Published in: Information Fusion journal

The Version of Record is available online at: https://doi.org/10.1016/j.inffus.2023.102016

A multi-objective segmentation method for chest X-rays based on collaborative learning from multiple partially annotated datasets☆

Hongyu Wang [a,b,c], Dandan Zhang [d], Jun Feng [d,*], Lucia Cascone [e], Michele Nappi [e], Shaohua Wan [f,*]

[a] School of Computer Science and Technology, Xi'an University of Posts and Telecommunications, Xi'an Shaanxi 710121, China
[b] Shaanxi Key Laboratory of Network Data Analysis and Intelligent Processing, Xi'an University of Posts and Telecommunications, Xi'an Shaanxi 710121, China
[c] Xi'an Key Laboratory of Big Data and Intelligent Computing, Xi'an University of Posts and Telecommunications, Xi'an Shaanxi 710121, China
[d] Department of Information Science and Technology, Northwest University, Xi'an Shaanxi 7101127, China
[e] Department of Computer Science, University of Salerno, Salerno, Italy
[f] Shenzhen Institute for Advanced Study, University of Electronic Science an Technology of China, Shenzhen 518110, China

ARTICLE INFO

Keywords:
Chest X-rays
Multi-object segmentation
Partially annotated data
Medical image analysis

ABSTRACT

Accurate segmentation of multiple targets, such as ribs, clavicles, heart, and lung fields, from chest X-ray images is crucial for diagnosing various lung diseases. Currently, mainstream deep learning methods heavily rely on fully annotated large-scale datasets. However, annotating all objects in chest X-ray images is a labor-intensive and time-consuming task. The publicly available partially annotated chest X-ray datasets have varying annotation objects and standards, and they are seldom utilized comprehensively in existing studies. To address these challenges, we propose A Multi-objective Segmentation Method for chest X-rays based on Collaborative Learning from Multiple Partially Annotated Datasets (called: MSM-CLMPAD). Our approach first utilizes an encoder constructed with densely connected blocks to extract multi-scale features from multiple partially annotated datasets. Then, leveraging the overlapping relationships among segmentation targets, we design a dual decoder guided by the attention mechanism. The novel attention-guided decoder effectively disentangles the features corresponding to various targets. Importantly, we propose an alternating training strategy for different datasets, facilitating collaborative learning of the same network model across datasets with diverse annotation targets. The experimental results performing on four public datasets demonstrate that our method achieves superior Dice and Jaccard coefficients compared to other popular methods, particularly for overlapping targets and unclear regions. We also explore the mutual influence of different targets in chest X-rays, offering a solution for interaction among partial datasets and further alleviating the difficulties of data annotation for multi-organ segmentation tasks.

1. Introduction

In recent years, the incidence and mortality rates of lung diseases have shown a steady increase, attributed to factors like air pollution, environmental changes, and industrial development. Lung cancer, in particular, has reached an alarming incidence rate of 11.4% and a mortality rate of 18% [1]. Chest X-ray imaging has become the primary screening and diagnostic tool for lung diseases due to its simplicity, speed, low radiation exposure, and cost-effectiveness [2]. It provides radiologists with crucial information about organ size, shape, and location, including heart, lung, and ribs. Automatic segmentation of multiple objects from X-ray images plays a vital role in the diagnosis of various lung diseases, such as pulmonary nodules, tuberculosis, and COVID-19 [3].

Early researchers often used edge detection, template matching, statistical shape model, active contour, and other machine learning based methods to segment medical images [8]. But such methods need to manually set feature descriptors, and the segmentation performance of the algorithm depends on the empirical knowledge of experts, and

☆ This research was supported by the National Natural Science Foundation of China (NO.62001380, 62073260); Key & D projects in Shaanxi Province (No. 2023-YBSF-493, 2023-YBSF-455); Key Project of Shenzhen City Special Fund for Fundamental Research (No. JCYJ20220818103200002).

* Corresponding authors.
E-mail addresses: hywang@xupt.edu.cn (H. Wang), 202021313@stumail.nwu.edu.cn (D. Zhang), fengjun@nwu.edu.cn (J. Feng), lcascone@unisa.it (L. Cascone), mnappi@unisa.it (M. Nappi), shaohua.wan@uestc.edu.cn (S. Wan).

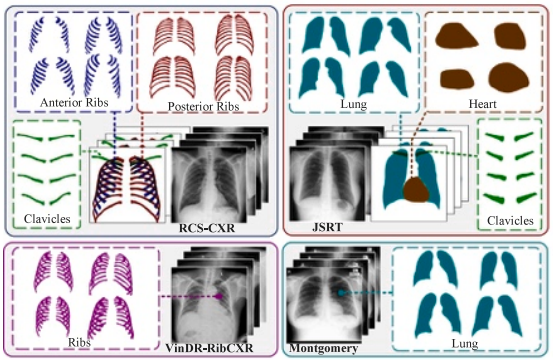

Fig. 1. Four chest X-ray partially annotated datasets, including RCS-CXR [4], JSRT [5], VinDR-RibCXR [6], Montgomery [7]. They have different annotation targets and data annotation standards.

the generalization performance is poor [9]. As deep learning technology advances rapidly, algorithms utilizing its strong feature extraction abilities have emerged as the predominant research focus in image segmentation [10,11]. However, these methods often require a large number of labeling samples with strong supervision [12]. And labeling multi-objects in chest X-ray images is a difficult, time consuming, and costly job. Consequently, obtaining sizable and reliable chest x-ray images for multi-object segmentation is a challenge task.

A significant bottleneck in the field of multi-target segmentation research is the limited sample size and mostly partial labeling of currently used datasets. It has become a practical research problem to effectively utilize these partially annotated data to improve segmentation accuracy and robustness. Fig. 1 illustrates the following challenges: (1) Partially annotated data: Most datasets only annotate specific objects, while other organs are labeled as background. This differs from the datasets for multi-object annotations in natural scene like PASCAL VOC. Learning comprehensive image features from these partially annotated data presents a key challenge. (2) Overlapping areas: Chest X-ray images exhibit overlapping regions, such as clavicle and rib, anterior and posterior ribs. Assigning multiple labels to pixels in these overlapping regions during segmentation significantly impacts accuracy and integrity. (3) Variations in imaging: Differences in imaging equipment, acquisition parameters, and institutional factors lead to variations in the images themselves. Additionally, datasets from different institutions may follow different labeling standards. For example, some datasets annotate complete ribs and distinguish between anterior and posterior ribs (eg. RCS-CXR), while others provide fewer rib annotations without such differentiation (eg. VinDr-RibCXR). These differences in datasets and labeling standards pose challenges for multi-object segmentation tasks.

