Article

An Innovative Deep Learning Approach to Spinal Fracture Detection in CT Images

Ann. Ital. Chir., 2024 95, 4: 657–668
https://doi.org/10.62713/aic.3498

Haiting Wu[1], Qingsong Fu[2]

[1]Department of Spinal Surgery, Ningbo No.2 Hospital, 315010 Ningbo, Zhejiang, China
[2]Department of Orthopaedics, Ningbo No.2 Hospital, 315010 Ningbo, Zhejiang, China

AIM: Spinal fractures, particularly vertebral compression fractures, pose a significant challenge in medical imaging due to their small-scale nature and blurred boundaries in Computed Tomography (CT) scans. However, advanced deep learning models, such as the integration of the You Only Look Once (YOLO) V7 model with Efficient Layer Aggregation Networks (ELAN) and Max-Pooling Convolution (MPConv) architectures, can substantially reduce the loss of small-scale information during computational processing, thus improving detection accuracy. The purpose of this study is to develop an innovative deep learning approach for detecting spinal fractures, particularly vertebral compression fractures, in CT images.

METHODS: We proposed a novel method to precisely identify spinal injury using the YOLO V7 model as a classifier. This model was enhanced by integrating ELAN and MPConv architectures, which were influenced by the Receptive Field Learning and Aggregation (RFLA) small object recognition framework. Standard normalization techniques were utilized to preprocess the CT images. The YOLO V7 model, integrated with ELAN and MPConv architectures, was trained using a dataset containing annotated spinal fractures. Additionally, to mitigate boundary ambiguities in compressive fractures, a Theoretical Receptive Field (TRF) based on Gaussian distribution and an Effective Receptive Field (ERF) were used to capture multi-scale features better. Furthermore, the Wasserstein distance was employed to optimize the model's learning process. A total of 240 CT images from patients diagnosed with spinal fractures were included in this study, sourced from Ningbo No.2 Hospital, ensuring a robust dataset for training the deep learning model.

RESULTS: Our method demonstrated superior performance over conventional object detection networks like YOLO V7 and YOLO V3. Specifically, with a dataset of 200 pathological images and 40 normal spinal images, our method achieved a 3% increase in accuracy compared to YOLO V7.

CONCLUSIONS: The proposed method offers an innovative and more effective approach for identifying vertebral compression fractures in CT scans. These promising findings suggest the method's potential for practical clinical applications, highlighting the significance of deep learning in enhancing patient care and treatment in medical imaging. Future research should incorporate cross-validation and independent validation and test sets to assess the model's robustness and generalizability. Additionally, exploring other deep learning models and methods could further enhance detection accuracy and reliability, contributing to the development of more effective diagnostic tools in medical imaging.

Keywords: deep learning; spinal fracture; CT images; accuracy; YOLO V7; agency; TRF

Introduction

The increasing ratio of traumatic diseases in our modern society is becoming more apparent. Severe neurologic deficits resulting from injuries, such as intramedullary hematoma and spinal cord contusion with associated edema, can be discerned through various imaging modalities. Computed Tomography (CT) is particularly effective in detecting spinal fractures and bony abnormalities due to its high-resolution imaging capabilities. CT can accurately capture physiological and morphological changes, including swelling and asymmetry, by producing detailed cross-sectional images of bony structures. Moreover, CT is highly effective in confirming damages incurred from traumatic spinal injuries, making it crucial for identifying spinal fractures and assessing their severity. Consequently, accurate diagnosis becomes vital for managing lesions on bony tissue and planning treatment, especially when there exists ambiguity in diagnostic evaluations.

The concept of automatically recognizing injuries on CT images has emerged as a critical research area in medical imaging due to its potential to alleviate workload of healthcare professionals considerably. Object detection in medical imaging involves employing computer algorithms to identify and locate specific objects or structures within medical images, such as radiographs and CT scans. This technology can significantly boost the efficiency and preci-

Correspondence to: Qingsong Fu, Department of Orthopaedics, Ningbo No.2 Hospital, 315010 Ningbo, Zhejiang, China (e-mail: fuqs2021@163.com).

sion of medical diagnoses and treatment planning. A major application of object detection in medical imaging is the detection of abnormalities or lesions within the body. For instance, object detection algorithms can identify tumors in CT or magnetic resonance imaging (MRI) images, assisting in cancer diagnosis [1, 2]. Similarly, these techniques can identify other abnormalities, such as fractures or anomalies in the heart or vascular system [3, 4]. The capability to automatically detect these abnormalities can reduce the reliance on radiologists' manual image interpretation, thereby enhancing the pace and accuracy of diagnoses [5]. Moreover, object detection algorithms can be used in monitoring disease progression and treatment response by tracking alterations in the size and shape of abnormalities over time [6, 7]. Another significant application of object detection is the automatic segmentation of organs and tissues. This process is helpful for activities like assessing organ size and volume, which is crucial for planning treatment and monitoring disease progression [8]. Object detection, for instance, can facilitate the measurement of a tumor's size to decide the appropriate treatment plan [9] or monitor size changes during chemotherapy to assess treatment response [10]. Furthermore, object detection can automatically label various structures in medical images, thereby enhancing the accuracy and efficiency of manual image annotation [11, 12].

One considerable challenge in object detection in medical imaging is the high variability in the appearance and shape of the objects of interest. For instance, tumors can differ significantly in size, shape, and intensity within an image, making it difficult for algorithms to detect them accurately [13]. Likewise, other abnormalities, such as fractures, may present with complex and variable shapes, rendering them challenging to detect and segment [14]. To address this obstacle, researchers have designed algorithms specifically for medical imaging, including deep learning-based methods capable of recognizing patterns within medical images [15, 16]. These methods often incorporate convolutional neural networks (CNNs), which have succeeded in various object detection tasks across different domains [17, 18].

Another challenge in object detection in medical imaging is the limited availability of annotated data for algorithm training. Medical images often contain patient information, making it challenging to retrieve large dataset required for training and evaluating object detection algorithms [19]. To address this problem, researchers have devised methods for synthesizing annotated data, such as utilizing computer simulations or generating synthetic images with deep learning techniques [20, 21]. These methodologies help create extensive datasets of annotated images, which are helpful for training and evaluating object detection algorithms.

Despite various challenges, object detection in medical imaging holds substantial potential to enhance the accuracy of medical diagnosis and the efficiency of treatment planning. Besides the aforementioned applications, object detection can be used in other areas of medical imaging, such as image registration [22, 23], image-guided surgery [24, 25], and radiation therapy planning [26, 27]. Creating more precise and resilient object detection algorithms will be crucial in fully realizing the potential of this technology in medical imaging.

Several directions for future research in object detection for medical imaging are emerging. One crucial area is the development of algorithms explicitly designed for the unique characteristics of medical images. Another promising direction is the integration of object detection with other medical imaging technologies. For example, combining object detection with image registration algorithms could enable accurate alignment of images from different modalities or time points, facilitating more precise measurement and tracking of abnormalities [28, 29]. Furthermore, object detection could improve image-guided surgery systems by providing real-time guidance and localization during procedures [30]. Alongside these technical challenges, regulatory and ethical considerations are crucial for object detection in medical imaging. The accuracy and reliability of these algorithms will need to undergo stringent validation before they can be applied clinically [31]. Furthermore, measures must be taken to ensure that these algorithms do not discriminate against specific patient groups [32].

