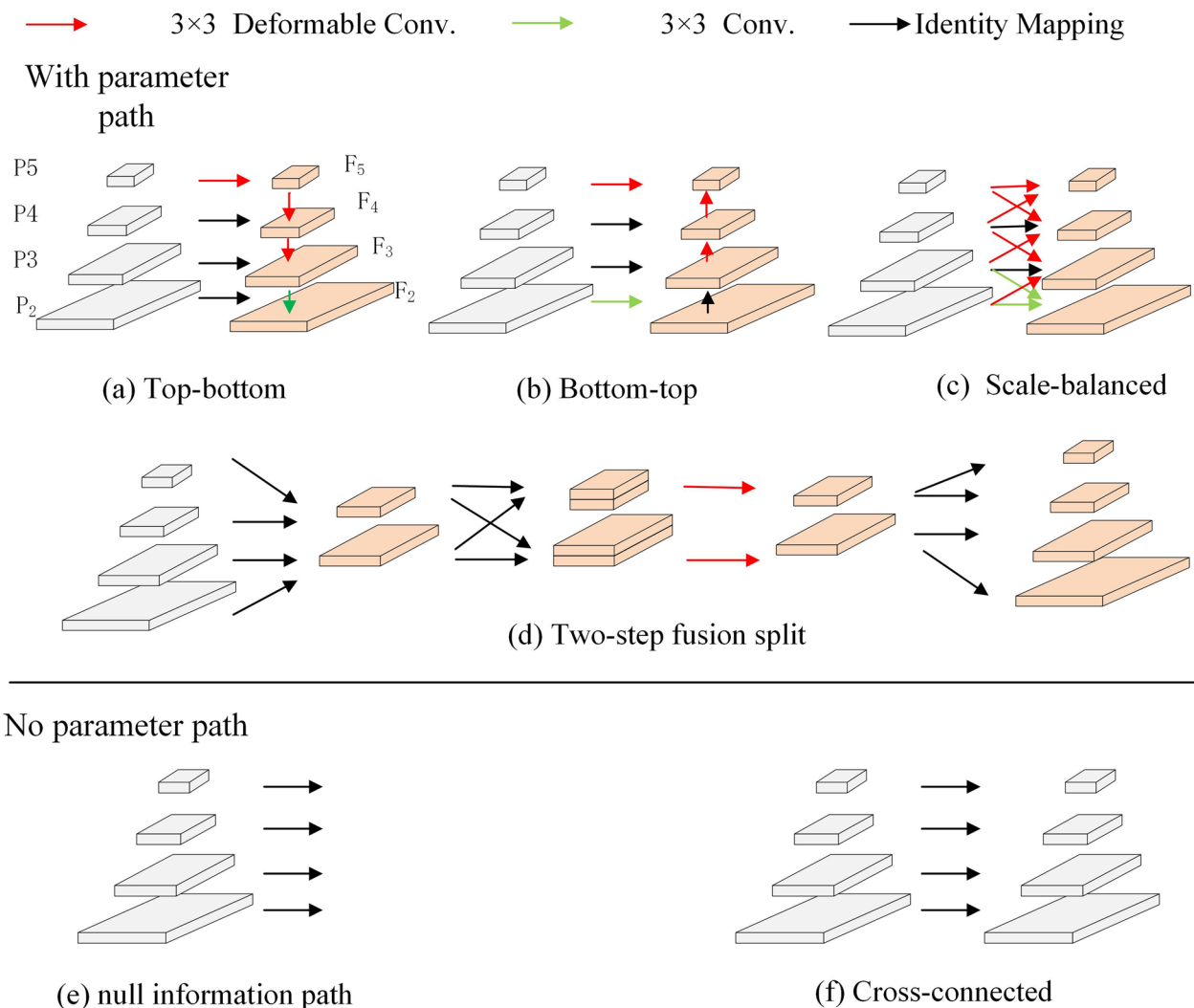


Fig. 4 4 message paths with parameters and 2 message paths without parameters. **a** Top-bottom information path; **b** Bottom-top information path; **c** Scale-balanced information path; **d** Two-step fusion split information path; **e** Null information path; **f** Cross-connected information path