At present, the prevailing methods for addressing the aforementioned challenges primarily involve partitioning the labeled dataset into multiple subsets and training separate segmentation networks on each subset. However, these methods bring about a significant increase in computational complexity and also present difficulties in fully and efficiently harnessing all partially labeled datasets, ultimately resulting in wastage of resources. Taking inspiration from multi-head networks, our study explores an alternative approach that involves training a single multi-object segmentation network using a fusion of partially labeled datasets. This methodology eliminates the need for additional computational resources. We propose a novel Multi-object Segmentation Method for chest X-rays based on Collaborative Learning from Multiple Partially Annotated Dataset, referred to as MSM-CLMPAD, as illustrated in Fig. 2. Our approach makes three primary contributions aimed at enhancing the segmentation performance of multi-objects in X-ray images:

(1) To the best of our knowledge, this is the first multi-organ segmentation framework for chest X-rays that is trained with multiple partially annotated datasets. It distinguishes itself by employing collaborative learning across multiple partially annotated datasets, which eliminates the strict requirement for fully annotated objects and a unified labeling standard across different datasets.
(2) We construct a shared encoder using densely connected blocks. Unlike existing methods, it incorporates a synergistic attention skip connection module and an attention-guided multi-scale feature selection module. This combination enables us to capture spatial details and rich contextual information. By leveraging these learned hierarchical features, we significantly enhance the robustness of the multi-object segmentation network.
(3) To achieve accurate segmentation of overlapping targets, we have designed an attention oriented dual decoder structure. This innovative approach addresses the challenge of low segmentation accuracy in overlapping areas, which is a limitation of traditional methods. Our deep supervised dual decoder allows for focused attention on the distinct characteristics of different targets.

We employ an alternating training strategy to incorporate multiple partially labeled datasets into the training of a single network. This strategy effectively addresses the challenge of multi-organ segmentation by integrating diverse partially annotated datasets within a unified framework. In the experimental section, we specifically investigate the interaction among different segmentation targets, a aspect that is often overlooked in existing methods.

2. Related work

2.1. Multi-object segmentation in X-ray images

(1) Rib and clavicle. Convolutional Neural Networks (CNNs) are the predominant approaches for rib and clavicle segmentation. For instance, Liu et al. [13] employed a fully convolutional network with a weighted loss function to optimize rib and clavicle segmentation. This specifically addresses the challenge of accurately segmenting vague edges. Wang et al. [4] improved feature reuse by incorporating densely connected blocks, alleviating the issue of limited feature extraction caused by small dataset sizes. Oliveira et al. [14] leveraged domain adaptation techniques to extract skeletal information from CT images and facilitate rib segmentation in chest X-rays without relying solely on rib labels. Nguyen et al. [6] enhanced the UNet++ architecture by utilizing EfficientNet-B0 [15] as the encoder, achieving promising results in rib segmentation. However, these methods ignore the semantic gap that exists between the encoder and decoder, assuming equal contribution from all pixels in the skip connections. The inclusion of irrelevant pixels, such as background and noise, can potentially hinder the network's segmentation performance. Challenges such as overlapping targets, low contrast, and limited annotations impact the progress of deep learning-based rib and clavicle segmentation in chest X-rays.

(2) Lung field, heart and clavicle. Deep learning-based methods for single-object or multi-object segmentation of lung fields, hearts, and clavicles have made significant advancements in recent years. Wang et al. [16] proposed a network based on Mask R-CNN [17] for segmenting lung fields, hearts, and clavicles. However, this method incurs high computational costs due to excessive region proposals. Kholiavchenko et al. [18] improved segmentation performance at the boundaries of lung fields, hearts, and clavicles by simultaneously supervising the network using masks and mask contours for each target. Peng et al. [19] proposed a two-stage lung field segmentation method, where post-processing played a crucial role in refining the output results. Pal

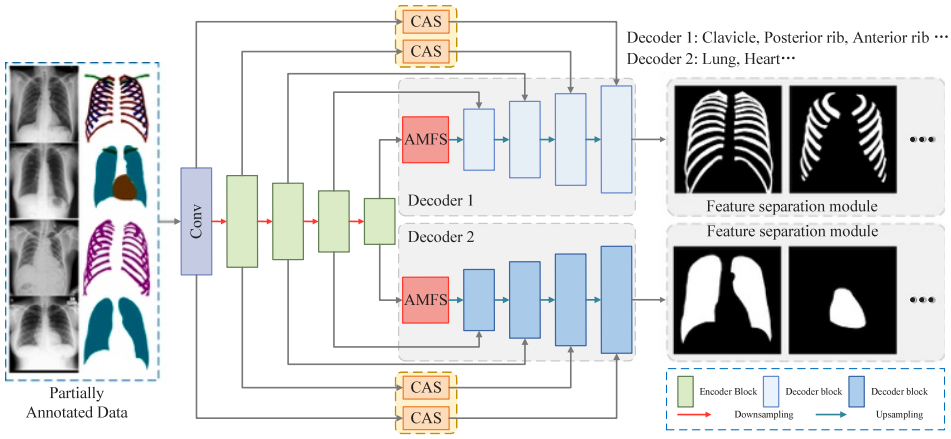

Fig. 2. The MSM-CLMPAD framework is designed to process partially annotated data. It consists of an encoder and two decoders. The encoder extracts hierarchical features that the decoders utilize to precisely segment various targets, including ribs, clavicles, lung fields, and hearts.

et al. [20] enhanced the UNet network [21] and incorporated attention gates to improve the segmentation performance of small targets such as clavicles and tracheas. Ullah et al. [22] introduced a dual encoder–decoder structure, where the first network generated coarse segmentation results and the second network refined the segmentation effects for smaller regions. While existing research achieves high accuracy in segmenting large, simple, and high-contrast targets like lung fields, it often exhibits poor segmentation performance for smaller targets such as clavicles.

2.2. Multi-objects segmentation from partially annotated datasets

Data annotation is a costly and scarce resource in medical image analysis, leading to a limited availability of fully annotated datasets. Furthermore, different research objectives may result in variations in annotated regions of interest for the same imaging technique. To effectively utilize existing annotated data, Petit et al. [23] introduced pseudo-labels into the training set and implemented an iterative re-labeling strategy using a self-supervised mechanism to address missing organ annotations in partially annotated datasets. However, the introduction of inaccurate pseudo-labels may have a negative impact on network performance. Fang et al. [24] introduced a feature extraction network with a pyramid input and pyramid output for multi-organ segmentation in abdominal CT. Their approach treats unknown labels as background, employs a target adaptive loss function, and utilizes a unified training strategy across multiple partially annotated datasets. Shi et al. [25] merged unlabeled organs with the background and imposed exclusive constraints on each voxel to train a multi-organ segmentation network using partially annotated datasets. Zhang et al. [26] proposed a dynamic on-demand network for multi-organ and tumor segmentation on partially annotated datasets, employing dynamic convolutional filters to control parameter changes in the dynamic head and to segment specific tumors or organs in abdominal CT images based on the task. Zhang et al. [27] proposed conditional nnUNet, which incorporates task-related information into the decoder to adjust the segmentation of different organs.

In summary, these studies have demonstrated successful results in organ segmentation tasks by treating the partial labeling problem as a multi-class segmentation task and assigning unlabeled organs to the background. However, these approaches are not suitable for chest X-ray images due to the overlapping nature of anatomical structures and the potential for a single pixel to belong to multiple targets simultaneously. Moreover, X-ray projection complexities and lung–rib overlap hinder single-decoder segmentation. Our method addresses these by using a single-encoder multi-decoder approach and collaborative learning. This enhances accuracy and generalization, effectively managing target differences and overlaps for precise multi-object segmentation.

3. Materials and method

3.1. Overall framework of the proposed method

To achieve comprehensive learning from multiple partially labeled datasets within a single network, we propose a novel Multi-object Segmentation Method for chest X-rays based on Collaborative Learning from Multiple Partially Annotated Datasets (MSM-CLMPAD). As depicted in Fig. 2, the framework of MSM-CLMPAD mainly comprises an encoder for shared convolution and two decoders for specific segmentation tasks.