In conclusion, object detection in medical imaging has the potential to substantially improve the accuracy and efficiency of medical diagnosis and treatment planning. Ongoing research in this field addresses the challenges of variability in object appearance and the limited availability of annotated data while also exploring the integration of object detection with other medical imaging technologies. Creating more accurate and robust object detection algorithms will be pivotal for fully harnessing this technology's potential in medical imaging.

This manuscript primarily focuses on vertebral compression fractures, a prevalent type of spinal fracture, and explores techniques for their automatic recognition. Additionally, we examine various applications of object detection in medical imaging, the current state of the field, future research trajectories, and the challenges that may arise.

Materials and Methods

Conventional object detection models, such as Single Shot MultiBox Detector (SSD), Faster Region-based Convolutional Neural Network (R-CNN), and You Only Look Once (YOLO) V2, have been effectively applied as advanced classifiers in detecting spinal injuries, brain tumors, and retinal lesions. However, compressive, bursting, shearing, and torsional tension injuries in spinal trauma, as represented in CT images, display characteristics such as small targets, blurry features, and indistinct bounding box boundaries. Consequently, these unique features often lead to traditional object detection models demonstrating suboptimal performance in identifying these spinal injuries. Over recent years, there have been considerable progress and mod-

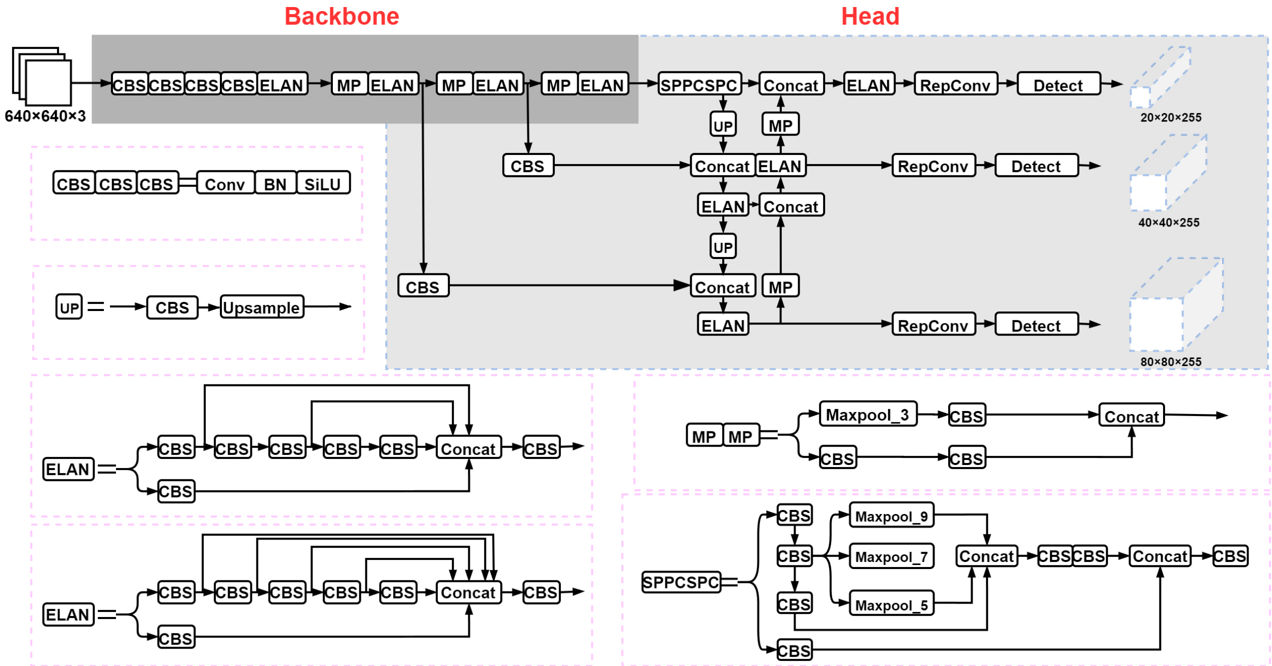

Fig. 1. Architectural overview of YOLO V7 model, the ELAN structure, and MP dimensionality reduction components. (All images in the Materials and Methods section were created using the software "Lark" (V7.22.5, Lark Technologies Pte. Ltd., Beijing, China).) YOLO, You Only Look Once; CBS, Convolution Batch Normalization and SiLU; ELAN, Efficient Layer Aggregation Networks; MP, Max-Pooling; SPPCSPC, Spatial Pyramid Pooling Cross Stage Partial Connections; UP, Upsampling; BN, Batch Normalization; SiLU, Sigmoid Linear Unit.

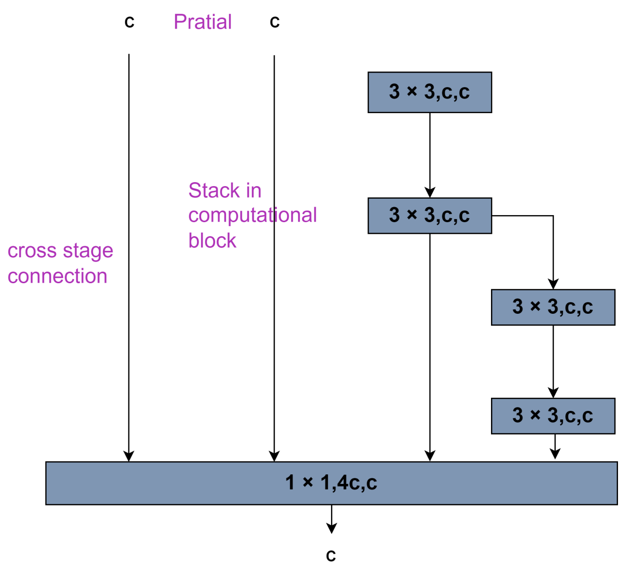

Fig. 2. Detailed illustration of the ELAN structure within the YOLO V7 model.

ifications in the structures of deep learning networks. They have evolved from traditional encoder-decoder designs to highly intricate architectures, giving rise to novel models like Transformer, Attention, and Shuffle. In object detection tasks, advancements like soft labels, auxiliary training, and Distance Intersection over Union (DIOU) have been integrated into these new models. Meanwhile, detecting small objects has become a focus in of deep-learning object recognition. Adjustments in the network, bounding box design, and similarity calculation have been optimized to enhance the recognition performance for small objects. Consequently, this study proposes a targeted method for identifying spinal fractures based on the latest YOLO V7 [33] and the Receptive Field Learning and Aggregation (RFLA) framework for small object recognition. Additionally, we compared our model with two conventional object recognition networks, YOLO V7 and YOLO V3 [34].

This study proposes a method for detecting spinal injuries by leveraging the YOLO V7 network architecture. To address the specific characteristics of spinal injuries in images, such as small targets, blurred edges, and the complexity of accurately determining the bounding box boundaries, this method integrates the Gaussian Receptive Field-based Label Assignment from the RFLA framework for small target detection. It uses the Wasserstein distance [35] to tweak the model structure. These adjustments help optimize the model's extraction performance.