Where W_s^f and W_l^f is 3×3 conformable convolution module, and $CON(\cdot)$ denotes the stitching along the channel dimension. After these operations, the feature map β_s, β_l carries the information from the fusion of features at all levels. Finally adapting it to a multiscale pyramidal feature map, we have

$$F_2^f = U(\beta_l), F_3^f = \beta_l, F_4^f = \beta_s, F_5^f = D(\beta_s) \tag{13}$$

- Null information paths and cross-connected information paths Another characteristic of bone marrow oedema is its relatively concentrated location, so it is not necessary to use superimposed information paths in its detection to extract too much location information. Our algorithm adds 2 unparallelled information paths to reduce the complex-

ity of the model. They are shown as Fig. 4e and f. The cross-connected path is used to perform direct mapping, while the null information path is used to eliminate redundant information paths. These two non-parametric information paths allow for a better balance between accuracy and efficiency.

Construction of the NAS-FPN model

In some recent studies, NAS has been applied to automatically search for FPN structures, the most common of which is the Single Path One-Shot [23] structure, as shown in Fig. 5a. However, there is only one information path between its nodes to connect and it has only sequential connections. From the information

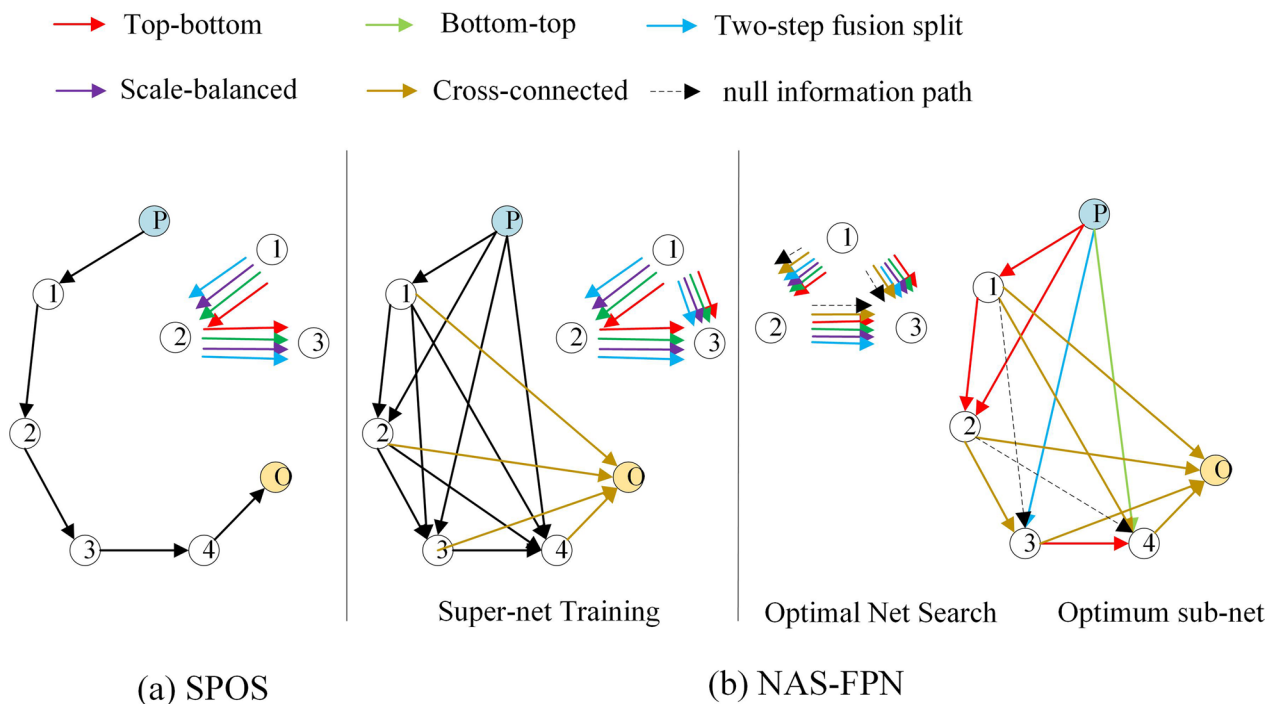


Fig. 5 a FPN super-net structure for SPOS b NAS-FPN search space super-net results

extraction point of view, an aggregated NAS can make more efficient use of the extracted information features.

