BMC Medical Imaging

RESEARCH

Open Access

Grayscale medical image segmentation method based on 2D&3D object detection with deep learning

Yunfei Ge[1], Qing Zhang[1], Yuantao Sun[1*], Yidong Shen[2] and Xijiong Wang[3]

Abstract

Background: Grayscale medical image segmentation is the key step in clinical computer-aided diagnosis. Model-driven and data-driven image segmentation methods are widely used for their less computational complexity and more accurate feature extraction. However, model-driven methods like thresholding usually suffer from wrong segmentation and noises regions because different grayscale images have distinct intensity distribution property thus pre-processing is always demanded. While data-driven methods with deep learning like encoder-decoder networks always are always accompanied by complex architectures which require amounts of training data.

Methods: Combining thresholding method and deep learning, this paper presents a novel method by using 2D&3D object detection technologies. First, interest regions contain segmented object are determined with fine-tuning 2D object detection network. Then, pixels in cropped images are turned as point cloud according to their positions and grayscale values. Finally, 3D object detection network is applied to obtain bounding boxes with target points and boxes' bottoms and tops represent thresholding values for segmentation. After projecting to 2D images, these target points could composite the segmented object.

Results: Three groups of grayscale medical images are used to evaluate the proposed image segmentation method. We obtain the IoU (DSC) scores of 0.92 (0.96), 0.88 (0.94) and 0.94 (0.94) for segmentation accuracy on different datasets respectively. Also, compared with five state of the arts and clinically performed well models, our method achieves higher scores and better performance.

Conclusions: The prominent segmentation results demonstrate that the built method based on 2D&3D object detection with deep learning is workable and promising for segmentation task of grayscale medical images.

Keywords: Grayscale medical image, Image segmentation, Deep learning, Object detection, Point cloud

Background

Medical imaging plays the key role in diagnosis or disease treatment by revealing internal structures with technologies mainly of computer tomography (CT), magnetic resonance imaging (MRI), ultrasound, and especially X-ray radiography [1]. Due to different absorption capability of various organs or tissues for radiations, waves, and etc., pixels belong to various object in grayscale medical images have diverse grayscale values usually from 0 to 255 [2] and meanwhile values of pixels of the same object always gather within a range.

Medical image segmentation has been widely applied to make images clearer with anatomical or pathological structures changes [3], such as bone segmentation [4], lung segmentation [5, 6], heart fat segmentation [7], liver or liver-tumor segmentation [8, 9] and Intracranial

*Correspondence: sun1979@sina.com
[1] School of Mechanical Engineering, Tongji University, Shanghai, China
Full list of author information is available at the end of the article

hemorrhage segmentation [10, 11], etc. They could be considered to divide origin images into several sub regions for picking up some crucial objects and extracting interesting features which improve the computer aided diagnostic efficiency. There has raised enormous approaches and they could be classified into two categories: model-driven techniques and data-driven techniques [5, 12].

Many model-driven methods for medical image segmentation, including thresholding, clustering, and region growing, were presented in particular before the widespread application of deep learning [12]. Thresholding was one of the most common used method in practice due to its efficiency [13]. The basic working of thresholding was to determine specific threshold values and each pixel in the image could be classified as the foreground or background depending on the comparison between their intensity values and threshold values [14–16]. Traditional thresholding methods always relied on single models for universal segmentation tasks which could lead to incorrect results. Also, segmentation objects often occupied only parts of whole images and pixels of different objects may share same intensity values, so noises could appear if image segmentation was applied overall.

With the era of big data coming, emerging data-driven technologies with deep learning have remarkably demonstrated in variety medical image segmentation task. Supervised learning methods and especially some convolutional neural network (CNN) based encoder-decoder structures such as fully convolutional networks (FCN) [17], U-Net [18], DeepLab [19] has practically proved [5]. Compared with traditional methods, deep learning could help analyze medical images more effectively and extract more detailed features.

Although these end-to-end structures was pragmatic for medical images semantic segmentation, the segmentation accuracy always relied on a large amount of training dataset. But medical image annotation could be time-consuming and quite expensive, thus transfer learning was used to solve the problem of limited labeled data and pre-trained networks on natural images as ImageNet [20] were often adopted for image segmentation [21, 22]. However, considering these datasets were mainly designed to train models for object detection or classification, they may be more suitable to pretrain networks for object detection. This inspired us to segment images with object detection. We find that grayscale images could be segmented according to the comparison of thresholding values with values of pixels in images and these pixels could be turned into 3D point cloud according to their positions and grayscale values. Thus, by applying 3D object detection in the point cloud, we could achieve groups of points within 3D bounding boxes. The top and the bottom of boxes represent the thresholding values for segmentation and after mapping these points into 2D images, corresponding pixels could compose segmented results. Besides, 2D object detection could determinate regions of interest (ROI) in grayscale medical images to reduce noises. Therefore, according to above strategy, we propose the grayscale medical image segmentation method based on 2D&3D object detection.

The remainder of this paper is organized as following: second section introduces the applied medical Image datasets and describes details of proposed technologies, while in third section the obtained results are displayed and the discussion is provided. Finally, forth section presents the conclusions as well as future work suggestions.

Methods

Image datasets

Since bone and chest X-ray images are the most common grayscale medical images in clinical, two typical sets of available datasets are prepared including musculoskeletal radiographs, and chest radiographs. Musculoskeletal radiographs about upper and lower extremity includes musculoskeletal radiographs (MURA) [23], lower extremity radiographs (LERA) [24] and prepared phalanx and forearm X-ray images. Chest radiographs mainly come from chest radiography (CheXpert) [25, 26]. MURA and LERA are large datasets of bone radiographs from Stanford University and they contain X-ray images about the upper and lower extremity respectively. CheXpert is a large dataset of chest radiographs and it is also from Stanford University. Besides, phalanx and forearm X-ray images are obtained with the portable X-ray machine as Fig. 1 Shown. Totally, 2509 cases among MURA and LERA, 3100 cases of CheXpert and 500 phalanx and forearm X-ray images are adopted for models training and validation.

The proposed grayscale medical image segmentation method is based on the supervised artificial intelligence

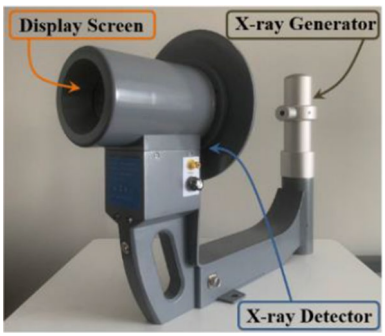


Fig. 1 The portable X-ray machine

techniques, and labels are performed manually in two types medical images for model training. Figure 2 shows origin images, and their respective Ground Truth (GT) images in different datasets.

Grayscale image segmentation framework

The proposed image segmentation method maps each pixel in the medical grayscale image to 3D coordinates as the pixel-features point cloud, according to their positions and gray values. By acquisition of foreground points and their corresponding bounding box using 3D object detection method, we could achieve threshold values and the segmentation result of the corresponding grayscale image. The whole pipeline and the implementation flow of this method are shown in Figs. 3 and 4 respectively. Given a grayscale medical image, after (1) obtaining interest regions of associated segmentation objects in the image, (2) generating 3D bounding box proposals in point cloud and (3) the regression of their locations and scales, the refined boxes could be achieved. The projection of points in refined bounding box into the 2D image is the segmentation result.

Related work

According to the proposed strategy and above pipeline, object detections play the central roles at each block of our method. Many researches about 2D&3D object detection has raised ever and they could perform well especially those with deep learning.

The current mainstream 2D object detection methods based on deep learning could be generally classified into two-stage and one-stage methods [27]. With two-stage methods, proposal bounding boxes are generated firstly and the further refinement of proposals and confidences is obtained in the second stage [28]. While using the one-stage methods [29, 30], the location and the classification of object bounding boxes could be estimated directly without refinement which means one-stage methods are usually faster than two-stage ones but have lower object detection accuracy [31].