A total of 240 CT images from patients diagnosed with spinal fractures were included in this study. This study was approved by the Ethics Committee of Ningbo No.2 Hospital (No. PJ-NBEY-KJ-2022-08-15). This study utilized medical images and data strictly following the highest standards of ethical conduct and patient confidentiality. We obtained informed consent from all study subjects. Additionally, the purpose and nature of the study design along with the poten-

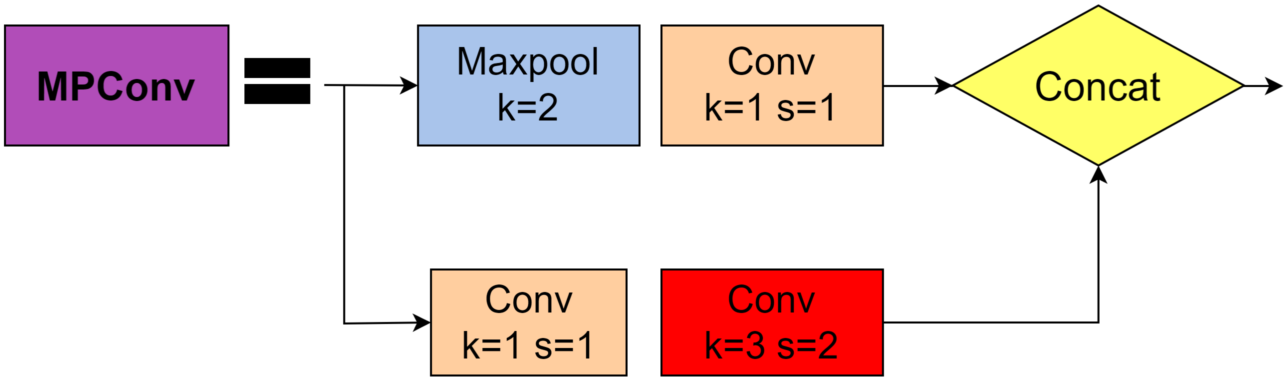

Fig. 3. Schematic presentation of the Max-Pooling Convolution (MPConv) structure within the YOLO V7 model.

tial risks and benefits associated with this study were thoroughly explained to the participants before obtaining their consent.

Network Structure

We utilized YOLO V7, a state-of-the-art object detection model recognized for its proficiency in detecting and classifying objects within images and videos. This model uses a grid-based approach, segmenting an image into cells. Each cell is responsible for determining whether it encompasses an object and, if so, classifying it. Unlike other object detection models that engage multiple CNNs, YOLO V7 relies on a singular CNN to make predictions. This approach enhances its speed and efficiency compared to other models, although it might sacrifice accuracy. The structure of YOLO V7 is summarized in Fig. 1, with its primary improvement being the Efficient Layer Aggregation Networks (ELAN) [36] structure and the Max-Pooling (MP) dimensionality reduction components.

Typically, networks for most deep learning tasks are designed for smaller samples with a 224 or 256 pixels width. However, object detection tasks often deal with larger images and targets, requiring data from a broader range of scales. Recently, the YOLO family has been tailored for larger samples with a width exceeding 400 pixels. For instance, the sample dataset for YOLO V7 includes images with widths of 640 and 1280 pixels. Given that the dataset in this study has dimensions of 1170 pixels $\times$ 2088 pixels, YOLO V7 is a suitable choice for this dataset.

Moreover, conventional convolutional networks are often designed for square input tensors, but in realistic scenarios (such as in this study), images frequently possess unique aspect ratios. To address this issue, the latest models, like YOLO V5 and YOLO V7, propose a scaling adaptive method. By allowing the original image to adapt by adding the minimum number of black borders, the inference speed experienced an enhancement of 37%.

The ELAN structure, a distinctive design feature of the YOLO V7 object detection model, is illustrated in Fig. 2. This structure balances efficiency and accuracy by employing a combination of convolutional layers, residual blocks, and skip connections to construct a deep, potent network capable of accurately detecting and classifying objects in images and videos. A prominent characteristic of the ELAN structure is its use of residual blocks. These blocks enable the network to learn intricate features by adding the output of one layer to the input of the subsequent layer rather than replacing it. This functionality allows the network to augment its existing knowledge, incrementally enhancing its performance. Another essential feature of the ELAN structure is the incorporation of skip connections, enabling the network to bypass specific layers and directly extract information from earlier layers. This helps the network better comprehend the context of an image, leading to more accurate predictions.

MP Dimensionality Reduction Components

Max-Pooling Convolution (MPConv) [37], a specific type of convolutional layer utilized in the YOLO V7 object detection model, is engineered for efficiency and efficacy, making it suitable for real-time applications like video surveillance and autonomous vehicles. Convolutional layers are a critical component of CNNs and are commonly employed for tasks like image classification and object detection. They apply a set of filters to an input image, subsequently detecting patterns and features within that image. The MPConv structure presents a variation on conventional convolutional layers, explicitly designed for enhanced efficiency and effectiveness in object detection applications. This is achieved through a combination of pointwise (or 1 $\times$ 1) convolutions and depthwise convolutions. Pointwise convolutions operate on individual pixels, while depthwise convolutions process groups of pixels. By integrating these two types of convolutions, the MPConv structure effectively detects patterns and features in images, maintaining greater efficiency and lighter computational weight compared to traditional convolutional layers. The structure of MPConv is depicted in Fig. 3.

Loss Function

The YOLO V7 loss function consists of two main components: classification loss and localization loss. The classification loss assesses the model's accuracy in classifying objects in an image, calculated by comparing the model's class predictions for each object to the ground truth annotations. The localization loss evaluates the model's precision in accurately locating objects in an image, calculated by comparing the model's predicted bounding boxes to the ground truth annotations. The overall loss for the YOLO V7 model is calculated as a weighted sum of these two losses. The model is trained to minimize this loss, aiming to make predictions that closely match the ground truth annotations. In this study, the weight of the localization loss is increased since there is only one category.

Receptive Field Modelling for Spinal Injury Detection

Having detailed the network architecture in the preceding section, we now focus on a novel optimization method for detecting spinal injuries, which is a small object in medical images with no explicit boundaries, making challenges for precise bounding box definition. Traditional object detection methods use a "divide and conquer" principle, employing a Feature Pyramid Network (FPN) to detect objects of different scales across various layers. While beneficial for objects with clear edges, this strategy becomes less effective for our target object—spinal injuries.

To overcome this limitation, we propose to utilize the concept of the Effective Receptive Field (ERF) [38] for label assignment, which, unlike the traditional anchor-based and anchor-free detectors, does not rely on heuristic anchor box preset or scale grouping. The ERF is a subset of the total input space that a feature in a CNN responds to, making it particularly suitable for detecting objects with blurred or non-existent boundaries.

We modeled the ERF using a Gaussian distribution to measure the degree of matching between the ERF and the ground truth region. This approach aims to equate the effective region that a CNN feature effectively 'sees' or responds to, with the statistical properties of a Gaussian distribution. The Theoretical Receptive Field (TRF) [39], which represents the maximum potential receptive field that a feature in a CNN includes, is calculated using the following formula:

$$tr_n = tr_{n-1} + (k_n - 1)\prod_{i=1}^{n-1} S_i \tag{1}$$

Where trn denotes the TRF of each point on the n-th convolution layer, and kn and sn denote the kernel size and stride of the convolution operation on the n-th layer.

Since the ERF and TRF share the same center points, but the ERF covers only a part of the entire TRF, we approximate the radius of ERF as half the radius of the TRF. This approximation is then used as the co-variance for a standard 2-D Gaussian distribution.