The NAS-FPN is a fully connected DAG (directed acyclic graph). Each of its nodes represents an FPN with six different information paths connected between every two nodes. Where node P represents the feature matrix extracted from the backbone network and node O represents the final output feature matrix. In network training, our algorithm first constructs a super network containing all information paths, and then searches for the optimal sub-network through evolutionary algorithm. The optimal sub-network represents the optimal combination of FPN that aggregate multiple information paths. The optimal sub-network is searched by reasoning with previously trained weights rather than retraining the weights, so it is more efficient. In the experiments, we find that the model achieves the best results when $N = 5$. When $N > 5$, NAS-FPN can not increase the accuracy of the detection, but significantly decreases the speed of the detection. Therefore $N = 5$ is used in the subsequent comparison experiments unless otherwise specified. In order to reduce the complexity of the network, we reduce the output channels of the FPN from 256 to 112. The corresponding channels of the network detection head are also reduced from 256 to 112. It is found experimentally that these channel reductions have less impact on the network accuracy, but can significantly reduce the weight size of

the network (see the ablation data in the next section for details).

Optimization of super-net and loss functions of networks

The optimal subnetworks in Fig. 5b are searched by an evolutionary algorithm after completing the training of the super-net. That is, S subnets are first randomly selected from the super-net and the performance of each subnet is ranked by inferring the data from the validation set. Then the best k subnets start the “crossover” and “mutation” operations [24], and the new subnets are generated and the best subnet are selected after repeating the previous steps n times. The “crossover” refers to the crossover of two randomly selected subnetworks to generate a new subnet, and the “mutation” refers to the mutation of a random subnet boundary with probability p to generate a new subnet. In our algorithm, $S = 50$, $k = 10$ and $p = 0.1$ are set. To help the convergence of the model, this algorithm adds an L_1 regularization function to the weights to balance the boundary losses. The total loss function is given by

$$L = L_{bbox} + \mu L_1 = \sum (L_{cls} + L_{loc}) + \mu L_1 \quad (14)$$

Where L_{cls} is the corresponding loss for lesion classification, which is the cross-entropy loss function; L_{loc} is the corresponding positioning loss function, which is the Smooth L_1 loss function for L_{loc} ; μ is the hyperparameter

of the regularisation function L_1 , and the network was found to work best when μ was taken as 0.11 in the experiments.

Results and discussion

Experimental data and experimental parameters

No publicly dedicated dataset is available for detecting bone marrow oedema in lumbar MRI images. The data for the research is obtained from the relevant department of Tianjin Hospital, and the lumbar MRI image dataset was obtained from 72 anonymous patients between 2019 and 2021. A total of 521 MRI images are screened, among which 416 images are placed in the training set and 105 images are placed in the testing set. The dataset was annotated according to the MS COCO [25] format, which was completed by three experienced radiologists. Our algorithm was trained using stochastic gradient descent. The RTX8000 GPU is used for accelerated training, and all experiments are carried out on this device. The training parameter batch-size is set as 4, the learning rate is set as 0.001, and a total of 30 training rounds were set. We used random flipping and random cropping as quality enhancement methods. Flip the mirror image of the original image randomly with a probability of 0.5. And clip the image randomly to 0.6–1.0 times of the original ones in size.

Ablation experiments

Ablation experiments are conducted to understand the contribution of each module in this paper’s algorithm. The ablation experiments use AP_{50} of the COCO evaluation criteria as the evaluation metric. Besides, there are experimental data of recall and F1-measure as auxiliary verification data. Where AP_{50} and recall are the average of five cross-validation experiments. The training set and verification set of each cross-validation are not repeated. The subjects of the ablation experiments include the number of super net nodes N , the deformable convolutional and deformable RoI layers and the NAS-FPN module.