The widespread application of 3D geometric data spurs the development of 3D object detection and it could be categorized into monocular/stereo image-based, point cloud-based and multimodal fusion-based methods in terms of the modality of input data [32]. Due to point clouds are the most regular data which could be achieved

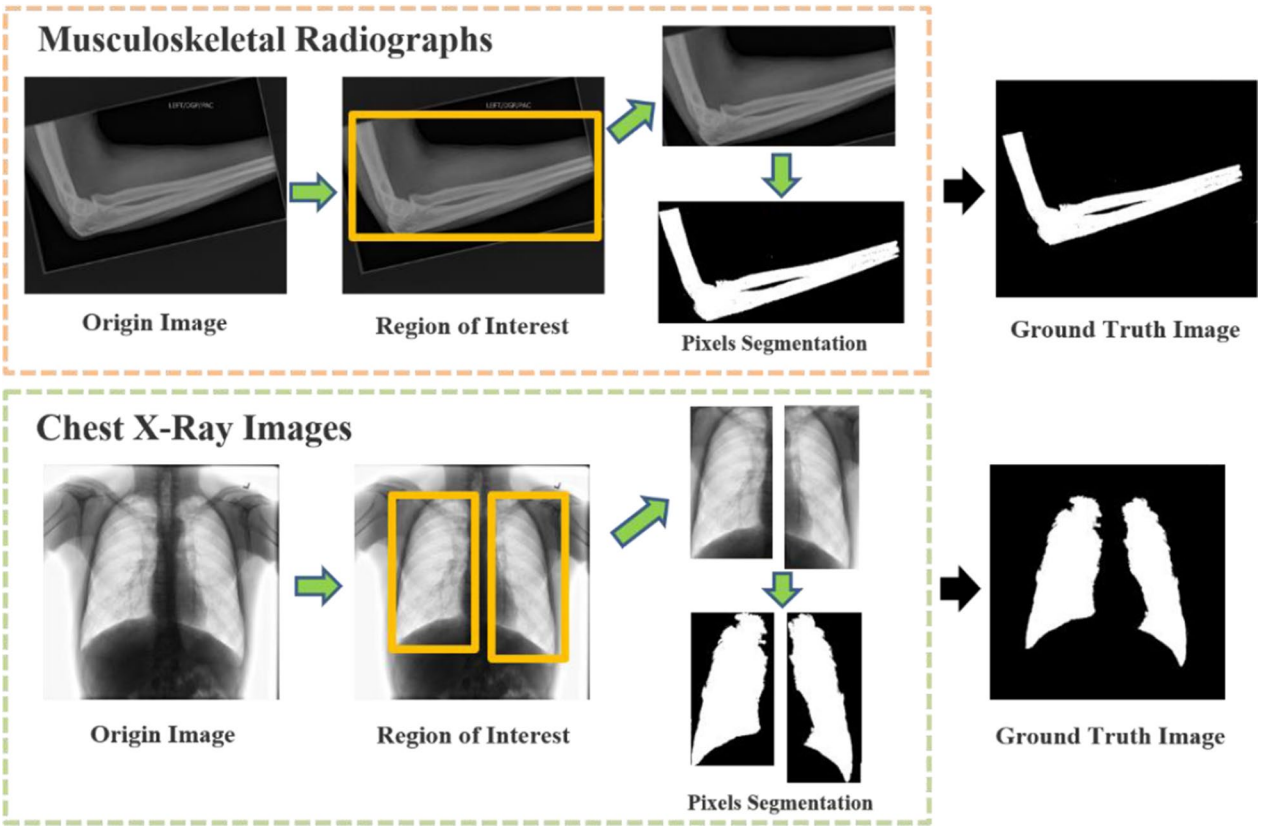


Fig. 2 Examples of medical images in two datasets and manual segmentation results

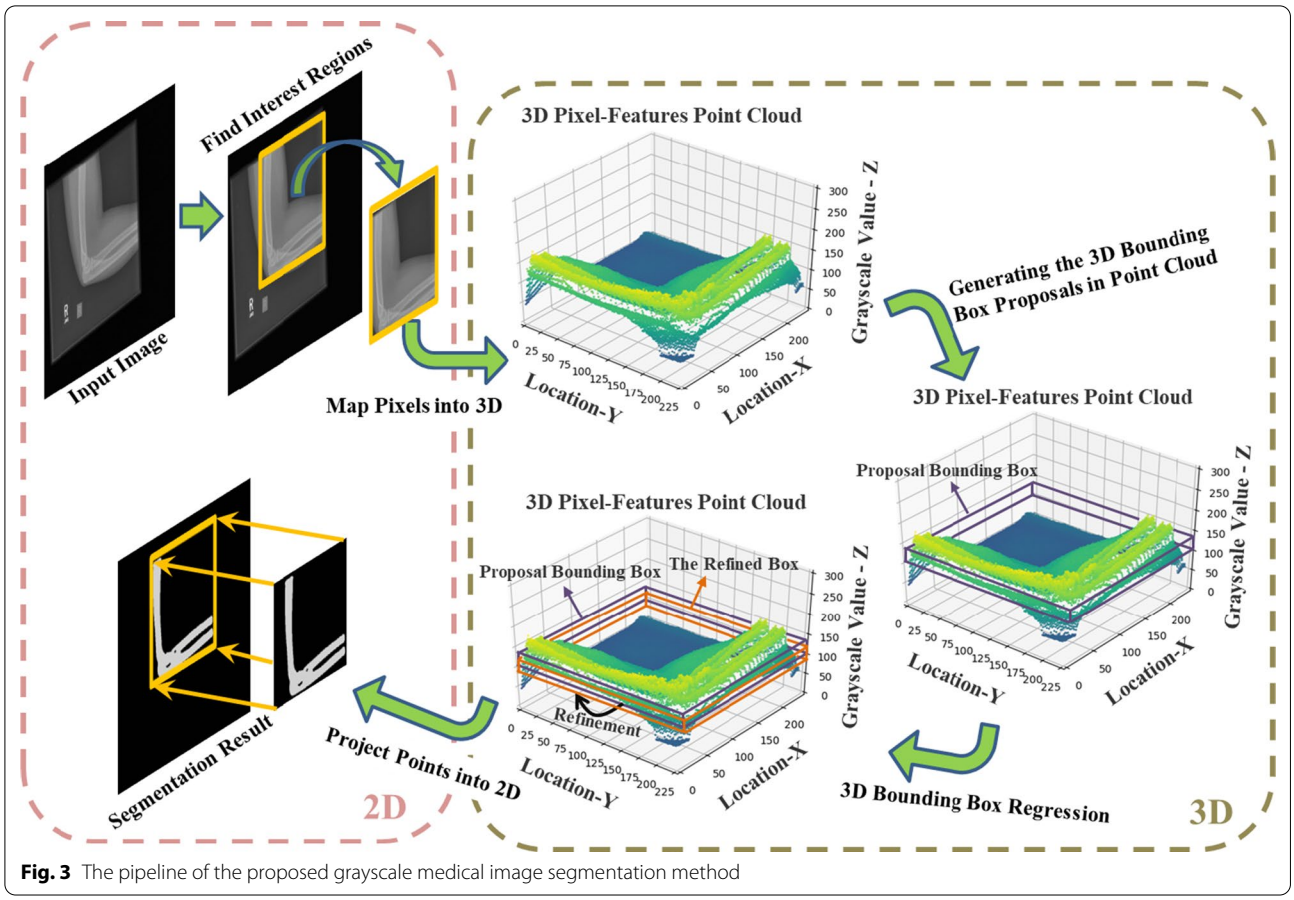


Fig. 3 The pipeline of the proposed grayscale medical image segmentation method

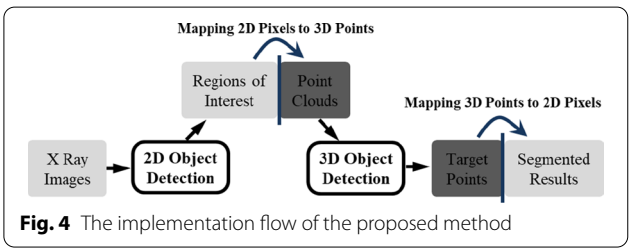


Fig. 4 The implementation flow of the proposed method

with different sensors, enormous researches of point cloud-based methods have raised [33–35]. Among these method, different data format like raw point clouds or 3D voxel grids transformed from points could be feed into deep net architectures to find targets with bounding boxes and their classes [36].

Achievement of interest regions in image

In a medical grayscale image, pixels of the segmentation object always just take up a part of the entire image and there may exists noisy pixels with the same gray values in irrelevant regions. Therefore, 2D object detection is adopted as the pre-processing procedure to identify the specially interest regions with segmentation objects and reduce noisy pixels as shown in Fig. 5.