Hence, we modeled the range of ERF into a 2-D Gaussian distribution Ne (μe, Σe) with:

$$\mu_e = \begin{bmatrix} x_n \\ y_n \end{bmatrix}, \Sigma_e = \begin{bmatrix} er_n^2 & 0 \\ 0 & er_n^2 \end{bmatrix} \tag{2}$$

Applying this method to our problem allows a more flexible and potentially more accurate detection of spinal injuries. By prioritizing the areas of the image that the network 'sees' most effectively, we can potentially improve the accuracy of our model, especially in cases where the injury does not have clear boundaries. Furthermore, this approach does not rely on strict bounding box definition, offering greater flexibility in detecting and classifying injuries of diverse sizes and shapes. The implementation details and empirical evidence indicating the advantages of our proposed method are elaborated in the subsequent sections.

Application of Wasserstein Distance for Spinal Injury Detection

Given the nature of spinal injury detection in CT scans, where typically only a single injury is present in each scan, the ground truth bounding box generally has little to no overlap the majority of the prior boxes and points. This feature makes the Wasserstein Distance particularly relevant to our scenario.

We utilized the Wasserstein Distance to measure the disparity between the distribution of the Effective Receptive Field (ERF) and the ground truth (gt) region. Since we have previously modeled both entities as Gaussian distributions, we can use the Wasserstein Distance to quantify their dissimilarity efficiently.

To understand this, Gaussian ERF represented as ne = Ne (μe, Σe) and Gaussian gt as ng = Ng (μg, Σg). Using these Gaussian distributions, the squared 2nd Wasserstein Distance can be simplified as:

$$W_2^2\left(n_e, n_g\right) = \left\| \left(\left[x_n, y_n, er_n, er_n\right]^T, \left[x_g, y_g, \frac{w_g}{2}, \frac{h_g}{2}\right]^T \right) \right\|_2^2 \tag{3}$$

Here, the Wasserstein Distance measures the discrepancy between the two non-overlapping distributions of the ERF and the gt region, providing a rational way to rank the priority of the different potential injury candidates for a specific ground truth. This unique property renders the Wasserstein Distance highly suited to our task, ensuring consistent reflection of the degree of match between all feature points and a specific ground truth box. The RFLA small target

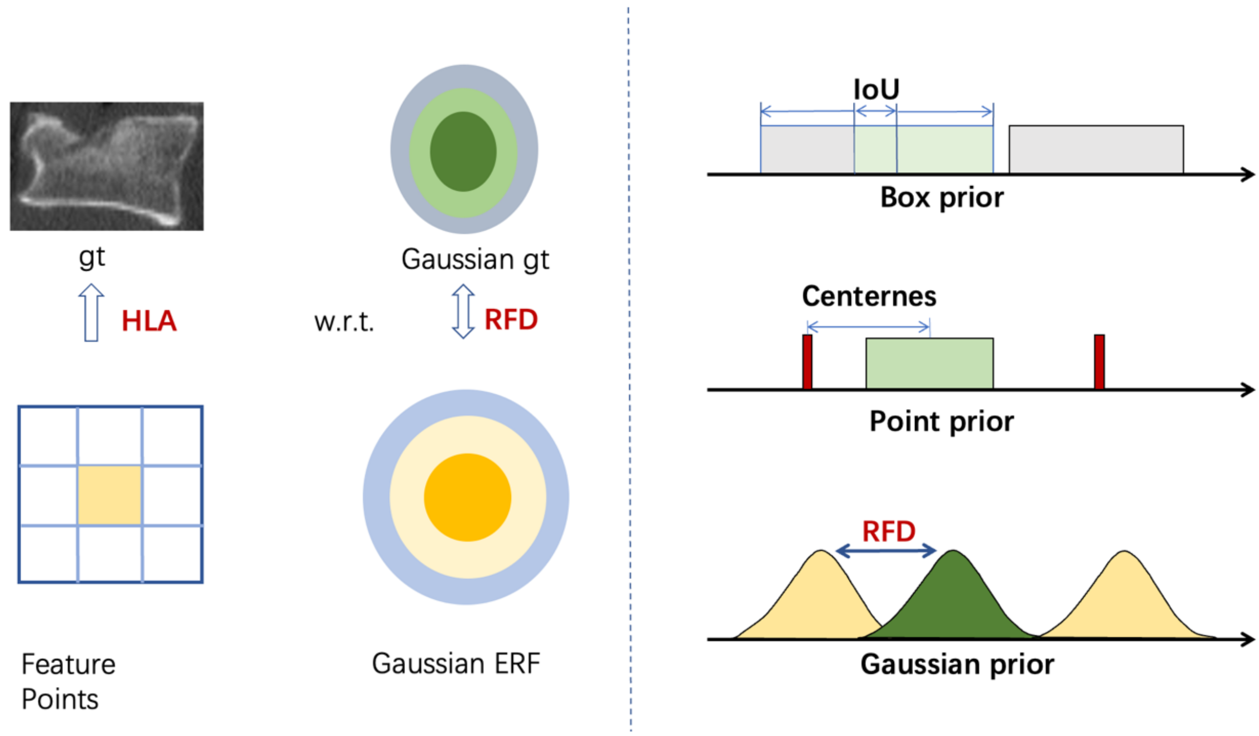

Fig. 4. Receptive Field Learning and Aggregation (RFLA) small target detection architecture. HLA, Hierarchical Localization and Attention; RFD, Receptive Field Dimension; IOU, Intersection over Union; ERF, Effective Receptive Field; gt, ground truth.

detection architecture, shown in Fig. 4, uses Gaussian Receptive Field-based Label Assignment and Wasserstein distance.

Interestingly, the two-dimensional Gaussian distribution derived from the bounding box can be intuitively understood as representing the 'influence' of the ground truth across the image, with the strongest influence at the center of the bounding box and decreasing influence as we move away. By reducing this distribution to one dimension along the path of the spine, we effectively condense the problem space while retaining the crucial characteristics essential for injury detection.

YOLO V3

YOLO V3 is a typical object detection model designed to identify and classify objects in images and videos. It works by dividing an image into a grid of cells, with each cell responsible for predicting whether it contains an object and, if so, assessing the class of the object. Unlike other models that use multiple CNNs, YOLO V3 uses a single CNN to make these predictions. The CNN consists of different layers, including convolutional layers, pooling layers, and fully connected (FC) layers. The convolutional layers are responsible for extracting features from the input image by applying a set of filters to the image and detecting patterns and features. The pooling layers down-sample the feature maps produced by the convolutional layers, reducing the size and complexity. The FC layers make predictions based on the features extracted by the convolutional and pooling layers by applying a set of weights and biases to the input features, producing a prediction for each cell in the grid. Compared to YOLO V7, YOLO V3 uses a single CNN for object detection predictions, while YOLO V7 uses a multi-scale network with an MPConv structure. This feature makes it faster and more efficient than other models, though it can result in lower accuracy.

Statistical Analysis

The model performance was assessed using three primary metrics: Accuracy, Kappa coefficient, and Mean Intersection over Union (MIOU). These metrics collectively provide a comprehensive assessment of the model's effectiveness in identifying spinal fractures.

• Accuracy: Accuracy is calculated as the ratio of correctly predicted instances to the total number of instances:

$$\text{Accuracy} = \frac{\text{Number of Correct Predictions}}{\text{Total Number of Predictions}}$$

• Kappa Coefficient: The Kappa coefficient measures the agreement between the predicted and actual classifications, adjusting for the possibility of the agreement occurring by chance:

$$\kappa = \frac{\text{Po} - \text{Pe}}{1 - \text{Pe}}$$

Where Po is the observed agreement and Pe is the expected agreement.