The number of super net nodes N represents the number of levels of the FPN structure. On the one hand, its size directly affects the performance of the network. On the other hand, a larger N represents more network parameters, which affects the training speed and the weights. The experimental results are shown in Table 1. After adding the NAS-FPN module, the value of AP_{50} increases 3.4% at most from 84.9%. The recall of all experiments remained above 91.2%. Meanwhile, F1-Measure has a maximum value of 91.0 when $N = 5$. Also, the weight reduces 11.37 MB–56.88 MB. It proves the effectiveness of the NAS-FPN module. However, when $N \geq 5$,

Table 1 Effect of super net node N on precision and weight

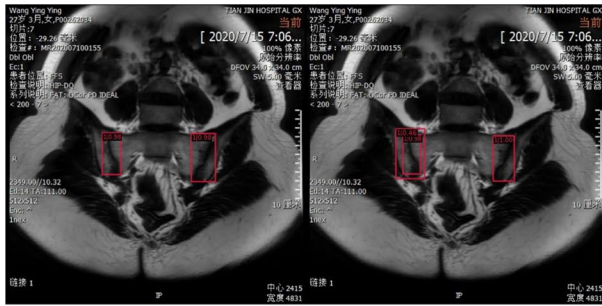
Algorithms	Number of nodes N	$AP_{50}/\%$	Recall/%	F1/%	Weight size/MB	Single image detection time/ms
Faster-RCNN +FPN	–	84.9	91.7	88.2	322.54	161
Ours	2	83.7	91.2	87.3	265.66	135
(Faster-RCNN + NAS-FPN)	3	85.3	91.7	88.4	266.69	140
	4	87.3	91.2	89.2	270.65	140
	5	88.3	91.2	91.0	277.18	147
	6	88.0	91.3	90.8	304.19	158
	7	86.7	93.6	90.4	311.17	166

Table 2 Enhancement of effects by deformable and NAS-FPN modules

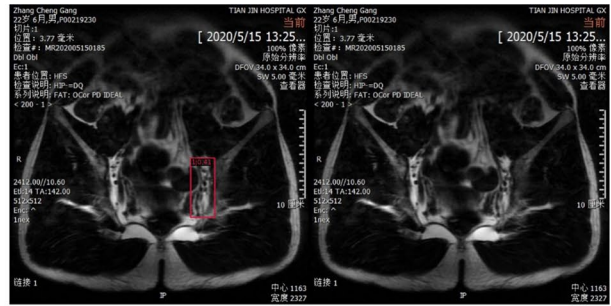
Algorithms	Deformable Conv	NAS-FPN	$AP_{50}/\%$	Recall/%	F1/%	Weight size/MB	Single image detection time/ms
Faster-RCNN +FPN	–	–	84.9	91.7	88.2	322.54	161
Ours(Faster-RCNN + NAS-FPN)	✓	–	85.7	90.4	88.0	327.10	140
	–	✓	88.3	91.2	91.0	277.18	147
	✓	✓	90.6	95.1	92.8	275.21	144

the AP_{50} and the F1-Measure are not increasing, but the weights still increase rapidly and the detection rate decreases. To balance precision and efficiency, subsequent experiments use the value of $N = 5$.

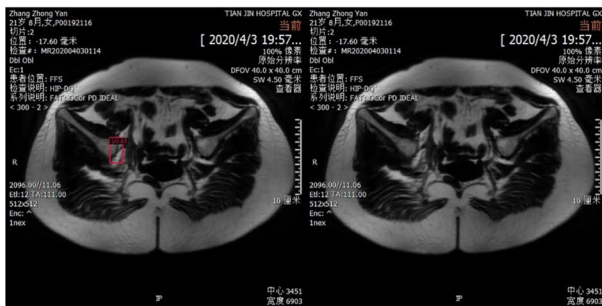
Deformable convolution modules and NAS-FPN modules
 In order to validate the improvement of the DCNAS structure on the detection of bone marrow oedema, we compare the performance of our algorithm with the