The encoder, depicted as the green boxed branch in Fig. 2, plays a crucial role in extracting reliable feature information from multiple partially annotated datasets and learning the association relationships between different tasks. The encoder's parameters remain consistent across different datasets, eliminating the need to train separate encoders for each dataset. This approach not only saves training time but also preserves computational resources efficiently. To optimize image features, we introduce a co-attentive connection module into the encoder, which is designed with densely connected blocks. This innovative module effectively enhances the representation and optimization of image features. Subsequently, we design dual decoders for specific segmentation tasks. In these decoders, an attention-guided multi-scale feature selection module is adopted to process features from different branches, thereby facilitating the separation of multiple targets. Specifically, Decoder 1 is dedicated to handling segmentation tasks of targets that have more overlap with the lung field, such as the rib and clavicle. Decoder 2 is intended to handle segmentation tasks of the lung field or targets that have less overlap with the lung field, such as the lung field, heart, and so on.

3.2. Encoder based on dense connection and residual block

Despite the similarities in the appearance of the original images across various partially annotated chest X-ray datasets, discrepancies inevitably persist due to diverse collection conditions. For instance, the JSRT dataset, which comprises scanned chest X-ray images, exhibits pronounced intensity at the chest boundary. However, the internal intensity diminishes, resulting in lower contrast. Moreover, the visibility of the rib is inferior compared to that in the RCS-CXR dataset.

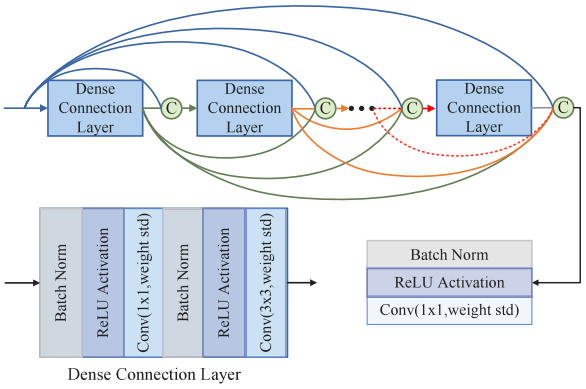

Fig. 3. Diagram of the dense connection block structure, consisting of multiple dense connection layers.

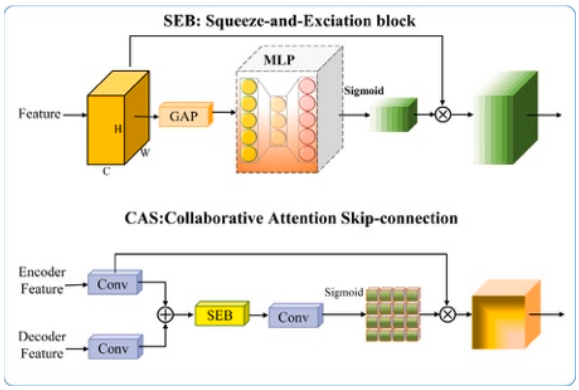

Fig. 4. Collaborative attention skip-connection module. In the SEB module, we perform global average pooling (GAP) on the feature maps. Then, a small Multi-Layer Perceptron (MLP) is used to learn the importance of each channel by capturing the interdependencies between channels.

Addressing this challenge necessitates a robust approach to feature extraction that is capable of capturing reliable information from different partially annotated datasets. Therefore, in this study, we employ dense connection blocks [28], renowned for their superior feature extraction capabilities, as the cornerstone of the encoder. These dense connection blocks empower the encoder to extract high-quality and reusable features from different datasets, while effectively mitigating the impact of inter-dataset variations. It ensures the flow of relevant information, enhancing the encoder's capacity to handle diverse partially annotated datasets and significantly improving performance in multi-target segmentation tasks.

As depicted in Fig. 3, the structure of a dense connection block comprises several dense connection layers. A key characteristic of these blocks is that each subsequent dense connection layer receives the outputs from all preceding layers as inputs, promoting feature exchange and preservation of the gradient during training. Ultimately, to reduce the channel size, the architecture includes a batch normalization layer, an activation layer, and a 1×1 convolutional layer that incorporates weighted standardization. This thoughtfully designed, densely connected block structure ensures the extraction of high-quality, reusable features across datasets, regardless of their differing collection conditions. Through this strategy, our model demonstrates improved performance and generalizability, significantly advancing the field of chest X-ray analysis.

As shown in Fig. 2, the encoder comprises five layers in total. Except for the first layer, which employs two 3×3 convolutional blocks, the remaining four layers are constructed using densely connected blocks, as depicted in Fig. 3. These densely connected blocks incorporate varying numbers of dense connection layers: 6, 12, 24, and 16, respectively. Between each layer in the encoder, max pooling is deployed for downsampling, aiming to capture semantic information at varying scales. To effectively preserve the image boundary and spatial detail information related to the segmentation area, a Collaborative Attention Skip-connection (CAS) module is designed following the encoder. As shown in the lower part of Fig. 4, this CAS module effectively extracts spatial detail information associated with the region to be segmented within the encoder. It diminishes the impact of background noise and semantic gaps that occur between the feature mappings of the encoder and the decoder on the segmentation.

Within the CAS module, the encoder's low-level features and the decoder's high-level features are transformed to the same dimensionality using a 1×1 convolution. In our experiment, this was set to half the channel number of the encoder. Subsequently, these features are fused using a pixel-wise addition operation, and then the fused features are forwarded to a Squeeze and Excitation Block (SEB) module [29] as input. As depicted in the upper part of Fig. 4, the SEB module enhances the information interaction between channels and promotes the reusability of key features. This module consists of two primary steps. Firstly, we perform global average pooling(GAP) to condense spatial information into channel-wise descriptors. Then, a small Multi-Layer Perceptron (MLP) is employed to learn inter-channel relationships and generate channel-wise scaling factors. After SEB module, a 1×1 convolution is then applied to reduce the feature channels to one, and a Sigmoid activation function is used to obtain the weight of each pixel in the feature map. Lastly, the weighted feature map is multiplied element-wise with the encoder features, emphasizing the low-level detail information associated with the region to be segmented.

Importantly, within the CAS module, the scale factor of the SEB module varies according to the input channel number. When the input channel number is less than 64, the scale factor is set to 1/8. Conversely, if the input channel number exceeds 64, the scale factor is set to 1/16. This arrangement is designed to facilitate information interaction across the majority of channels, thus enhancing the expressive capability of the features. This design enables our model to effectively handle diverse partially annotated datasets and achieve superior performance in chest X-ray image segmentation tasks.

3.3. Multi-object segmentation based on dual decoders

3.3.1. Dual decoders architecture

To achieve accurate segmentation results for various targets, we have implemented a dual-decoder architecture. This design facilitates the separation of targets into two groups based on their overlap degree with the lung field and their attribute similarity, such as grayscale information. Then, the dual-decoder architecture processes the features extracted by the encoder, thereby diminishing the impact of excessive overlap areas on target segmentation. Both Decoder 1 and Decoder 2 employ residual blocks with weight normalization and bilinear interpolation operations to progressively restore semantic information to its original image size. Considering that the densely connected blocks extract features with higher semantic information in deeper layers, our approach only applies the CAS module in the first two layers to extract fine-grained spatial detail information.