Compared with the accuracy, the proposed pre-processing procedure cares more about the detection speed, so we adopt the one-stage method YOLOv3 [37, 38] as the backbone network. And considering the scarcity of labeled medical grayscale images, we apply the fine tuning—a transfer learning method [31] to migrate most layers of the backbone model which was pretrained on ImageNet, Pascal VOC (Pattern analysis, statistical modeling and computational learning visual object classes) and MS COCO (Microsoft common objects in context) datasets [39, 40]. As Fig. 6. shown, with fine tuning method, we could freeze N–M layers of pre-trained model and only train the last M layers on local dataset. In order to retain the detection ability of pre-trained model as much as possible, and ensure the stability of the loss change during the training process, the proposed image segmentation pre-processing method only unfreeze the last 3 layers of pre-trained network for training.

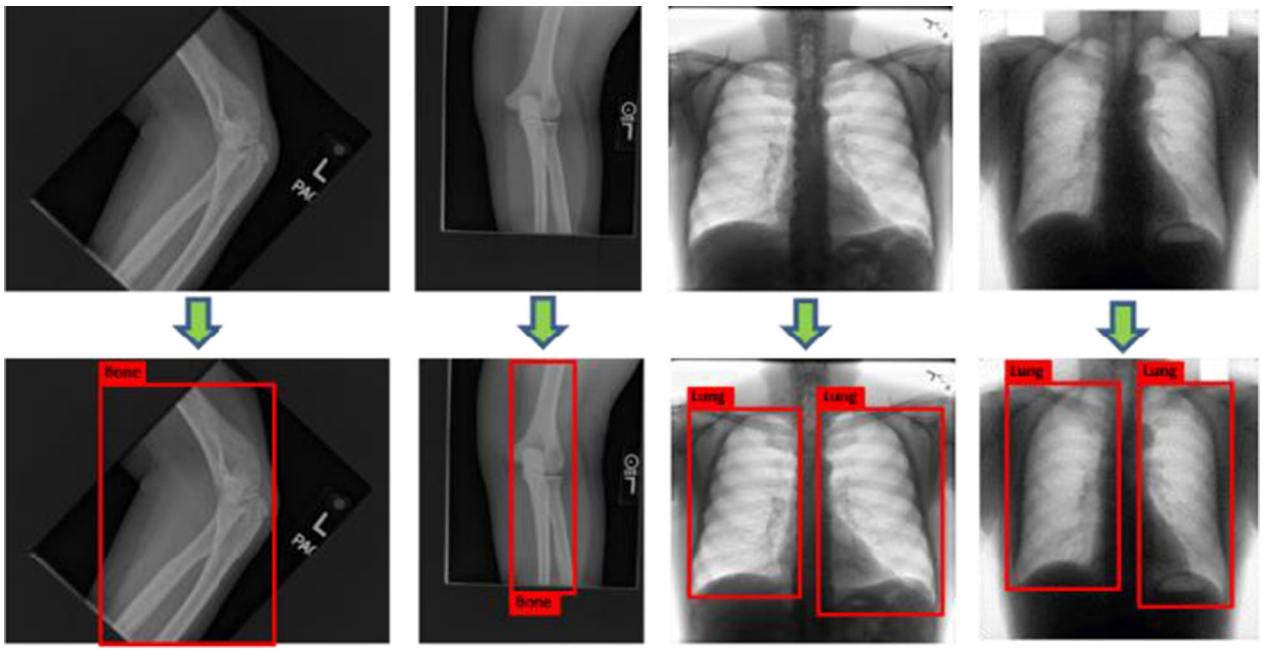


Fig. 5 Achievement of interest regions in 2D images

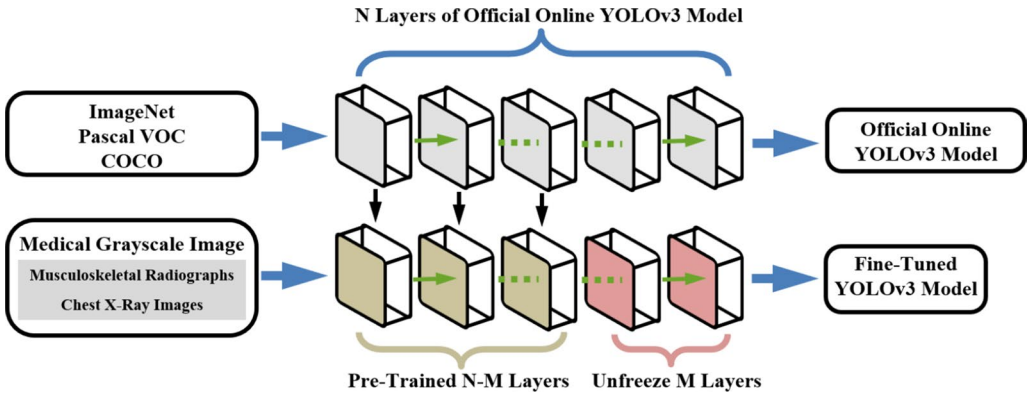


Fig. 6 The proposed 2D object detection network with fine-tuning method

Generation of proposal bounding box in pixel-features point cloud

The grayscale value of each pixel in interest regions represents their brightness [41]. Pixels compose the same tissues in particular image always share the grayscale value ranges and we could recognize them manually. All values range from 0 to 255 (Typically zero is taken to be black, and 255 is taken to be white). Darker pixels represent structures like soft tissues having less attenuation to the beam, while light ones represent structures like bones having high attenuation. Due to the lack of detailed gray values of pixels displayed on 2D images, it is hard to determinate their specific grayscale value ranges.

Thus, we turn pixels in 2D interest regions into the 3D representations as Fig. 7 shown. In Fig. 7. the first two dimensions represent pixels locations and the third dimension represents their grayscale values. The 3D data could be considered as the pixel-features point cloud and it is distinct and intuitive to obtain points which represent pixels belong to the same tissues. This helps us translate the 2D image segmentation task into the 3D object detection with point cloud. We only need to determine locations and widths of 3D bounding boxes which contain the foreground points during