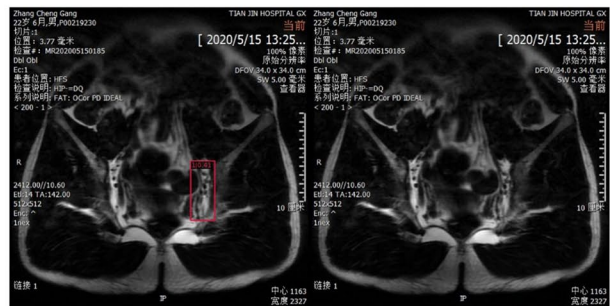
Carafe



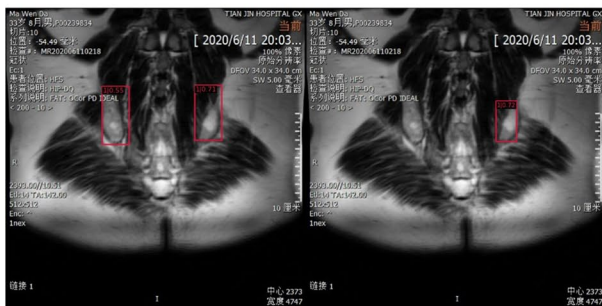
Faster-RCNN-Resnet101



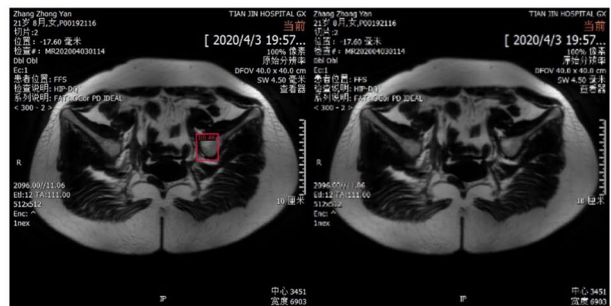
Hernet



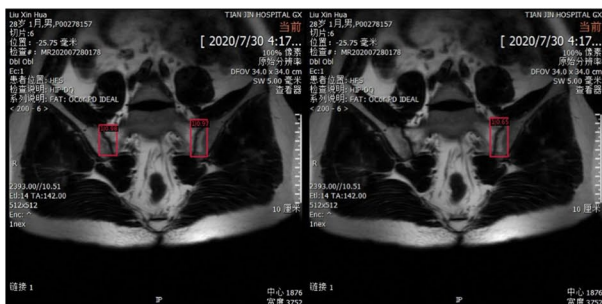
PAFPN



Retina-Net



SSD



Yolo V3



Yolo X

Fig. 6 Comparison with other algorithms. The left side of each graph shows the results of our algorithm and the right side shows the results of other algorithms

Table 3 Comparison of current working algorithm with existing networks

Algorithms	AP ₅₀ /%	Recall/%	F1/%	Weight size/MB	Single image detection time/ms
Carafe	83.3	88.4	85.8	366.34	144
Faster-RCNN-Resnet101	85.9	91.2	88.4	471.24	147
Hernet	88.0	93.4	85.8	367.41	178
PAFPN	83.8	88.9	86.3	350.21	140
Retina-Net	88.5	95.4	91.8	283.33	136
SSD	89.4	94.4	91.8	185.76	102
Yolo V3	87.0	90.9	88.9	481.40	109
Yolo X	86.9	93.4	90.1	105.30	109
Ours	90.6	95.1	92.8	275.21	144

Faster-RCNN+FPN. As shown in Table 2, when the deformable convolution module is added, AP₅₀ improves from 84.9 to 85.7%, and the detection time is reduced to 140ms per image. When the NAS-FPN module is added, AP₅₀ improves to 88.3%. When the DCNAS module is added, AP₅₀ improves to 90.6% and the detection time becomes 144ms per image, while the parameter weight is reduced from 322.54 MB to 275.21 MB compared to the unaltered network. When the deformable convolution and NAS-FPN are added, recall is also increased to 95.1%, and F1-Measure is increased to 92.8%. The experimental results show that the deformable module and NAS-FPN are effective.

Algorithm comparison and analysis

In order to verify the effectiveness of the algorithm in this paper, the research performs comparative experiments with existing algorithmic models for target examination. Including classical networks such as Faster-RCNN-ResNet101, Retina-Net [26], SSD [27], Yolo V3 [28] and Yolo X [29] and some improved detection networks such as Carafe [30], Hernet [31], and PAFPN. The results are shown in Table 3 and Fig. 6.

As can be seen from Table 3, our algorithm has the best accuracy and F1-Measure for bone marrow oedema detection. and our algorithm also has good detection recall and speed. And our network weight is only 275.21 MB, which means it can be more easily deployed in the detection equipment of clinical medicine. Of course, it might be possible to use other lightweight network backbones with less weight. As can be seen from Fig. 6, our algorithm is able to detect bone marrow oedema in different styles of MRI images with few missed and false detections.