• Mean Intersection over Union (MIOU): MIOU is a common evaluation metric for segmentation tasks, calculated as the average of the Intersection over Union (IOU) for all classes:

$$\mathrm{IOU} = \frac{\text{Area of Overlap}}{\text{Area of Union}}$$

$$\mathrm{MIOU} = \frac{1}{N}\sum_{\mathrm{i}=1}^{N} \mathrm{IOUi}$$

Where N is the number of classes.

Experiment

In this study, we collected 200 datasets of spinal compressive fractures along with 40 images of healthy spines. The original image had dimensions of approximately 1170 × 2088 px. Typically, each image contained only one fractured area, with the damaged bone having a pixel width of approximately 240 px, and the fracture itself is about 40 px wide. These characteristic regions occupy a smaller portion of the overall image. Additionally, the features within the fracture area are often fragmented, lacking distinct edges, making it challenging for accurate annotation and identification. In our dataset, when assigning labels, we exclusively outlined the core area of the bone injury.

Summary of Methods

We developed a novel method for detecting spinal fractures in CT images using the YOLO V7 model, enhanced with ELAN and MPConv architectures. The dataset consisted of 240 CT images from Ningbo No.2 Hospital, including 200 pathological and 40 normal spinal images. We employed standard normalization techniques for preprocessing. To address the challenges of small-scale and blurred boundary fractures, we integrated Theoretical Receptive Field (TRF) and Effective Receptive Field (ERF) frameworks. Additionally, the Wasserstein distance was utilized to optimize the model's learning process. The performance of our method was compared against conventional object detection networks like YOLO V7 and YOLO V3.

Results

Fig. 5 depicts five distinct cases of spinal compression fractures, analyzed using three models: YOLO V7, YOLO V3, and our proposed method for fracture location identification. The ground truth, determined by orthopedic experts, serves as the benchmark. Table 1 provides the statistical analysis of these methods, including accuracy, Kappa, and MIOU. The results indicate that our method outperforms the others, achieving the highest accuracy and precisely outlining the area of bone injury. YOLO V7 follows closely with reasonable precision, while YOLO V3 occasionally misses detections and misinterprets scenarios.

Table 1. Accuracy, Kappa, and MIOU statistics for YOLO V7, YOLO V3, and our method.

Method	Accuracy	Kappa	MIOU
Our method	93.10%	0.867	0.734
YOLO V7	89.50%	0.832	0.706
YOLO V3	82.20%	0.765	0.682

MIOU, Mean Intersection over Union.

Spinal compressive fractures occur when pressure causes a section of the spinal bone to break. This condition is characterized by distinct fracture lines visible on CT scans, thinning of vertebral bone matter, and compression of fractures. Imaging shows localized damage and fractures at the bone's edge. However, these features are difficult to distinguish, and manual annotation often requires marking the entire spinal bone, even though the target region is much smaller. This complicates the accurate identification of spinal compressive fractures.

To address this issue, YOLO V7 uses the ELAN and MPConv structures, which preserve more information during the network's forward propagation, thereby better balancing large-scale and small-scale features. Additionally, to mitigate the issue of ambiguous boundaries in compressive fractures, this study employs a TRF based on Gaussian distribution to measure distance. Consequently, features at a rectangle's central location are given more weight. This method amplifies the image features predominantly concentrated in the label's center, enhancing recognition accuracy and yielding results that closely match the ground truth. Conversely, YOLO V3, rooted in the traditional Encoder/Decoder architecture, may lose some small-scale information during propagation, potentially leading to misjudgments and missed detections.

In practical applications, we often rely on coarsely annotated labels. The standard practice involves annotating the entire cross-section of the spine. Fig. 6 depicts prediction results from a dataset with coarse annotations, with accuracy, Kappa, and MIOU statistics for these methods (Table 2). Notably, the injured area may only constitute a minor portion of the image. In such situations, conventional methods like YOLO V7 and YOLO V3 compute the distance between the bounding box and the ground truth linearly. This causes the target detection network to learn the complete cross-sectional features of the spine, identifying all vertebrae. Conversely, our proposed method employs TRF, assigning more weight to the central features when measuring distance. This approach demonstrates enhanced adaptability to datasets with coarse annotations by focusing on the injury's more crucial central features rather than the entire cross-section of the spine. This nuanced approach helps better distinguish between injured and non-injured regions, thereby enhancing detection precision in real-world scenarios.

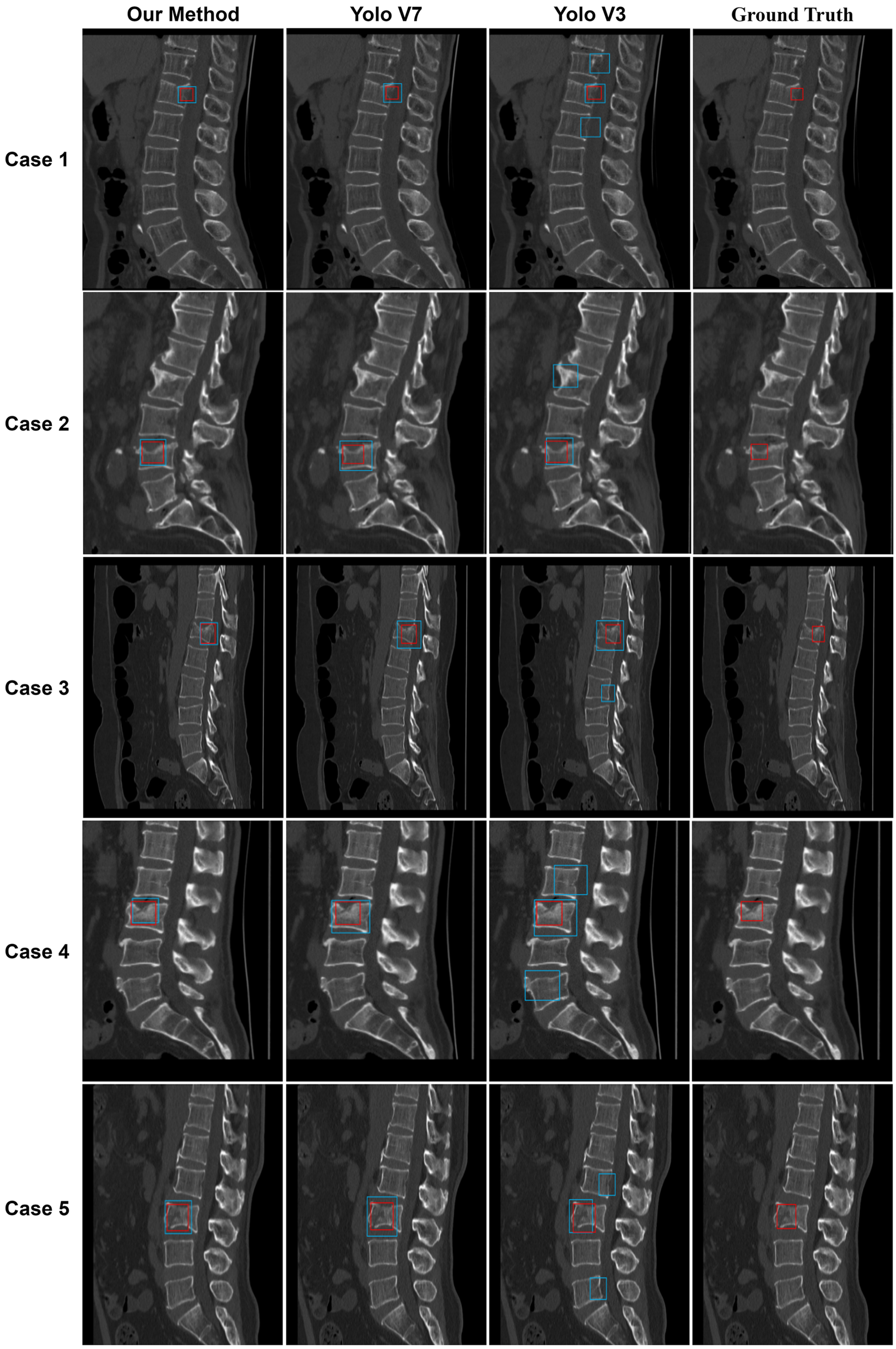

Fig. 5. Analysis of spinal compression fractures using YOLO V7, YOLO V3, and our proposed method. Comparison of spinal fracture identification in five different cases using YOLO V7, YOLO V3, and our method (the red boxes represent the ground truth locations of the compressive fractures, while the blue boxes indicate the results calculated or recognized by the various models).