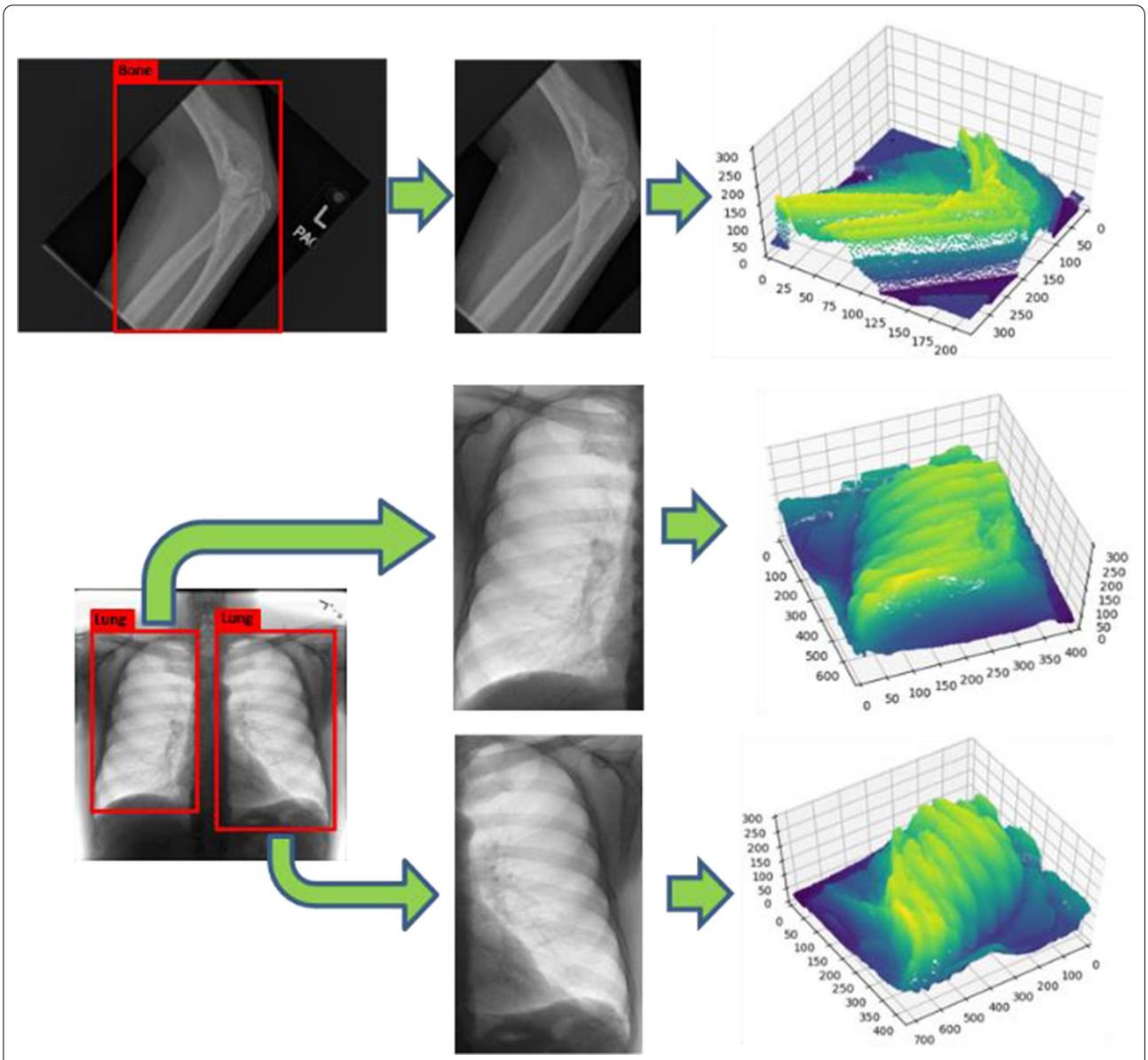


Fig. 7 Turning pixels in interest regions into the pixel-features point cloud

the object detection. Then bottoms and tops of bounding boxes could represent the segmentation required threshold values for 2D images.

Inspired by two-stage 2D object detection methods, we present a novel two-stage 3D object detection method, which is operated on pixel-features point cloud. In the first stage of existing popular two-stage 2D object detection method, the proposal bounding boxes with its classification scores are generated with convolutional neural network and the refinements of those boxes are obtained in the following stage after the Non-Maximum Suppression (NMS). While in our proposed 3D object detection method, based on two-stage strategy, the proposal 3D bounding boxes with the classification scores of points inside them are estimated firstly and these proposals are refined with regression in second stage.

The generation of proposal bounding boxes in pixel-features point cloud has three modules. As shown in Fig. 8, These modules include localization of anchor boxes, classification of points inside boxes utilizing PointNet [36] as backbone network and Non-Maximum Suppression with 3D Intersection-over-Union (IoU).

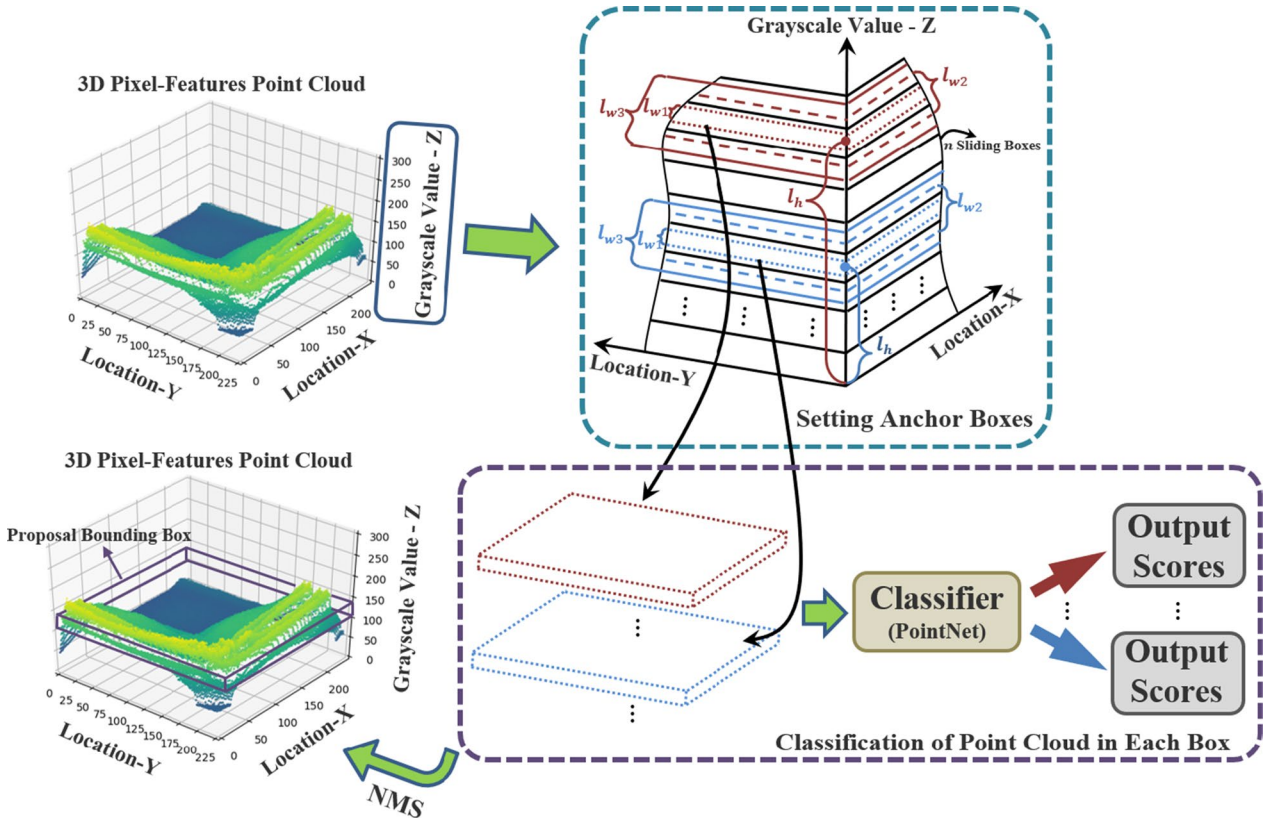


Fig. 8 The generation of proposal bounding boxes in pixel-features point cloud

Anchor boxes Proposal bounding boxes generation takes the $l_x \times l_y \times 255$ point cloud representation as input where l_x and l_y respectively indicate the length and width of 2D interest region. In order to avoid high overlap rate of predict boxes and the low search efficiency using selective search as Region Convolutional Neural Network (RCNN) method, inspired by the Region Proposal Networks (RPN) in Faster RCNN, we apply the anchor boxes method for electing predict boxes.

To generate proposals, we slide a small network over the input by a shared 3D convolutional layer referred to RPN and Single Shot MultiBox Detector (SSD) method as Fig. 8 shown. At each sliding-box location, we could predict multiple proposals simultaneously, and we denote the maximum number of possible proposals as k. These proposals are parameterized relative to k 3D anchor boxes. Each anchor is centered at its corresponding sliding box and is associated with a scale. Each anchor is defined with coordinates (l_h, l_w) where l_h and l_w represent its location and scale. We apply 3 scales by default, deciding $k = 3$ anchors at each sliding box and $n \times k$ anchors in total.

Classification of point cloud Anchor boxes with different scales share the same box-length l_x and box-width l_y, and they are distinguished by their center locations and box-heights. In order to determine the proposal bounding box from numerous anchor boxes, we utilize the PointNet as our backbone network and apply the fine-tuning method for training our classification module.

The classification network in Fig. 8 indicates that raw point clouds are directly taken as the input and each point is processed independently at the initial stage. Due to point clouds could be easily applied rigid or affine transformations, input points are sorted into a canonical order with the first affine transformation by a mini-net (T-net) and moreover, after points features extraction with multi-layer perceptron (mlp), features from different points could also be aligned using another alignment network by feature transformation matrix. Then, the max pooling layer aggregates all points features extracted from the second mlp and outputs the global features. The final fully connected layers set the global feature as input and outputs k scores for all the k candidate classes.