The dual-decoder architecture enhances the deep learning model's ability to accurately segment multiple targets in chest X-ray images. Its proficiency in dealing with overlapping targets and single-pixel multi-target challenges makes it highly practical in authentic clinical applications. These advancements are crucial for diagnosing various lung ailments and also contribute valuable insights to solving the data scarcity issue in other medical image segmentation tasks.

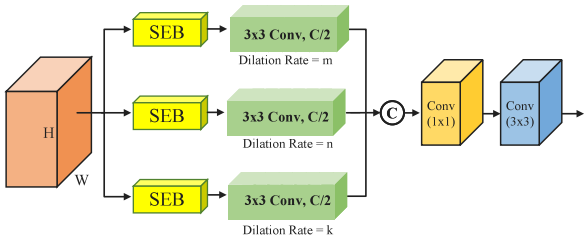

Fig. 5. Attention-guided multi-scale feature selection module. The number of channels in the three branches is C/2, and their corresponding dilation rates are m, n, and k.

3.3.2. Attention-guided multi-scale feature selection

In chest X-ray images, factors such as anatomical overlap, lung lesions, and gastric bubbles can easily cause local grayscale abnormalities in the clavicle and rib areas. A focus on local features within a small range may result in inaccurate segmentation of ribs or clavicles. Conversely, if the analysis is primarily relies on contextual information within a large range, misclassified pixels may lead to inaccurate segmentation. Inspired by multi-scale networks such as ASPP [30], we have designed an Attention-guided Multi-scale Feature Selection (AMFS) module. This module uses dilated convolutions of various sizes to capture multi-scale features, thereby enhancing the segmentation performance of our network.

As shown in Fig. 5, the AMFS module comprises three branches, each containing an SEB module and a 3×3 convolution with different dilation rates. The scale factor of the SEB module in each branch is set to 16. In Decoder 1, the dilation rates are set to 1, 2, and 6, respectively, enabling features extraction at scales of 3×3, 5×5, and 13×13. In Decoder 2, the dilation rates are set to 1, 3, and 7, enabling feature extraction at scales of 3×3, 7×7, and 15×15. The two decoders use different dilation rates, resulting in distinct receptive fields. This enables feature extraction at various scales, accommodating the requirement for more comprehensive feature extraction in complex target areas. During rib segmentation, the primary concern of the AMFS module is to address the issue of extreme grayscale unevenness within individual ribs caused by overlap or lesions. It helps tackle challenges in essential or hard-to-segment anatomical areas. In contrast, when segmenting targets like lung fields and hearts, the AMFS module focuses on both potential grayscale variations due to lesions and the influence of varying target sizes on segmentation.

To balance computational cost and segmentation performance, the module reduces the number of channels in each branch to half of the input channels, denoted as C/2 in Fig. 5. For instance, if the AMFS module takes an input of 512 channels, then each branch would consist of 256 channels. Subsequently, multi-scale information is fused through simple concatenation operations, followed by a 1×1 convolution to reduce the output channel count to the number of input channels. Then, a 3×3 convolution is used to further integrate and generate the final output features. This design not only enhances feature representation ability but also guarantees computational efficiency.

Both Decoder 1 and Decoder 2 utilize the Dice loss function for network training. In addition to the use of Dice loss function, this study proposes an alternating training strategy for different partially annotated datasets. A customized loss calculation method was employed for each segmentation task, excluding unannotated targets from the loss computation. In the scenario where both RCS-CXR and JSRT datasets are used for training, these two datasets are input into the network sequentially. Each target's Dice loss function weight is set to 1, with each decoder branch calculating the average. This mechanism allows the network to switch its focus between datasets, fostering a more balanced understanding of the diversity in each dataset and promoting more generalized feature learning. The alternating training strategy ensures that the model is not biased towards one particular dataset but rather, gains a comprehensive understanding of the underlying patterns across multiple datasets.

In situations where RCS-CXR, JSRT, VinDr-RibCXR, and Montgomery datasets are all incorporated into the training process, all four training sets are alternated. The alternating approach enables the encoder to adapt to diverse target features and dataset variations, leading to improved generalization across multiple datasets. During this process, the weight of the clavicle in Decoder 1 is set to 1.5, and the weight of the heart in Decoder 2 is set to 2. The weights for the Dice loss function of all other targets are set to 1. This diverse weighting strategy ensures that our network pays particular attention to certain anatomical regions that are critical or difficult to segment. It balances the trade-off between various targets and their complexities, allowing the model to effectively learn to handle diverse scenarios, thereby improving its segmentation ability. Overall, the two decoders have three main differences: (1) Different processing targets; (2) Different CAS module parameter settings; (3) Different loss weight settings during the training process. These designs enable the network to better learn and handle the features of different targets within various datasets. The convergence of the composite loss function ensures effective feature learning across different datasets. Other network training parameters are as follows: A maximum of 150 iterations were performed, with a learning rate decay of 0.1 every 40 epochs after 60 epochs. The initial learning rate was set to 0.0003, and the Adam optimizer was utilized. During training, the batch size was set to 2, while it was set to 1 during the validation and testing phases.

4. Experiments

4.1. Dataset and evaluation index

The algorithm utilized four partially annotated datasets for training, as shown in Fig. 1. The dataset division consisted of the RCS-CXR dataset for rib and clavicle segmentation [4], the JSRT dataset for lung, heart, and clavicle segmentation [5], the VinDr-RibCXR dataset for rib segmentation [6], and the Montgomery dataset for lung field segmentation [7]. To ensure uniformity in data, all images were resized to 512×512 pixels. A four-fold cross-validation approach was utilized. This entails four repetitions, where one subset serves as validation while the remaining three form the training set. Performance metrics from each repetition were averaged for an overall model effectiveness assessment. The segmentation algorithm's performance was evaluated using the Dice coefficient (DSC) and Jaccard coefficient (Jaccard).

During the data preprocessing stage, augmentation techniques were applied, including rotation, translation, shearing, horizontal flipping, and contrast transformation and so on. The augmentation factor was set to 10 for the RCS-CXR and Montgomery datasets, and 5 for the VinDr-RibCXR and JSRT datasets. The algorithms were implemented on the Ubuntu 20.04 operating system, using an NVIDIA GeForce RTX 3090. The training process followed an alternating training strategy, incorporating multiple datasets simultaneously. Experimental figures displayed annotations in different colors: predicted results were shown in red, ground truth in blue, and overlapping regions between predicted results and ground truth were highlighted in yellow.

4.2. Performance of ablation experiments

We conducted a series of ablation experiments using the RCS-CXR and JSRT datasets. Starting from the baseline network, we gradually added the CAS module, AMFS module, and data augmentation. The experimental results were summarized in Table 1. In the baseline network, the common features extracted from multiple partially annotated datasets differed from the target regions of interest in each specific annotated dataset. For example, the posterior ribs and anterior ribs mainly exist within the lung field. Including the JSRT dataset increased the extraction of features from the entire lung field by the encoder. The

Table 1
Ablation results of each target on two different test sets. The evaluation index is DSC (%).