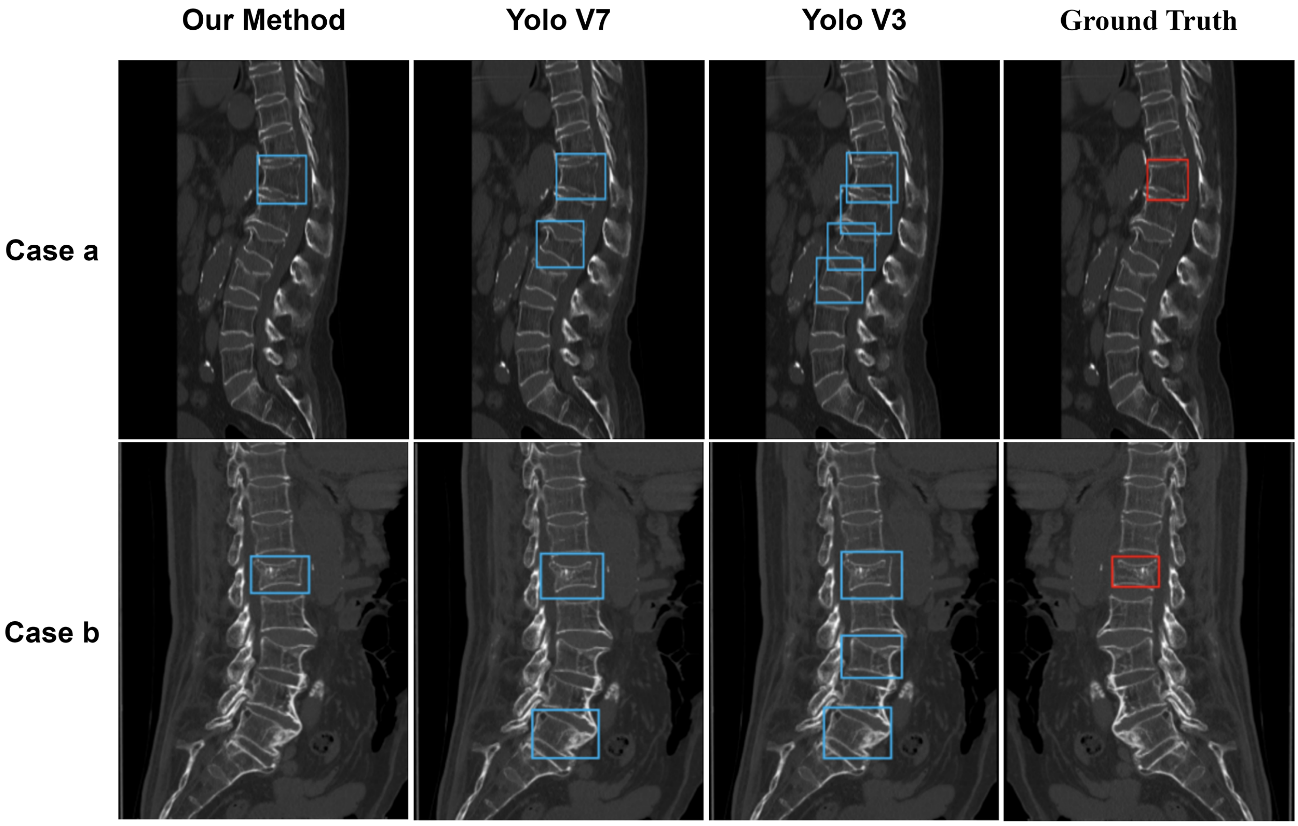

Fig. 6. Prediction results from a dataset with coarse annotations using YOLO V7, YOLO V3, and our proposed method. (The red boxes represent the ground truth locations of the compressive fractures, while the blue boxes indicate the results calculated or recognized by the various models.)

Table 2. Accuracy, Kappa, and MIOU statistics for YOLO V7, YOLO V3, and our method in a coarsely annotated dataset.

Method	Accuracy	Kappa	MIOU
Our method	78.70%	0.714	0.652
YOLO V7	69.50%	0.636	0.579
YOLO V3	63.40%	0.572	0.534

Discussion

Identifying vertebral compression fracture areas in CT images utilizing intelligent algorithms has been a major challenge in medical imaging. Deep learning-based object detection networks, presently among the most advanced algorithms in this area, have been widely applied to various medical imaging recognition tasks [40, 41]. Existing research primarily utilizes the traditional Encoder-Decoder architecture of fully convolutional networks [42]. However, these networks often grapple with information loss during the encoding and decoding and managing multi-scale information [36, 43]. In spinal injuries, CT images frequently show discontinuities in bone matter, localized changes in bone density, and irregular vertebral edge lines. Due to their relatively small scale and obscure boundaries, fracture areas in these images pose significant recognition challenges. Traditional object detection methods often fail to achieve high-precision detection in these scenarios [44].

In response to these challenges, we propose a novel approach for targeted spinal injury detection. This method combines the state-of-the-art YOLO V7 network architecture with the Gaussian Receptive Field framework. By providing a more precise definition of the distance between the ground truth and prior boxes, this framework effectively addresses the issue of identifying small spinal injury targets with blurred boundaries.

Our study included a dataset consisting of 200 pathological images and 40 normal spinal images. To benchmark our method, we used YOLO V7 and YOLO V3, two widely used object detection networks. Our findings indicated that the YOLO V7 network, featuring the ELAN and MPConv frameworks, retained more small-scale information, thus facilitating the detection of a larger number of targets and improving the accuracy by 3%.

Furthermore, the method proposed in this study offers a more precise definition of the distance between the ground truth and prior boxes. Compared to the traditional YOLO V7, it enhances accuracy by an additional 3%. These findings facilitate improved detection and diagnosis of spinal injuries, revealing the potential for deep learning-based object detection networks in medical imaging applications.

Future research can improve detection accuracy further and enable the practical deployment of these models in clinical settings.

Our findings align with previous research demonstrating the advantages of advanced deep-learning models in medical imaging tasks. For instance, Ronneberger *et al*. [45] (2015) illustrated that integrating multi-scale feature extraction techniques in U-Net significantly enhances the segmentation of biomedical images, including small structures. Similarly, Liao *et al*. [46] (2019) reported that using advanced architectures like ResNet improves the accuracy of pulmonary nodule detection in chest CT scans. Our integration of the ELAN and MPConv frameworks with YOLO V7 further refines the detection of small-scale spinal fractures.