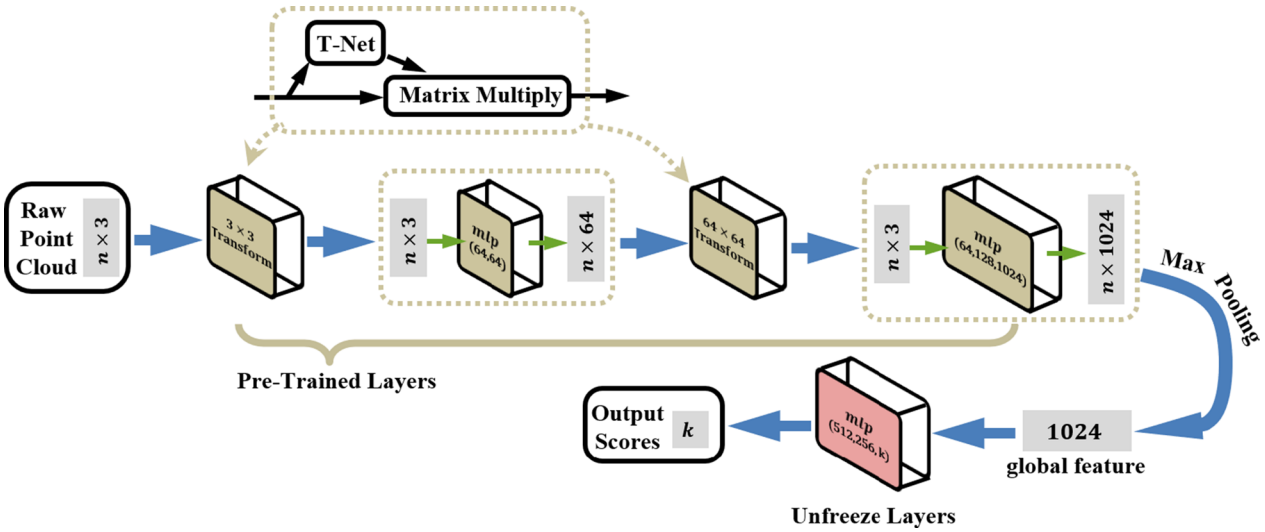


Fig. 9 The proposed point cloud classification network with fine-tuning method

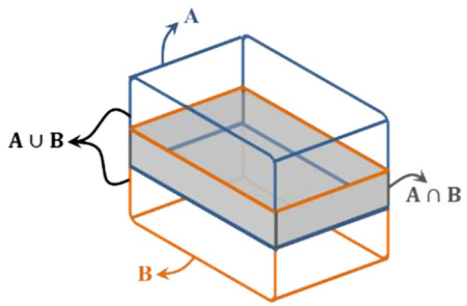


Fig. 10 IoU computation for 3D. The intersection volume is highlighted in gray

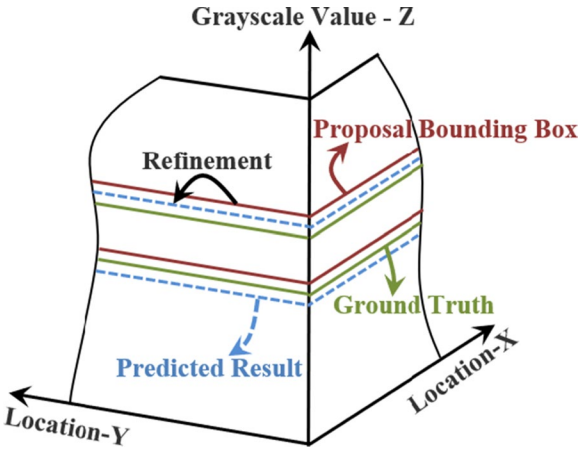


Fig. 11 Refinement of proposal bounding box

It should be noted that models-based point clouds datasets which mapped from grayscale medical images is scarce, thus we apply the fine-tuning method again. With the migration of PointNet model pretrained on ModelNet40 [42], we freeze most layers of the network except the final fully connected layers as shown in Fig. 9.

NMS with 3D IoU After the above module, the classification results of point cloud in each anchor box could be achieved with scores. But as many 2D object detection method, there exists some repeated proposals of one object. They belong to the same candidate class and overlap with the local highest-score box. For reducing the redundancy, we adopt the non-maximum NMS on these proposals with 3D intersection over union (3D IoU). Different from the IoU computation for 2D based on the relationships of areas between box A and B [43], like Fig. 10 shows, volumes of two boxes are applied for 3D IoU calculation [44] which could be formulated as:

$$\text{3D IoU}(A,B) = \frac{A_v \cap B_v}{A_v \cup B_v} = \frac{A_v \cap B_v}{|A_v| + |B_v| - A_v \cap B_v} \tag{1}$$

Through the setting of 3D IoU threshold for NMS and ranking with classification scores, it remains only one box for each candidate class which could be considered as the proposal bounding box.

Refinement of proposal bounding box

Even though high classification scores of the proposal bounding boxes, the location and scale errors between

them and ground truth exist. We train and implement a class-specific bounding box linear regression model to reduce errors and improve detection performance.

On the assumption that we achieve one proposal bounding box P^i and its nearby ground-truth box G^i as shown in Fig. 11, where $P^i = (P^i_{l_h}, P^i_{l_w})$ specifies height l_h of the center of proposal bounding box together with its width l_w. Meanwhile, the ground-truth bounding box G^i is specified in the same way: $G^i = (G^i_{l_h}, G^i_{l_w})$. The goal of the bounding box regressor is to learn a transformation which could map each proposal bounding box P to the ground-truth box G.

The transformation could be parameterized in terms of two functions $d_{l_h}(P)$ and $d_{l_w}(P)$. The first function specifies the translation of bounding box P's center which is scale-invariant, while the second specifies the log-space translation of its width. By applying the transformation as following equations, an input proposal bounding box P could be transformed into a predicted ground-truth box $\widehat{G}$.

$$\hat{G}_{l_h} = P_{l_w} \times d_{l_h}(P) + P_{l_h} \tag{2}$$

$$\hat{G}_{l_w} = P_{l_w} \times \exp\left(d_{l_w}(P)\right) \tag{3}$$

where exp is the natural exponential function.

Inspired by the 2D object detection, the bounding box regression of our method is performed on global features which is max pooled from PointNet model. Above two functions $d_{l_h}(P)$ and $d_{l_w}(P)$ could be modeled as linear functions of the global features of proposal bounding box P, denoted as $f_{mp}(P)$. Therefore, we have $d_*(P) = \mathrm{T}_* \times f_{mp}(P)$, where $*$ represents l_h or l_w, and T_* is a vector composed of learnable model parameters.

The transformation targets t_* between proposal bounding box P and the real ground-truth box G could be defined as:

$$t_{l_h} = \frac{G_{l_h} - P_{l_h}}{P_{l_w}} \tag{4}$$

$$t_{l_w} = \log\left(\frac{G_{l_w}}{P_{l_w}}\right) \tag{5}$$

Thus, after setting the loss function and by optimizing the regularized least squares objective as following, we could learn T_* and achieve the transformation to refine the proposal bounding box.

$$\boldsymbol{Loss} = \sum_i^N \left(t^i_* - \hat{\mathrm{T}}_* \times f_{mp}\left(P^i\right)\right)^2 \tag{6}$$

$$\mathrm{T}_* = \mathrm{argmin}_{\hat{\mathrm{T}}_*} \boldsymbol{Loss} + \lambda \hat{\mathrm{T}}^2_* \tag{7}$$

where argmin means T_* depends on the minimum of ***Loss***.

Obtaining segmentation results

As shown in Fig. 3, 3D bounding boxes in space could represent 3D positions range of pixel-feature points among them. Since 3D point cloud is mapped from pixels in 2D images according to positions and grayscale values, 3D bounding boxes could also present positions range and grayscale values range of 2D pixels corresponding to 3D points among boxes. After the refinement, 3D bounding boxes with accurate height and weight could be achieved. The top and the bottom of refined boxes represent the thresholding values for segmentation, while the front, back, left and right of boxes describe regions for segmentation. By remapping points among refined 3D bounding boxes to pixels in 2D images, these 2D pixels could compose segmentation results.

Training strategy

The proposed grayscale medical image segmentation method is based on 2D and 3D object detection models. Transfer learning and piecewise learning rate are applied for object detection models training. In the training of 2D object detection model, YOLOv3 which is trained on datasets including ImageNet, Pascal VOC and MS COCO is selected as the pretrained model. In the first stage, all but last 3 layers are frozen and the model is trained with prepared datasets including musculoskeletal and chest radiographs in the learning rate as 0.001 for 25 epochs. In the second stage, all layers of the network are unfrozen and they are trained in the learning rate as 0.0001 for 25 epochs. It takes 1.75 h to train the 2D object detection model. 3D object detection is composed of point cloud classifier and 3D bounding box regressor. In the training of point cloud classifier, PointNet which is trained on ModelNet40 is chosen as the pretrained model. In the first stage, all layers except last mlp module are frozen and the model is trained with point could datasets of bone and chest in the learning rate as 0.001 for 100 epochs. Then, all layers are unfrozen and the model is trained in the learning rate as 0.0001 for 100 epochs in the second stage. It takes 2.25 h to train the point cloud classifier; In the training of 3D bounding box regressor, due to the simple model and the requirement of bounding box refinement, the model is trained with one-stage strategy in a small learning rate as 0.0001 for 200 epochs and the training takes 0.25 h. Besides, Adam is selected as the optimizer for training of 2D object detection model and point cloud classifier, while Stochastic Gradient

Descent is chosen as the optimizer for 3D bounding box regressor training.