Dataset	RCS-CXR				JSRT		
Target	All bone	Clavicle	Posterior rib	Anterior rib	Lung	Heart	Clavicle
Baseline	87.97	93.84	86.03	79.96	97.59	93.70	93.73
Baseline + CAS	90.16	94.82	89.25	84.62	97.57	94.16	93.86
Baseline + CAS + AMFS	90.24	94.93	89.36	84.60	97.77	94.20	94.11
Baseline + CAS + AMFS + Aug	**90.94**	**95.27**	**90.13**	**85.58**	**97.80**	**94.36**	**94.43**

Table 2
Comparative experiments on different training strategies on the RCS-CXR test set, where R, J, V, and M represent the RCS-CXR, JSRT, VinDr-RibCXR, and Montgomery dataset, respectively.

Target	Dataset for training	DSC	Jaccard
All Bone	R	90.45	82.63
	R,J	90.57	82.83
	R,J,V,M	**90.94**	**83.44**
Clavicle	R	95.05	90.63
	R,J	95.14	90.79
	R,J,V,M	**95.27**	**91.11**
Posterior rib	R	89.65	81.31
	R,J	89.83	81.61
	R,J,V,M	**90.13**	**82.10**
Anterior rib	R	84.86	73.90
	R,J	84.90	73.98
	R,J,V,M	**85.58**	**75.05**

Table 3
Comparative experiments on different training strategies on the JSRT test set, where R, J, V, and M represent the RCS-CXR, JSRT, VinDr-RibCXR, and Montgomery dataset, respectively.

Target	Dataset for training	DSC	Jaccard
Lung	J	97.60	95.35
	R,J	97.80	95.71
	R,J,V,M	**97.86**	**95.84**
Heart	J	93.52	87.99
	R,J	**94.36**	**89.45**
	R,J,V,M	94.25	89.28
Clavicle	J	93.64	88.15
	R,J	**94.43**	**89.52**
	R,J,V,M	94.30	89.32

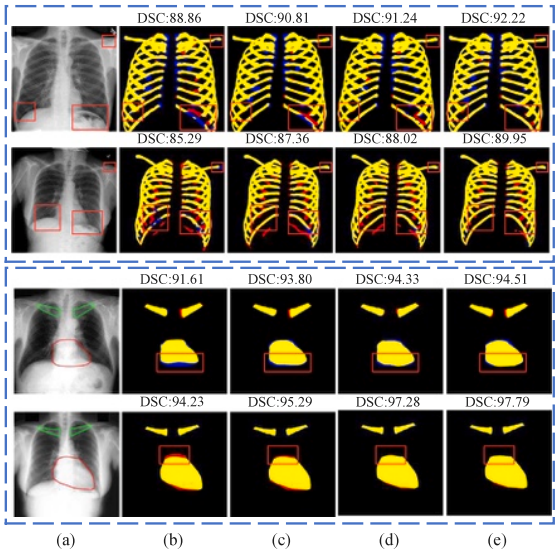

Fig. 6. Results of ablation experiments on RCS-CXR (upper half) and JSRT (lower half) test sets, (a) input images, (b) baseline, (c) baseline+CAS, (d) baseline+CAS+AMFS, (e) baseline+CAS+AMFS+Aug.

introduction of the CAS module selectively filtered important information related to bone structures such as ribs, as well as other targets like the lung field and heart. This significantly improved the segmentation performance of ribs on the RCS-CXR test set. The DSC metrics for anterior ribs, posterior ribs, and all bones improved by 4.66%, 3.22%, and 2.19% respectively. It also improved the segmentation performance of the heart and clavicle in the JSRT dataset.

As shown in Table 1, the introduction of the AMFS module improved the segmentation performance of the clavicle, posterior ribs, and lung field. Furthermore, with the addition of data augmentation in conjunction with the CAS and AMFS modules, the segmentation performance reached its peak. In summary, our experimental findings indicate that the CAS module, AMFS module, and data augmentation positively contributed to enhancing segmentation performance. These findings demonstrate the significance of these modules and techniques in enhancing the accuracy of the segmentation process.

To comprehensively assess the effect of each module on object segmentation, we visualized the results of the ablation experiments in Fig. 6. When examining the RCS-CXR dataset, we noticed that areas with poor segmentation results were mainly concentrated in regions of low contrast and uneven grayscale values. However, with the integration of the CAS module, the network's ability to perceive details in low contrast areas improved significantly, effectively aiding the network in tackling intricate features within the image. Additionally, the AMFS module demonstrated excellent performance in maintaining the continuity of rib segmentation. It aided the network in preserving the integrity of ribs and reducing discontinuities in the segmentation results. When data augmentation was applied, the segmentation results became more accurate and effectively reduced the occurrence of false-positive regions. As shown in Fig. 6, the JSRT dataset's segmentation task encountered difficulties due to low contrast and indistinct boundaries between the heart and surrounding tissues. Nevertheless, the integration of the CAS and AMFS modules enabled the network to capture finer details in the edge regions of the heart and produce more accurate segmentation results. This further showcased the effectiveness of these modules in improving segmentation performance. By visualizing the experimental outcomes, we observed that the incorporation of these modules positively impacted segmentation accuracy, enhanced the network's perception capabilities, and improved segmentation continuity in specific areas.

4.3. Performance using different training strategy

To evaluate the effect of incorporating other partially annotated datasets on multi-object segmentation, we trained a single-encoder single-decoder network separately on the RCS-CXR and JSRT datasets (indicated as "R" or "J" in Tables 2 and 3). Then, we augmented the training process with the four datasets. This strategic training approach enables more effective utilization of partially annotated datasets that strongly correlate with the segmentation targets. Fig. 7 illustrates the segmentation results under different training strategies. Each subfigure includes the original image in the first column, the results obtained using a single dataset for training in the second column, the results obtained using two datasets for training in the third column, and the results obtained using all four datasets for training in the fourth column.

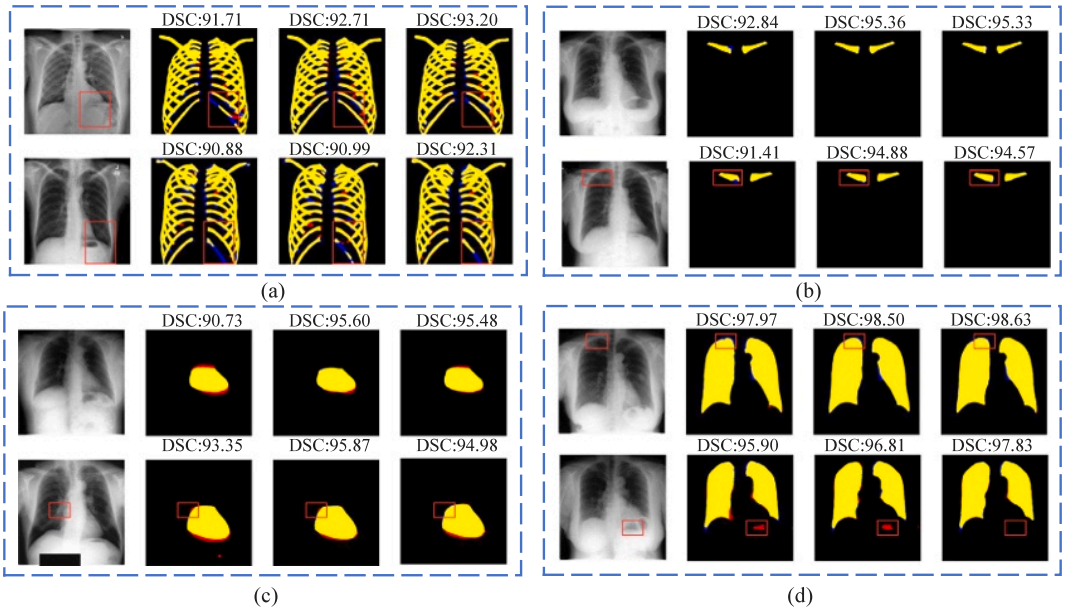

Fig. 7. Comparative experiments were conducted to analyze the impact of different training strategies on the segmentation results of various targets in the test set, including (a) all bones in the RCS-CXR test set, (b) clavicle in the JSRT test set, (c) heart in the JSRT test set, and (d) lung field in the JSRT test set. The segmentation results in each sub-figure corresponds to the three training strategies in Tables 2 and 3.