Despite our model's promising performance on the training set, the absence of validation and test sets may lead to several limitations: first, the model may overfit the training data, resulting in inadequate generalization. Second, the lack of independent datasets to evaluate the model's performance might introduce bias in the results. The rationale for not using validation and test sets is as follows: due to the limited size of our dataset, we prioritized using all available data for training to maximize the model's learning capacity. Additionally, this study is exploratory in nature, aiming to verify the feasibility and potential performance of the new method.

Despite these promising findings, our study has several limitations. Firstly, although the dataset is robust, it remains limited compared to the vast diversity of spinal fracture cases in the real world. Secondly, our method primarily focuses on compressive fractures and may need to be adjusted to effectively detect other types of spinal injuries. Lastly, the computational complexity of the proposed model may pose challenges for real-time clinical applications, necessitating additional optimization. These limitations indicate that future research should incorporate validation and test sets to comprehensively evaluate the model's performance and verify its robustness and generalizability.

These findings pave the way for improved detection and diagnosis of spinal injuries, indicating the potential for deep learning-based object detection networks in medical imaging applications. Our method can be used as a reference for future research to improve detection and accuracy and facilitate the practical deployment of such models in clinical settings.

Conclusions

This study addresses a significant challenge in medical imaging: identifying vertebral compression fractures in CT scans using advanced algorithms. Our novel approach uses the YOLO V7 architecture, combined with the Gaussian Receptive Field framework, to enhance the detection accuracy of small-scale spinal injuries with blurred boundaries. The effectiveness of this approach is validated using a dataset of 200 pathological images and 40 normal spinal images. The findings demonstrate superior performance than traditional object detection networks like YOLO V7 and YOLO V3, achieving a 3% increase in accuracy compared to YOLO V7. These promising results highlight the potential of our proposed method to improve the detection and diagnosis of vertebral compression fractures, paving the way for further research and optimization for practical application in clinical settings. The ultimate goal is to harness the potential of deep learning to improve patient care and treatment in medical imaging.

To further validate the generalizability and applicability of our results, future research should incorporate several key strategies. Firstly, cross-validation within the dataset should be performed to assess the model's robustness. Secondly, the introduction of independent validation and test sets is essential for comprehensive evaluation of the model's performance. These steps will help verify the model's robustness and generalizability, ensuring its practical utility in diverse clinical scenarios. Additionally, exploring other deep learning models and methods could further enhance detection accuracy and reliability. By addressing these areas, subsequent studies can build upon our findings and contribute to the development of more effective diagnostic tools in medical imaging.

Availability of Data and Materials

The datasets used during the current study are available from the corresponding author on reasonable request.

Author Contributions

HTW: Methodology, Data collection, Analysis, Investigation, Writing—Original Draft; QSF: Conceptualization, Formal analysis, Visualization, Supervision, Writing—Reviewing. Both authors revised the manuscript critically for important intellectual content. Both authors read and approved the final manuscript. Both authors have participated sufficiently in the work and agreed to be accountable for all aspects of the work.

Ethics Approval and Consent to Participate

The present study followed the Declaration of Helsinki. This study was approved by the Ethics Committee of Ningbo No.2 Hospital (No. PJ-NBEY-KJ-2022-08-15), and written informed consent was obtained from all subjects participating in the trial, and their information was stored and used for research anonymously.

Acknowledgment

Not applicable.

Funding

This work was supported by Medical Scientific Research Foundation of Zhejiang Province (2019KY597), Ningbo Public Service Technology Foundation (2022S063), Ningbo Medical Key Supporting Discipline: Orthopaedics (2022-F15) and Zhu Xiushan Talent Award Fund (2023HMYQ15).

Conflict of Interest

The authors declare no conflict of interest.

References

[1] Bhandary A, Prabhu GA, Rajinikanth V, Thanaraj KP, Satapathy SC, Robbins DE, *et al.* Deep-learning framework to detect lung abnormality–A study with chest X-Ray and lung CT scan images. Pattern Recognition Letters. 2020; 129: 271–278.

[2] Levine AB, Schlosser C, Grewal J, Coope R, Jones SJM, Yip S. Rise of the Machines: Advances in Deep Learning for Cancer Diagnosis. Trends in Cancer. 2019; 5: 157–169.

[3] Yoon AP, Lee YL, Kane RL, Kuo CF, Lin C, Chung KC. Development and Validation of a Deep Learning Model Using Convolutional Neural Networks to Identify Scaphoid Fractures in Radiographs. JAMA Network Open. 2021; 4: e216096.

[4] Brutti F, Fantazzini A, Finotello A, Müller LO, Auricchio F, Pane B, *et al.* Deep Learning to Automatically Segment and Analyze Abdominal Aortic Aneurysm from Computed Tomography Angiography. Cardiovascular Engineering and Technology. 2022; 13: 535–547.

[5] Biswas M, Kuppili V, Saba L, Edla DR, Suri HS, Cuadrado-Godia E, *et al.* State-of-the-art review on deep learning in medical imaging. Frontiers in Bioscience (Landmark Edition). 2019; 24: 392–426.

[6] Acharya UR, Fernandes SL, WeiKoh JE, Ciaccio EJ, Fabell MKM, Tanik UJ, *et al.* Automated Detection of Alzheimer's Disease Using Brain MRI Images- A Study with Various Feature Extraction Techniques. Journal of Medical Systems. 2019; 43: 302.

[7] Li W, Jia F, Hu Q. Automatic segmentation of liver tumor in CT images with deep convolutional neural networks. Journal of Computer and Communications. 2015; 3: 146–151.

[8] Yang D, Xu D, Zhou SK, Georgescu B, Chen M, Grbic S, *et al.* Automatic liver segmentation using an adversarial image-to-image network. In Medical Image Computing and Computer Assisted Intervention- MICCAI 2017: 20th International Conference, Quebec City, QC, Canada, September 11-13, 2017, Proceedings, Part III 20. 2017; 507–515.

[9] Sun C, Guo S, Zhang H, Li J, Chen M, Ma S, *et al.* Automatic segmentation of liver tumors from multiphase contrast-enhanced CT images based on FCNs. Artificial Intelligence in Medicine. 2017; 83: 58–66.

[10] Lin L, Dou Q, Jin YM, Zhou GQ, Tang YQ, Chen WL, *et al.* Deep Learning for Automated Contouring of Primary Tumor Volumes by MRI for Nasopharyngeal Carcinoma. Radiology. 2019; 291: 677–686.

[11] Ebsim R, Naqvi J, Cootes TF. Automatic detection of wrist fractures from posteroanterior and lateral radiographs: a deep learning-based approach. In Computational Methods and Clinical Applications in Musculoskeletal Imaging: 6th International Workshop, MSKI 2018, Held in Conjunction with MICCAI 2018, Granada, Spain, September 16, 2018, Revised Selected Papers 6. Springer International Publishing. 2019; 114–125.

[12] Baskaran L, Al'Aref SJ, Maliakal G, Lee BC, Xu Z, Choi JW, *et al.* Automatic segmentation of multiple cardiovascular structures from cardiac computed tomography angiography images using deep learning. PloS One. 2020; 15: e0232573.

[13] Esteva A, Kuprel B, Novoa RA, Ko J, Swetter SM, Blau HM, *et al.* Dermatologist-level classification of skin cancer with deep neural networks. Nature. 2017; 542: 115–118.