Performance assessment

In this study, we evaluate the segmentation performance by following four metrics: Dice similarity coefficient (DSC) scores [6], intersection over union (IoU), False negative (FN) and False positive (FP) [7]. Ranges of DSC and IoU are between 0 and 1, higher values of them and lower values of FN and FP indicate the higher accuracy. The calculation formula of DSC is defined as:

$$\mathrm{DSC} = \frac{2|T \cap G|}{|T| + |G|} \tag{8}$$

where T is the detected region and G is the ground truth region.

Results

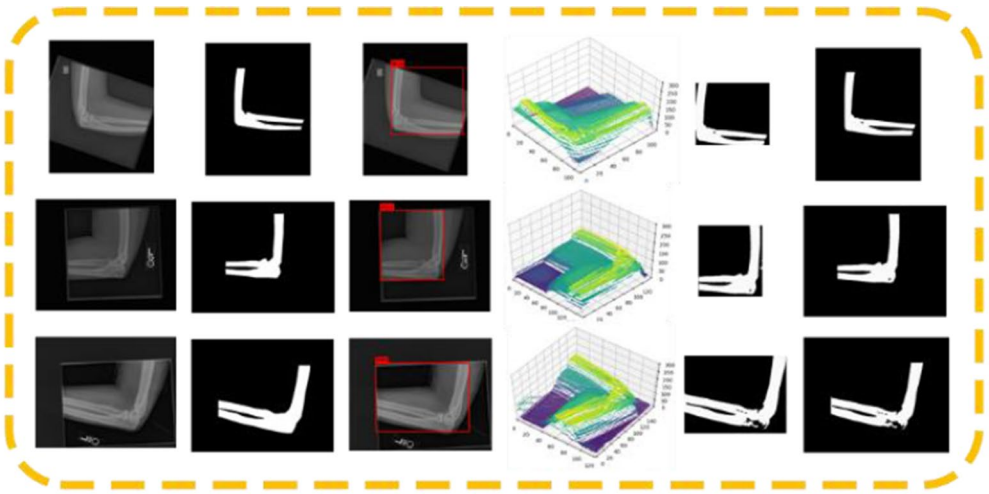
Our model is implemented with Pytorch [45] and its entire training process is performed on a computer with Windows 10 operating system, Intel Core i7 processor with 3.0 GHz, 64 GB of RAM and a single NVIDIA GPU (Quadro RTX 4000).

After training process, by applying the proposed method with the given grayscale medical images input and following the method pipeline as Fig. 3. shown, regions of target issues could be segmented. Each block in Fig. 12. presents several examples of segmentation performance from different kinds of datasets, as well as processing results after each stage, where white represents true positive pixels and black is for true negatives pixels. Moreover, according to evaluation criteria, Table 1 shows four metrics including IoU, DSC, FN and FP to assess the segmentation performance of images in different datasets.

As shown in Fig. 12 and Table 1, we could obtain high IoU and DSC scores with satisficed segmentation results on different datasets. This indicates that based on the proposed method, 2D interest regions and 3D bounding boxes containing target pixel-features point cloud during the processing could be successfully achieved.

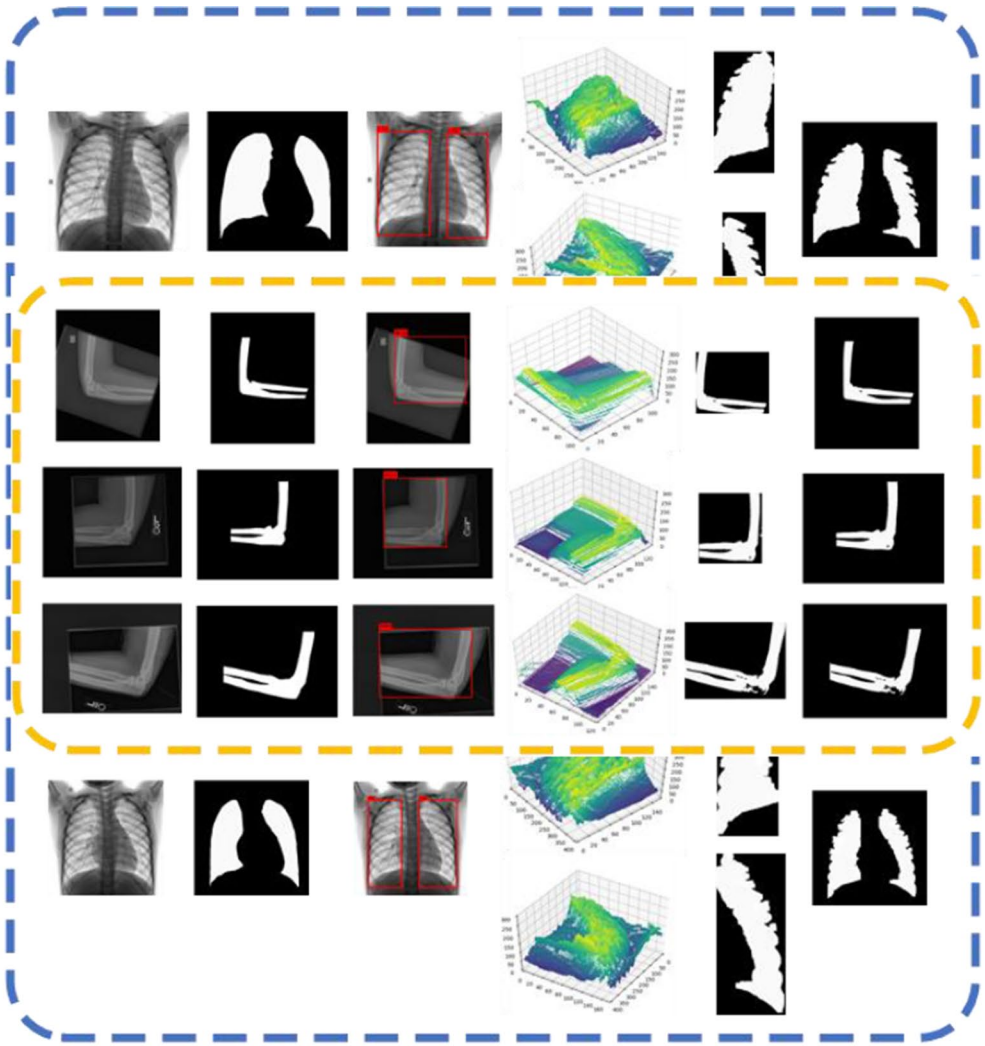
Discussion

In this section, we compare the image segmentation performance of the proposed method with multiple famous and clinically performed well models. As well known, CNN based models are among the most successful and widely used for medical image processing. Besides the milestone FCN model, UNet built on top of the fully convolutional networks with a U-shaped architecture to capture context information, and based on it, Res-UNet [46] improved the segmentation results using residual blocks as the building block and UNet++ [47] enhanced segmentation quality of varying-size objects. Also, Attentiom UNet [48] achieved the better performance with the attention gate. We train these models in the same dataset as our proposed method and Table 2 presents the comparison results. Meanwhile, Fig. 13. shows results of each case in Fig. 12. with different methods by visualization. It indicates that compared with other models, our proposed

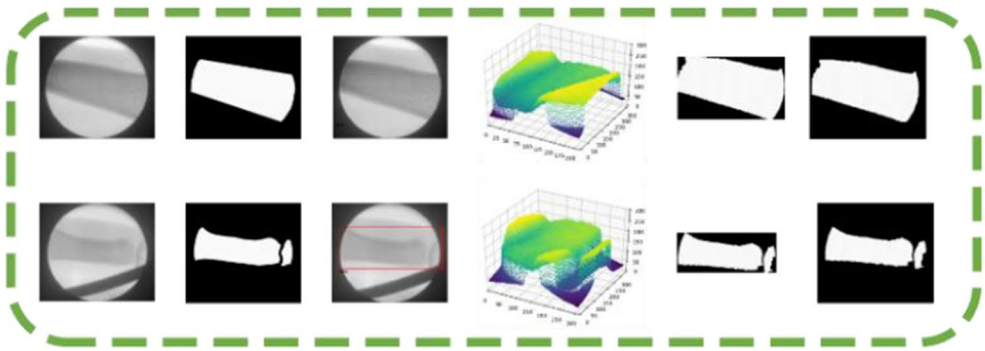


(a) Examples of segmentation performance in musculoskeletal radiographs dataset.