At present, the research on the detection of bone marrow edema in MRI images mainly focuses on CNN

classification model. For example, Klontzas et al. [8] designed a CNN ensemble to detect bone marrow edema in bone joints. The accuracy of this ensemble for classification tasks is 82.18–93.68%. Lee et al. [9] designed an optimization model based on ResNet18, which is used to detect bone marrow edema in sacroiliac joints. But this network also belongs to the category of image classification, and the detection accuracy of their model in the experiment is 93.55%. Unlike their research, our work belongs to target detection. Our optimized neural network can not only classify abnormal MRI images accurately, but also give the location information of bone marrow edema as accurately as possible. This can better provide clinicians with reference information for auxiliary diagnosis. By now, there is almost no target detection model for bone marrow edema in MRI images. This is where our research leads. Compared with the classical target detection models, our model has achieved excellent results.

Conclusion

Based on the Faster-RCNN, we redesigned part of the backbone network and the feature pyramid module. Extensive experiments have demonstrated that our algorithm can accurately detect bone marrow oedema in MRI images. At the same time, our algorithm balances accuracy and detection efficiency, and can also be relatively easy to deploy on hospital detection equipment. Therefore, our work has practical clinical implications and can be used as a diagnostic aid for doctors. In principle, our work can be used not only for the detection of bone marrow oedema, but also for the detection of other pathologies. In particular, the application of MRI images as a background for detection has good prospects. We will continue to design new dedicated networks and carry out new experiments in collaborating laboratories.

Abbreviations

- MRI: Magnetic resonance imaging
- CAD: Computer aided detection
- AUC: Area under curve
- RoI: Region of interest
- FPN: Feature pyramid networks
- NAS: Neural architecture search
- DAG: Directed acyclic graph

Acknowledgements

We are grateful for the financial support of the Tianjin Health Science and Technology Project Committee.

Author contributions

CS was responsible for designing algorithm and writing manuscript. SZ and WZ involved in designing experimental method. ZW, WL, ZS, PZ and ST participated in labeling data and analyzing experimental data. YL made

contributions to pictures making and manuscripts checking. All the authors reviewed and approved the manuscript.

Funding

This study was supported by the fund of "Tianjin Health Science and Technology Project Committee", Certificate No. KJ20091. The funding body played no role in the design of the study and collection, analysis, interpretation of data, and in writing the manuscript.

Availability of data and materials

The experimental data for this work has been presented in the paper. The dataset and algorithms have been deposited with the corresponding author; please contact the corresponding author for details if you wish to use them.

Declarations

Ethics approval and consent to participate

This study was approved by the Medical Ethics Committee of Tianjin Hospital, Tianjin, China on 14 September 2020. All participants provided written informed consent to participate in the study. All the methods used in this study are in line with relevant guidelines and regulations.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Received: 10 September 2022 Accepted: 21 March 2023