As shown in Tables 2 and 3, the network trained by jointly utilizing two datasets shows improved segmentation performance on various targets. On the JSRT test set, the challenging task of clavicle segmentation demonstrates significant improvement, with an increase of 0.79 and 1.37 percentage points in the DSC and Jaccard values, respectively. This improvement can be attributed to the supplemental information about clavicle samples from the RCS-CXR dataset, indicating that even with different annotation standards for the target, segmentation performance can influence each other. Additionally, there was an improvement in the segmentation performance of the heart within the JSRT dataset, with the accuracy increasing from 93.52% to 94.36%. From Fig. 7(a), the enhanced continuity of the ribs is clearly noticeable. This improvement can be mainly attributed to the supplementary positional information derived from the lung field annotations present in the JSRT dataset. This enhancement can be attributed to the reference positional information provided by the ribs, which plays a crucial role in guiding the segmentation of the heart (Fig. 7(c)). There is a noticeable improvement in the accuracy of the segmented edges in the region where the heart and ribs overlap. As shown in Fig. 7(d), the lung field, occupying the largest area, is relatively easier to segment. The introduction of differently annotated data significantly mitigated misclassification.

As indicated in Table 2, the inclusion of the VinDr-RibCXR dataset and Montgomery dataset further improved the segmentation performance of all bones, clavicles, posterior ribs, and anterior ribs. This further supports the idea that even with different annotation standards for the target, segmentation performance can influence each other. Nevertheless, with the incorporation of the Montgomery dataset, the segmentation performance of the heart and clavicles exhibited a slight decline, as illustrated in Table 3. This phenomenon primarily arises from the introduction of the Montgomery dataset, which augmented the quantity of lung field segmentation data. Consequently, an imbalance emerged among the samples utilized in training for the lung field, heart, and clavicles. This imbalance led to a decrease in the segmentation performance of the heart and clavicles.

In summary, the experiments demonstrate that the introduction of additional relevant datasets could provide more information and context, aiding the network in better understanding and segmenting the target. This has a significant positive impact. Especially when the dataset has a strong correlation with the segmentation targets, it can markedly enhance segmentation performance. This confirms the efficacy of using multiple datasets for target segmentation tasks and indicates that such a training strategy can be applied in practice to improve the quality of segmentation results. However, they might also introduce some issues, such as sample imbalance, which may negatively affect the segmentation performance of specific targets. These observations validate the efficacy of training with multiple datasets and suggest the adoption of these training strategies in practice to enhance the quality of segmentation results.

4.4. Comparison with other popular methods

To validate the state-of-the-art performance of our algorithm in multi-object segmentation of chest X-ray images, we compared its experimental results with several popular medical image segmentation networks. The segmentation results for the RCS-CXR dataset are summarized in Table 4 and Fig. 8. Notably, the most favorable performance was attained in clavicle segmentation, followed by all bones, posterior ribs, and anterior ribs. Our algorithm outperformed other networks in multi-object segmentation, with average DSC and Jaccard index values of 90.48% and 82.90%, respectively. While nnUNet showed slightly lower performance compared to our algorithm, our approach yielded superior DSC for anterior rib segmentation. Attention UNet, Ce-net, and CPF-Net demonstrated similar performance. However, BiSeNetV2 and DDRNet performed poorly, exhibiting jagged edges in the segmentation of objects in Fig. 8. This can be attributed to their direct upsampling of the image to the original size, which resulted in the loss of fine image details. It is worth noting that both CPF-Net and our algorithm employed residual blocks in the encoder stage. Moreover, we improved skip connections and utilized different methods to capture contextual information. Our algorithm consistently outperformed CPF-Net in segmenting all bones, clavicles, posterior ribs, and anterior ribs. It effectively mitigated false-positive regions in CPF-Net, further emphasizing the efficacy of our proposed algorithm in rib segmentation. In conclusion, our algorithm demonstrated advanced performance in multi-object segmentation of chest X-ray images, particularly in

Table 4
Comparison of some popular segmentation methods on RCS-CXR test set.

Index	All bone		Clavicle		Posterior rib		Anterior rib		Avg	
	DSC	Jaccard	DSC	Jaccard	DSC	Jac	DSC	Jaccard	DSC	Jaccard
UNet [21]	89.34	80.81	93.95	88.67	88.83	79.98	81.81	69.56	88.48	79.75
Attention UNet [31]	89.72	81.41	94.07	88.92	88.87	80.05	83.95	72.54	89.15	80.73
UNet++ [32]	89.52	81.09	93.71	88.31	88.65	79.71	83.65	72.13	88.88	80.31
nnUNet [33]	89.95	81.87	95.07	90.67	89.60	81.26	83.67	72.12	89.58	81.48
Ce-net [34]	89.92	81.74	93.89	88.60	89.05	80.33	83.33	71.63	89.05	80.58
CPF-Net [35]	89.94	81.78	94.14	89.00	89.12	80.45	83.88	72.45	89.27	80.92
BiseNetV2 [36]	87.65	78.08	91.70	84.87	86.56	76.39	80.51	67.58	86.60	76.73
DDRNet [37]	87.17	77.32	91.19	83.93	86.00	75.52	79.18	65.74	85.88	75.63
Our method	**90.94**	**83.44**	**95.27**	**91.03**	**90.13**	**82.10**	**85.58**	**75.05**	**90.48**	**82.90**

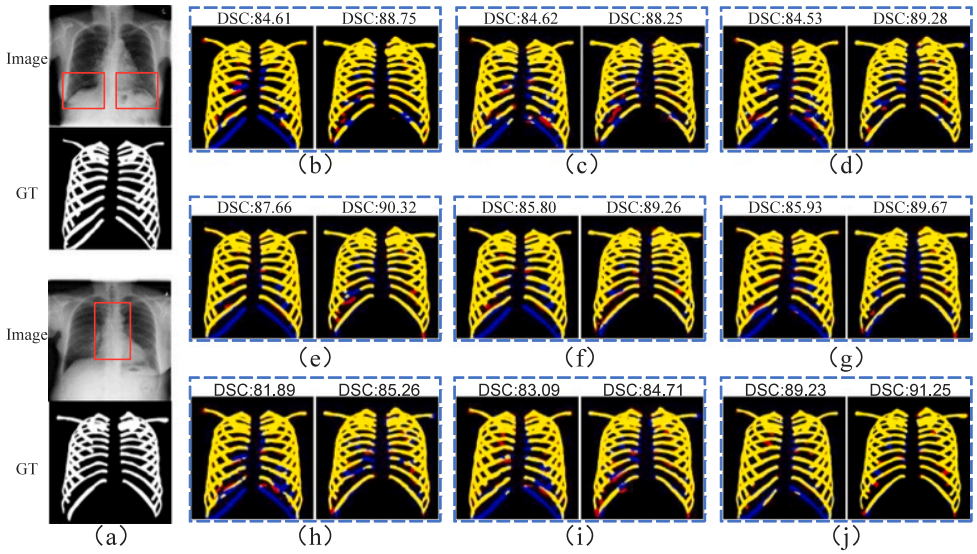

Fig. 8. The segmentation performance of different popular methods on the RCS-CXR test set. (a) Image and GT, (b) UNet [21], (c) Attention UNet [31], (d) UNet++ [32], (e) nnUNet [33], (f) Ce-net [34], (g) CPF-Net [35], (h) BiSeNetV2 [36], (i) DDRNet [37], (j) Our Method.