[14] Cheng CT, Ho TY, Lee TY, Chang CC, Chou CC, Chen CC, *et al.* Application of a deep learning algorithm for detection and visualization of hip fractures on plain pelvic radiographs. European Radiology. 2019; 29: 5469–5477.

[15] Derkatch S, Kirby C, Kimelman D, Jozani MJ, Davidson JM, Leslie WD. Identification of Vertebral Fractures by Convolutional Neural Networks to Predict Nonvertebral and Hip Fractures: A Registry-based Cohort Study of Dual X-ray Absorptiometry. Radiology. 2019; 293: 405–411.

[16] Litjens G, Ciompi F, Wolterink JM, de Vos BD, Leiner T, Teuwen J, *et al.* State-of-the-Art Deep Learning in Cardiovascular Image Analysis. JACC. Cardiovascular Imaging. 2019; 12: 1549–1565.

[17] Voulodimos A, Protopapadakis E, Katsamenis I, Doulamis A, Doulamis N. A Few-Shot U-Net Deep Learning Model for COVID-19 Infected Area Segmentation in CT Images. Sensors (Basel, Switzerland). 2021; 21: 2215.

[18] Talo M, Baloglu UB, Yıldırım Ö, Acharya UR. Application of deep transfer learning for automated brain abnormality classification using MR images. Cognitive Systems Research. 2019; 54: 176–188.

[19] Perone CS, Cohen-Adad J. Promises and limitations of deep learning for medical image segmentation. Journal of Medical Artificial Intelligence. 2019; 2.

[20] Shin HC, Tenenholtz NA, Rogers JK, Schwarz CG, Senjem ML, Gunter JL, *et al.* Medical image synthesis for data augmentation and anonymization using generative adversarial networks. In Simulation and Synthesis in Medical Imaging: Third International Workshop, SASHIMI 2018, Held in Conjunction with MICCAI 2018, Granada, Spain, September 16, 2018, Proceedings 3. 2018; 1–11.

[21] Frid-Adar M, Diamant I, Klang E, Amitai M, Goldberger J, Greenspan H. GAN-based synthetic medical image augmentation for increased CNN performance in liver lesion classification. Neurocomputing. 2018; 321: 321–331.

[22] Boveiri HR, Khayami R, Javidan R, Mehdizadeh A. Medical image registration using deep neural networks: a

comprehensive review. Computers & Electrical Engineering. 2020; 87: 106767.
[23] Cao X, Fan J, Dong P, Ahmad S, Yap PT, Shen D. Image registration using machine and deep learning. Handbook of medical image computing and computer assisted intervention (pp. 319–342). Elsevier: Academic Press. 2020.
[24] Ward TM, Mascagni P, Ban Y, Rosman G, Padoy N, Meireles O, *et al*. Computer vision in surgery. Surgery. 2021; 169: 1253–1256.
[25] Chen JW, Lin W J, Lin CY, Hung CL, Hou CP, Cho CC, *et al*. Automated classification of blood loss from transurethral resection of the prostate surgery videos using deep learning technique. Applied Sciences. 2020; 10: 4908.
[26] Samarasinghe G, Jameson M, Vinod S, Field M, Dowling J, Sowmya A, *et al*. Deep learning for segmentation in radiation therapy planning: a review. Journal of Medical Imaging and Radiation Oncology. 2021; 65: 578–595.
[27] Jarrett D, Stride E, Vallis K, Gooding MJ. Applications and limitations of machine learning in radiation oncology. The British Journal of Radiology. 2019; 92: 20190001.
[28] Li Z, Dong M, Wen S, Hu X, Zhou P, Zeng Z. CLU-CNNs: Object detection for medical images. Neurocomputing. 2019; 350: 53–59.
[29] Baumgartner M, Jäger PF, Isensee F, Maier-Hein KH. nnDetection: a self-configuring method for medical object detection. In Medical Image Computing and Computer Assisted Intervention–MICCAI 2021: 24th International Conference, Strasbourg, France, September 27–October 1, 2021, Proceedings, Part V 24. 2021; 530–539.
[30] Alam F, Rahman SU, Ullah S, Gulati K. Medical image registration in image guided surgery: Issues, challenges and research opportunities. Biocybernetics and Biomedical Engineering. 2018; 38: 71–89.
[31] Kelly CJ, Karthikesalingam A, Suleyman M, Corrado G, King D. Key challenges for delivering clinical impact with artificial intelligence. BMC Medicine. 2019; 17: 1–9.
[32] Currie G, Hawk KE, Rohren E, Vial A, Klein R. Machine Learning and Deep Learning in Medical Imaging: Intelligent Imaging. Journal of Medical Imaging and Radiation Sciences. 2019; 50: 477–487.
[33] Wang CY, Bochkovskiy A, Liao HYM. YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for realtime object detectors. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. arXiv. 2023; 7464–7475. (preprint)
[34] Redmon J, Farhadi A. Yolov3: An incremental improvement. arXiv. 2018. (preprint)
[35] Adler J, Lunz S. Banach wasserstein gan. Advances in Neural Information Processing Systems. 2018; 31.
[36] Chen G, Zheng YD, Chen Z, Wang J, Lu T. ELAN: Enhancing Temporal Action Detection with Location Awareness. In 2023 IEEE International Conference on Multimedia and Expo (ICME). 2023; 1020–1025.
[37] Sori WJ, Feng J, Liu S. Multi-path convolutional neural network for lung cancer detection. Multidimensional Systems and Signal Processing. 2019; 30: 1749–1768.
[38] Luo W, Li Y, Urtasun R, Zemel R. Understanding the effective receptive field in deep convolutional neural networks. Advances in Neural Information Processing Systems. 2016; 29.
[39] Kulasingham JP, Simon JZ. Algorithms for Estimating Time-Locked Neural Response Components in Cortical Processing of Continuous Speech. IEEE Transactions on Bio-medical Engineering. 2023; 70: 88–96.
[40] Litjens G, Kooi T, Bejnordi BE, Setio AAA, Ciompi F, Ghafoorian M, *et al*. A survey on deep learning in medical image analysis. Medical Image Analysis. 2017; 42: 60–88.
[41] Chan HP, Samala RK, Hadjiiski LM, Zhou C. Deep learning in medical image analysis. Deep Learning in Medical Image Analysis: Challenges and Applications. 2020; 3–21.
[42] Long J, Shelhamer E, Darrell T. Fully Convolutional Networks for Semantic Segmentation. In Proceedings of the IEEE conference on computer vision and pattern recognition. 2015; 3431–3440.
[43] Huang G, Liu Z, Van Der Maaten L, Weinberger KQ. Densely connected convolutional networks. In Proceedings of the IEEE conference on computer vision and pattern recognition. 2017; 4700–4708.
[44] He K, Zhang X, Ren S, Sun J. Delving deep into rectifiers: Surpassing human-level performance on imagenet classification. In Proceedings of the IEEE international conference on computer vision. 2015; 1026–1034.
[45] Ronneberger O, Fischer P, Brox T. U-Net: Convolutional Networks for Biomedical Image Segmentation. In Medical Image Computing and Computer-Assisted Intervention. Springer: Cham. 2015; 234–241.
[46] Liao F, Liang M, Li Z, Hu X, Song S. Evaluate the Malignancy of Pulmonary Nodules Using the 3-D Deep Leaky Noisy-OR Network. IEEE Transactions on Neural Networks and Learning Systems. 2019; 30: 3484–3495.

Publisher's Note: *Annali Italiani di Chirurgia* stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.