Fig. 12 Segmentation results from different kinds of datasets. From the first to the last column are origin images, ground truth, achievements of interest regions, representations of pixel-feature point cloud, local segmentation results, and segmentation results in original image size, respectively



(b) Examples of segmentation performance in chest radiographs dataset.



(c) Examples of segmentation performance in X ray images with the portable X ray machine.

Fig. 12 continued

Table 1 The values of evaluative metrics from experiments in different datasets

Datasets	IoU	DSC	FN	FP
Musculoskeletal radiographs	0.92	0.96	0.05	0.02
Chest radiographs	0.88	0.93	0.11	0.15
Images from X-ray machine	0.94	0.94	0.06	0.08

approach improves the segmentation performance and it obtains the highest IoU scores of 0.92, 0.88 and 0.94 with three datasets respectively. In our approach, 2D and 3D object detection models could be both trained with transfer learning method which makes it possible to achieve a quite accurate image segmentation model with small training datasets. While other sematic segmentation methods may be sensitive to the scale of datasets because the pre-trained model could only help simplify the downsample training procedure, and the training of upsample still requires a number of datasets. This indicates that it is impossible to adapt them for every application task well because training data is scare especially in medical image field. Moreover, in grayscale images, grayscale values of pixels are important features to distinguish different objects, and the intuitive logic of grayscale image segmentation could be considered as the collection of pixels with similar grayscale values. So, the proposed image segmentation model which obtains the purpose ranges of grayscale values with 3D object detection have better explicability and segmentation effect.

Under different medical imaging devices and environment in clinical, ranges of grayscale values of pixels which compose the same segmentation target in different medical images are always different. But our proposed method could settle this and we could obtain thresholding values (top and bottom of 3D bounding boxes) by mapping pixels in 2D images into 3D point clouds and adopting 3D object detection with features of pixels.

Conclusions

In this paper, we present a new grayscale medical image segmentation method with object detection models. In this method, 2D object detection model is applied to achieve interest regions of segmentation objects. After mapping 2D pixels in interest regions to 3D point cloud

Table 2 Comparison between segmentation performance (IoU) of the proposed approach with other methods

Datasets	Proposed	FCN	UNet	UNet++	Res-UNet	Attention Unet
Musculoskeletal radiographs	0.92	0.82	0.85	0.84	0.91	0.90
Chest radiographs	0.88	0.76	0.81	0.83	0.88	0.86
Images from X-ray machine	0.94	0.72	0.82	0.87	0.85	0.91

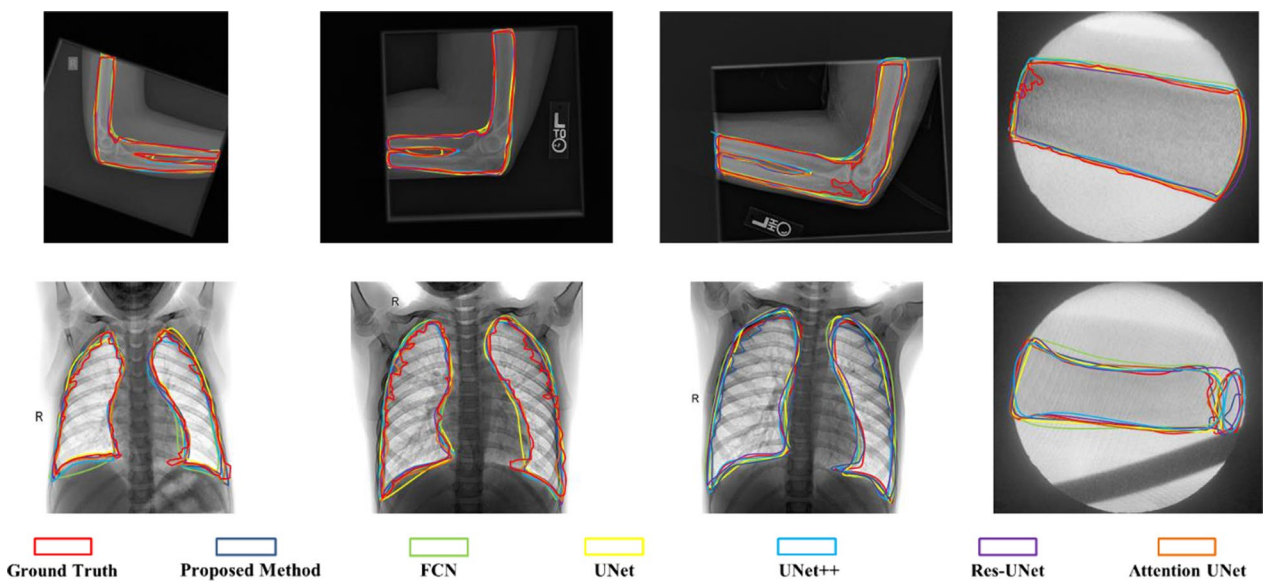


Fig. 13 Comparisons of segmentation performance on each case in Fig. 12. with different methods

according to their positions and grayscale values, 3D object detection model is adopted to obtain bounding boxes containing target pixels-feature points. After projecting these points to 2D images, they could composite segmentation results. Experiments results prove the better effectiveness and accuracy of our method than the other compared models. In clinical applications, more than improving segmenting performance with bone and chest X-ray images, the proposed segmentation method could also be carried over for other kinds of grayscale medical images in diagnosis efficiently and conveniently. This is because two object detection models could be trained separately with little labeling cost based on transfer learning method. Besides, pretrained models for both 2D and 3D object detection in our method could be changed and upgraded flexibly for further accurate segmentation results.

Abbreviations

CNN(s): Convolutional neural network(s); FCN: Fully convolutional networks; MURA: Musculoskeletal radiographs; LERA: Lower extremity radiographs; CheXpert: Chest radiography; GT: Ground truth; YOLO:: You only look once; PASCAL VOC: Pattern analysis, statistical modeling and computational learning visual object classes; MS COCO: Microsoft common objects in context; NMS: Non-maximum suppression; IoU: Intersection-over-Union; RCNN: Region convolutional neural network; RPN: Region proposal networks; SSD: Single shot multiBox detector; mlp: Multi-layer perceptron; DSC: Dice similarity coefficient; FN: False negative; FP: False positive.

Acknowledgements

Not applicable.

Authors' contributions

QZ conceived the research. YG and YS analyzed the clinical and imaging data. YS, YG, and XW designed the study. YG and YS performed the experiments and collected the results. YG and YS drafted the manuscript. QZ reviewed the final manuscript. All authors read and approved the final manuscript.

Funding

This work was supported by the project of Tongji University Sheng Feiyun College Student Science and Technology Innovation Practice Found.

Availability of data and materials

Musculoskeletal radiographs and chest radiographs which support our research are available from Stanford ML Group. But restrictions apply to the availability of these data, which were used under license for the current study, and so are not publicly available. Data are however available from the authors upon reasonable and with permission of Stanford ML Group. While phalanx and forearm X-ray images are available only upon request by emailing authors due to the ethical restrictions on sharing these data which could contain potentially sensitive information of patients.

Declarations

Ethics approval and consent to participate

We declare that all of us obey the principles of the Declaration of Helsinki. In other words, all experiments and methods in this paper are in accordance with these principles. The study was approved by the Ethics Committee of the First people's Hospital of Yancheng ([2021]-(K-54)). The fully anonymized phalanx and forearm X-ray images were received by authors on 2 April, 2021 and the requirement for informed consent was waived for this study by the Ethics Committee of the First people's Hospital of Yancheng ([2021]-(K-54)) because of the anonymous nature of the data.

Consent for publication

Not applicable for this paper.

Competing interests

All authors declare that they have no competing interests.

Author details

[1]School of Mechanical Engineering, Tongji University, Shanghai, China. [2]Department of Orthopaedics, The First People's Hospital of Yancheng, Yancheng, China. [3]Shanghai Bojin Electric Instrument and Device Co., Ltd, Shanghai, China.