Published online: 28 March 2023

References

- Walker BF. The prevalence of low back pain: a systematic review of the literature from 1966 to 1998. *Clin Spine Surg.* 2000;13(3):205.
- Murray C, Lopez AD. Measuring the global burden of disease. *N Engl J Med.* 2013;369(5):448–57.
- Gebrewold Y, Tesfaye B. Does lumbar mri predict degree of disability in patients with degenerative disc disease? a prospective cross-sectional study at university of gondar comprehensive specialized hospital, north west ethiopia, 2020. *BMC Med Imaging.* 2022;22(1):1–8.
- Watanabe H, Okada M, Kaji Y, Satouchi M, Arai Y. New response evaluation criteria in solid tumours - revised recist guideline (version 1.1). *Gan to kagaku ryoho. Cancer Chemother.* 2009;36(13):2495–501.
- Liao F, Liang M, Li Z, et al. Evaluate the Malignancy of Pulmonary Nodules Using the 3D Deep Leaky Noisy-or Network. *IEEE Transactions on Neural Networks & Learning Systems*, 2017.
- Ribli, Dezső, Horváth, Anna, Unger Z, et al. Detecting and classifying lesions in mammograms with Deep Learning. *entific Reports*, 2018;8(1).
- Huang T, Yang R, Shen L, Feng A, Li L, He N, Li S, Huang L, Lyu J. Deep transfer learning to quantify pleural effusion severity in chest x-rays. *BMC Med Imaging.* 2022;22(1):1–11.
- Klontzas ME, Vassalou EE, Kakkos GA, Spanakis K, Zibis A, Marias K, Karantanas AH. Differentiation between subchondral insufficiency fractures and advanced osteoarthritis of the knee using transfer learning and an ensemble of convolutional neural networks. *Injury.* 2022;53(6):2035–40. <https://doi.org/10.1016/j.injury.2022.03.008>.
- Lee KH, Choi ST, Lee GY, Ha YJ, Choi SI. Method for diagnosing the bone marrow edema of sacroiliac joint in patients with axial spondyloarthritis using magnetic resonance image analysis based on deep learning. *Diagnosics.* 2021;11:1156.
- Dai J, Qi H, Xiong Y, Li Y, Zhang G, Hu H, Wei Y. Deformable convolutional networks. 2017 IEEE International Conference on Computer Vision (ICCV), Venice, Italy, 2017, pp. 764–773. <https://doi.org/10.1109/ICCV.2017.89>.
- Lin TY, Dollar P, Girshick R, He K, Hariharan B, Belongie S. Feature pyramid networks for object detection. In: 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 2017.
- Liu S, Qi L, Qin H, Shi J, Jia J. Path aggregation network for instance segmentation. In: 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). 2018.
- Pang J, Chen K, Shi J, Feng H, Ouyang W, Lin D. Libra r-cnn: Towards balanced learning for object detection. In: 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). 2020.
- Wang X, Zhang S, Yu Z, Feng L, Zhang W. Scale-equalizing pyramid convolution for object detection. In: 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). 2020.
- Tan M, Pang R, Le QV. Efficientdet: Scalable and efficient object detection. In: 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). 2020.
- Liang T, Wang Y, Tang Z, Hu G, Ling H. Opanas: One-shot path aggregation network architecture search for object detection. *arXiv e-prints*, <https://arxiv.org/abs/2103.04507>, 2021.
- Ghiasi G, Lin TY, Le QV. Nas-fpn: Learning scalable feature pyramid architecture for object detection. In: 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). 2019.
- Wang N, Gao Y, Chen H, Wang P, Zhang Y. Nas-fcos: Fast neural architecture search for object detection. In: 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). 2020.
- Girshick R. Fast r-cnn. 2015.
- Cordts M, Omran M, Ramos S, Rehfeld T, Schiele B. The cityscapes dataset for semantic urban scene understanding. In: 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 2016.
- Girshick R, Donahue J, Darrell T, Malik J. Rich feature hierarchies for accurate object detection and semantic segmentation. In: IEEE Computer Society. 2013.
- Du X, Lin TY, Jin P, Ghiasi G, Tan M, Cui Y, Le QV, Song X. Spinenet: Learning scale-permuted backbone for recognition and localization. 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Seattle, WA, USA, 2020, pp. 11589–11598. <https://doi.org/10.1109/CVPR42600.2020.01161>.
- Guo Z, Zhang X, Mu H, Wen H, Sun J. Single path one-shot neural architecture search with uniform sampling. 2019.
- Kendall M. A new measure of rank correlation. *Biometrika.* 1938;30(1–2):81–93.
- Library WP. European conference on computer vision.
- Lin TY, Goyal P, Girshick R, He K, Dollár P. Focal loss for dense object detection. *IEEE Trans Pattern Anal Mach Intell.* 2017;99:2999–3007.
- Berg A C, Fu C Y, Szegedy C, et al. SSD: Single Shot MultiBox Detector. *arXiv e-prints*, <https://arxiv.org/abs/1512.02325v5>, 2015.
- Redmon J, Farhadi A. Yolov3: An incremental improvement. *arXiv e-prints.* 2018.
- Ge Z, Liu S, Wang F, Li Z, Sun J. YoloX: Exceeding yolo series in 2021. 2021.
- Wang J, Chen K, Xu R, Liu Z, Chen CL, Lin D. Carafe: Content-aware reassembly of features. In: 2019 IEEE/CVF International Conference on Computer Vision (ICCV). 2020.
- Sun K, Xiao B, Liu D, Wang J. Deep high-resolution representation learning for human pose estimation. *arXiv e-prints.* 2019.

Publisher's Note

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