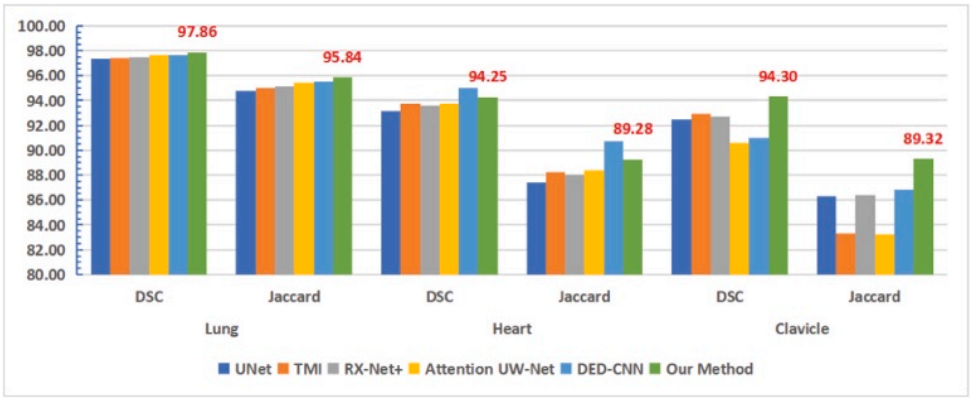

Fig. 9. Comparison of some popular segmentation methods on JSRT test set.

the segmentation of skeletal structures. These findings provide valuable insights for further improving the performance of chest X-ray segmentation techniques.

The JSRT dataset's segmentation results are summarized and compared in Figs. 9 and 10. Overall, the proposed algorithm outperforms other methods in terms of lung field and clavicle segmentation on the JSRT test set. Lung field segmentation achieves the highest accuracy, with all compared methods having DSC values exceeding 97%. The proposed MSM-CLMPAD exhibits better segmentation performance in terms of false-positive regions. Given the significant overlap between clavicles and ribs, clavicle segmentation presents the greatest challenge. The MSM-CLMPAD surpasses the TMI algorithm, the second-best method for clavicle segmentation, by 1.40%. At the edges of the clavicles, the proposed method demonstrates superior segmentation integrity compared to other methods. In heart segmentation, the proposed MSM-CLMPAD slightly lags behind the DED-CNN, which

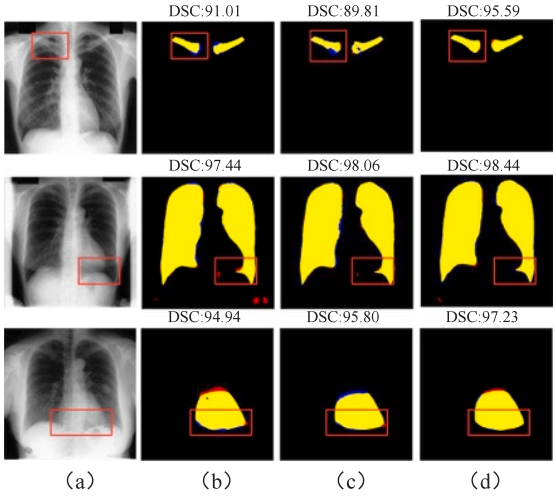

Fig. 10. The segmentation performance of different popular methods on the JSRT test set. (a) Input image, (b) UNet [21], (c) Attention UW-Net [20], (d) Our Method.

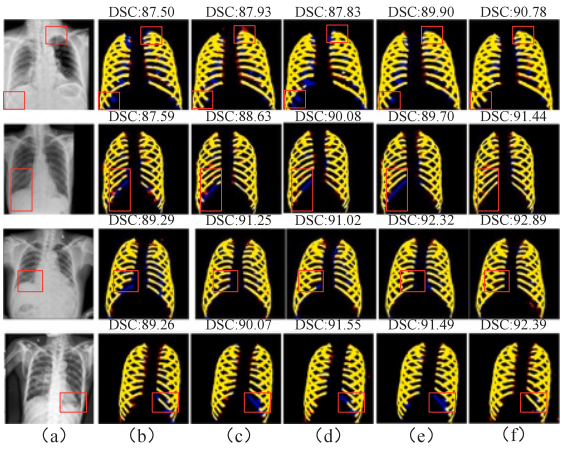

Fig. 11. The segmentation performance of different popular methods on VinDr-RibCXR test set. (a) Input Image, (b) UNet [21], (c) UNet w.EfficientNet [15], (d) UNet++ [32], (e) UNet++ w.EfficientNet [15], (f) Our Method.

utilizes a dual encoder–decoder structure and continuously optimizes the accuracy of the target through a two-stage approach. The DED-CNN performs well on independent targets but is less effective on clavicles with overlapping regions. Given that the proposed method has limited heart data and the additional partially annotated datasets introduced lack heart-related annotations to offer strong assistance, the segmentation performance for the heart is slightly lower. However, both the lung field and clavicles experience significant improvement, demonstrating the feasibility and benefits of jointly using multiple part-annotated datasets for multi-target segmentation in chest X-ray images.

The results of the VinDr-RibCXR, which is a single-target segmentation dataset, are summarized in Figs. 11 and 12. The proposed MSM-CLMPAD outperforms other methods in rib segmentation on the VinDr-RibCXR test set. Among the compared methods, the literature [15] utilizes EfficientNet-B0 as the encoder and the decoder follows the original UNet architecture. Similarly, the model labeled as "UNet++ w. EfficientNet-B0" also employs EfficientNet-B0 as the encoder. By employing EfficientNet-B0 as the encoder and scaling the network in terms of depth, width, and resolution to enhance feature extraction across multiple dimensions, the network's performance improves in

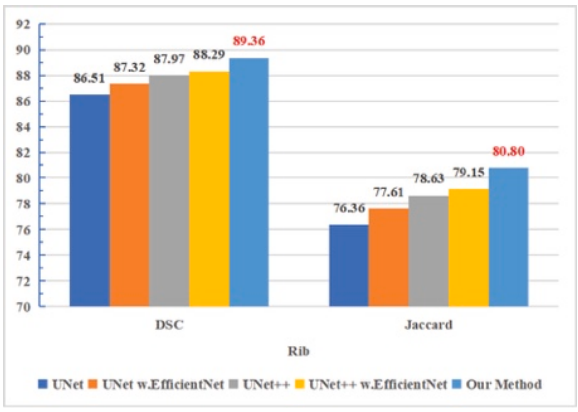

Fig. 12. Comparison of some popular segmentation methods on VinDr-RibCXR test set.