Received: 26 October 2021 Accepted: 22 February 2022
Published online: 27 February 2022

References

1. Wallyn J, Nicolas A, Salman A, et al. Biomedical imaging: principles, technologies, clinical aspects, contrast agents, limitations and future trends in nanomedicines. Pharm Res. 2019;36(6):78–108.
2. Yeo WK, Yap DFW, et al. Grayscale medical image compression using feedforward neural networks. In: 2011 IEEE international conference on computer applications and industrial electronics (ICCAIE). 2011. p. 633–8.
3. Lei T, et al. Medical image segmentation using deep learning: a survey. arXiv. 2020. p. 13120.
4. Rathnayaka K, Sahama T, Schuetz MA, et al. Effects of CT image segmentation methods on the accuracy of long bone 3D reconstructions. Med Eng Phys. 2011;33(2):226–33.
5. Wang S, Zhou Mu, Zaiyi L, et al. Central focused convolutional neural networks: developing a data-driven model for lung nodule segmentation. Med Image Anal. 2017;40:172–83.
6. Liu H, Lei W, Yandong N, et al. SDFN: segmentation-based deep fusion network for thoracic disease classification in chest X ray images. Comput Med Imaging Graph. 2019;75:66–73.
7. de Albuquerque VHC, Rodrigues DA, Ivo RF, et al. Fast fully automatic heart fat segmentation in computed tomography datasets. Comput Med Imaging Graph. 2020;80:101674.
8. Wen Li, et al. Automatic segmentation of liver tumor in CT images with deep convolutional neural networks. J Comput Commun. 2015;3(11):146.
9. Vivanti R, Ephrat A, Joskowicz L, et al. Automatic liver tumor segmentation in follow-up CT studies using convolutional neural networks. In: Proceedings of the patch-based methods in medical image processing workshop, vol 2. 2015. p. 2.
10. Mansour R F, Escorcia-Gutierrez J, Gamarra M, et al. Artificial intelligence with big data analytics-based brain intracranial hemorrhage e-diagnosis using CT images. Neural Comput Appl. 2021; 1–13.
11. Mansour RF, Aljehane NO. An optimal segmentation with deep learning based inception network model for intracranial hemorrhage diagnosis. Neural Comput Appl. 2021;33:13831–43.
12. Masood S, Muhammad S, Afifa M, et al. A survey on medical image segmentation. Curr Med Imaging. 2015;11(1):3–14.
13. Khandare ST, Isalkar AD. A survey paper on image segmentation with thresholding. Int J Comput Sci Mob Comput. 2014;3(1):441–6.
14. Sezgin M, Sankur B. Survey over image thresholding techniques and quantitative performance evaluation. J Electron Imaging. 2004;13(1):146–65.
15. Maolood IY, Al-Salhi YEA, Lu S. Thresholding for medical image segmentation for cancer using fuzzy entropy with level set algorithm. Open Med. 2018;13(1):374–83.
16. Hao D, Qiuming Li, Chengwei Li. Histogram-based image segmentation using variational mode decomposition and correlation coefficients. SIViP. 2017;11(8):1411–8.
17. Long J, Shelhamer E, Darrell T. Fully convolutional networks for semantic segmentation. In: Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR). 2015. p. 3431–40.
18. Ronneberger O, Fischer P, Brox T. U-net: Convolutional networks for biomedical image segmentation. In: International conference on medical image computing and computer-assisted intervention (MICCAI). 2015. p. 234–41.

19. Chen LC, Papandreou G, Kokkinos I, et al. Deeplab: semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs. IEEE Trans Pattern Anal Mach Intell. 2017;40(4):834–48.
20. Deng J, Dong W, Socher R, et al. Imagenet: a large-scale hierarchical image database. In: Proceedings of the IEEE conference on computer vision and pattern recognition. 2009. p. 248–55.
21. Kalinin AA, Iglovikov VI, Rakhlin A, et al. Medical image segmentation using deep neural networks with pre-trained encoders. In: Arif Wani M, Kantardzic M, Sayed-Mouchaweh M, editors., et al., Deep learning applications. Springer; 2020. p. 39–52.
22. Conze P-H, Brochard S, Burdin V, et al. Healthy versus pathological learning transferability in shoulder muscle MRI segmentation using deep convolutional encoder-decoders. Comput Med Imaging Graph. 2020;83:101733.
23. Rajpurkar P, Irvin J, Bagul Aarti, et al. Mura: large dataset for abnormality detection in musculoskeletal radiographs. arXiv. 2017; 1712.06957.
24. LERA—Lower extremity radiographs. https://aimi.stanford.edu/lera-lower-extremity-radiographs-2.
25. Irvin J, Rajpurkar P, Ko M, et al. Chexpert: A large chest radiograph dataset with uncertainty labels and expert comparison. Proc AAAI Conf Artif Intell. 2019;33(01):590–7.
26. Cohen J-P, Morrison P, Dao L, et al. Covid-19 image data collection: prospective predictions are the future. arXiv. 2020; 2006.11988.
27. Jiao L, Zhang F, Liu F, et al. A survey of deep learning-based object detection. IEEE Access. 2019;7:128837–68.
28. Girshick R, Donahue J, Darrell T, et al. Rich feature hierarchies for accurate object detection and semantic segmentation. In: Proceedings of the IEEE conference on computer vision and pattern recognition. 2014. p. 580–7.
29. Redmon J, Divvala S, Girshick R, et al. You only look once: unified, real-time object detection. In: Proceedings of the IEEE conference on computer vision and pattern recognition. 2016. p. 779–88.
30. Liu W, Anguelov D, Erhan D, et al. SSD: single shot multibox detector. In: European conference on computer vision. 2016. p. 21–37.
31. Shin HC, Roth HR, Gao M, et al. Deep convolutional neural networks for computer-aided detection: CNN architectures, dataset characteristics and transfer learning. IEEE Trans Med Imaging. 2016;35(5):1285–98.
32. Qian R, Lai X, Li X. 3D object detection for autonomous driving: a survey. arXiv. 2021. 2106.10823.
33. Zhou Y, Tuzel O. Voxelnet: end-to-end learning for point cloud based 3d object detection. In: Proceedings of the IEEE conference on computer vision and pattern recognition. 2018. p. 4490–9.
34. Chen Y, Liu S, Shen X, et al. Fast point r-cnn. In: Proceedings of the IEEE/CVF international conference on computer vision. 2019. p. 9775–84.
35. Shi S, Wang X, Li H P. 3d object proposal generation and detection from point cloud. In: Proceedings of the IEEE conference on computer vision and pattern recognition, Long Beach, CA, USA. 2019. p. 16–20.
36. Qi CR, Su H, Mo K, et al. Pointnet: deep learning on point sets for 3d classification and segmentation. In: Proceedings of the IEEE conference on computer vision and pattern recognition. 2017. p. 652–60.
37. Redmon J, Farhadi A. Yolov3: an incremental improvement. arXiv. 2018; 1804.02767.
38. Rothe R, Guillaumin M, Van Gool L. Non-maximum suppression for object detection by passing messages between windows. In: Asian conference on computer vision. 2014. p. 290–306.
39. Everingham M, Van Gool L, Williams CK, et al. The pascal visual object classes (voc) challenge: a retrospective. Int J Comput Vis. 2014;111:98–136.
40. Lin T Y, Maire M, Belongie S, et al. Microsoft COCO: common objects in context. In: European conference on computer vision. 2014. p. 740–55.
41. Tan L, Jiang J. Digital signal processing: fundamentals and applications. Academic Press; 2019.
42. Wu Z, Song S, Khosla A, et al. 3d shapenets: a deep representation for volumetric shapes. In: Proceedings of the IEEE conference on computer vision and pattern recognition. 2015. p. 1912–20.
43. Rezatofighi H, Tsoi N, Gwak JY, et al. Generalized intersection over union: a metric and a loss for bounding box regression. In: Proceedings of the IEEE conference on computer vision and pattern recognition. 2019. p. 658–66.
44. Zhou D, Fang J, Song X, et al. Iou loss for 2d/3d object detection. In: International conference on 3D vision (3DV). 2019. p. 85–94.
45. Paszke A, Sam G, Francisco M, et al. Pytorch: an imperative style, high-performance deep learning library. Adv Neural Inf Process Syst. 2019;32:8026–37.
46. Xiao X, Lian S, Luo Z, Li S. Weighted Res-UNet for high-quality retina vessel segmentation. In: 2018 9th international conference on information technology in medicine and education (ITME). 2018. p. 327–31.
47. Zhou Z, Siddiquee MMR, Tajbakhsh N, Liang J. UNet++: redesigning skip connections to exploit multiscale features in image segmentation. IEEE Trans Med Imaging. 2020;39(6):1856–67.
48. Oktay O, Jo S, et al. Attention U-Net: learning where to look for the pancreas. arXiv. 2018; 1804.03999.

Publisher's Note