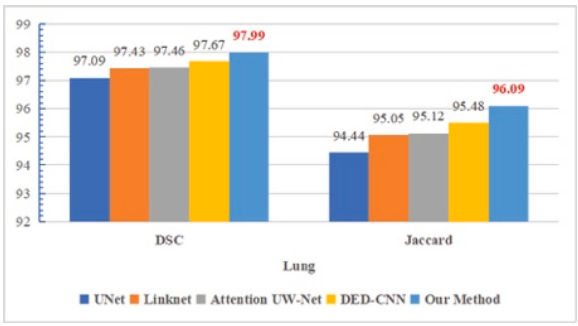

Fig. 13. Comparison of some popular segmentation methods on Montgomery test set.

comparison to the original UNet and UNet++. However, it still lags behind the proposed MSM-CLMPAD. Our method achieves a 1.07% DSC improvement over the second-best algorithm in rib segmentation. However, compared to the lung field and clavicles, rib segmentation poses greater difficulty due to its blurry edges and low contrast. As shown in Fig. 11, the improved EfficientNet-B0 network contributes to the continuity of rib segmentation to some extent. However, it does not fully rectify the issue of under-segmentation at rib boundaries, as observed in UNet and UNet++, especially for patients with pneumonia in the first and fourth rows.

As shown in Fig. 13, the proposed algorithm outperforms other methods in terms of lung field on the Montgomery test set. It achieves a 0.32% improvement in DSC and a 0.61% improvement in Jaccard index compared to DED-CNN. In contrast to the outcomes from the JSRT dataset (Fig. 9), DED-CNN achieves favorable results for independent lung field and heart segmentation via a multi-stage target optimization approach. As illustrated in Fig. 14, the Attention UW-Net improves the overall segmentation accuracy compared to the UNet by incorporating attention mechanisms into the UW-Net. However, over-segmentation is observed at the costophrenic angles (highlighted in red). When the lung texture is intensified in the patient's lungs (as shown in the second row of Fig. 14) and the contrast in the lung field region decreases, the proposed algorithm effectively addresses the under-segmentation issue (highlighted in blue). This enhancement is attributed to the supplementary information derived from multiple partially annotated datasets.

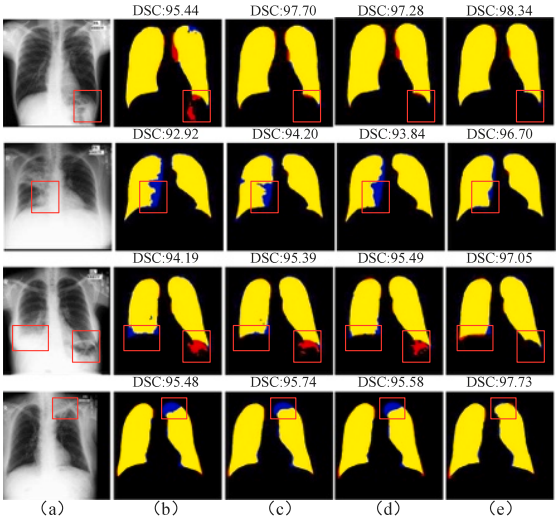

Fig. 14. The segmentation performance of different popular methods on Montgomery test set. (a) Input Image, (b) UNet [21], (c) LinkNet [38], (d) Attention UW-Net [20], (e) Our Method.

5. Conclusion

This study addresses the challenging task of multi-object segmentation in chest X-ray images in the absence of large-scale fully annotated datasets. This framework utilizes an encoder constructed with densely connected blocks and decoders built with weight-normalized residual blocks. During the feature extraction process, we employ a collaborative attention skip connection module and an attention-guided multi-scale feature selection module. These modules capture both spatial detail information and rich context information related to the segmentation task. Our collaborative learning strategy facilitates the joint participation of four partially annotated datasets in training a unified network. The experimental results indicate that our approach outperforms current popular methods, particularly when dealing with overlapping targets and lower contrast scenarios.

However, there are certain limitations to this approach. The utilization of dense connections and attention operations in the model to capture intricate image features has introduced some level of increased computational complexity. Additionally, the design of MSM-CLMPAD is tailored for chest X-ray image, with certain model parameters set manually. In future research, we plan to gather supplementary data to validate the method's robustness and performance.

CRediT authorship contribution statement

Hongyu Wang: Algorithms, Investigation, Writing, Revision. **Dandan Zhang:** Preprocessing, Coding, Algorithms, Editing. **Jun Feng:** Algorithms, Methods, Revision. **Lucia Cascone:** Investigation, Coding, Writing. **Michele Nappi:** Investigation, Coding, Visualization. **Shaohua Wan:** Methods, Writing, Editing, Revision.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

References

[1] N. Jeganathan, E. Nguyen, M. Sathananthan, Rheumatoid arthritis and associated interstitial lung disease: mortality rates and trends, Ann. Am. Thorac. Soc. 18 (12) (2021) 1970–1977.

[2] Y. Wu, Q. Kong, L. Zhang, A. Castiglione, M. Nappi, S. Wan, Cdt-cad: Context-aware deformable transformers for end-to-end chest abnormality detection on x-ray images, IEEE/ACM Trans. Comput. Biol. Bioinform. (2023).

[3] N. Subramanian, O. Elharrouss, S. Al-Maadeed, M. Chowdhury, A review of deep learning-based detection methods for COVID-19, Comput. Biol. Med. (2022) 105233.

[4] W. Wang, H. Feng, Q. Bu, L. Cui, Y. Xie, A. Zhang, J. Feng, Z. Zhu, Z. Chen, Mdu-net: A convolutional network for clavicle and rib segmentation from a chest radiograph, J. Healthc. Eng. 2020 (2020).

[5] B. Van Ginneken, M.B. Stegmann, M. Loog, Segmentation of anatomical structures in chest radiographs using supervised methods: a comparative study on a public database, Med. Image Anal. 10 (1) (2006) 19–40.

[6] H.C. Nguyen, T.T. Le, H.H. Pham, H.Q. Nguyen, VinDr-RibCXR: A benchmark dataset for automatic segmentation and labeling of individual ribs on chest X-rays, 2021, arXiv preprint arXiv:2107.01327.

[7] S. Jaeger, S. Candemir, S. Antani, Y.-X.J. Wáng, P.-X. Lu, G. Thoma, Two public chest X-ray datasets for computer-aided screening of pulmonary diseases, Quant. Imaging Med. Surg. 4 (6) (2014) 475.

[8] Y. Mo, Y. Wu, X. Yang, F. Liu, Y. Liao, Review the state-of-the-art technologies of semantic segmentation based on deep learning, Neurocomputing 493 (2022) 626–646.

[9] H. Wang, D. Zhang, S. Ding, Z. Gao, J. Feng, S. Wan, Rib segmentation algorithm for X-ray image based on unpaired sample augmentation and multi-scale network, Neural Comput. Appl. (2021) 1–15.

[10] B. Ni, Z. Liu, X. Cai, M. Nappi, S. Wan, Segmentation of ultrasound image sequences by combing a novel deep siamese network with a deformable contour model, Neural Comput. Appl. 35 (20) (2023) 14535–14549.

[11] S. Ding, H. Wang, H. Lu, M. Nappi, S. Wan, Two path gland segmentation algorithm of colon pathological image based on local semantic guidance, IEEE J. Biomed. Health Inf. 27 (4) (2022) 1701–1708.

[12] D. Zhang, H. Wang, J. Deng, T. Wang, C. Shen, J. Feng, CAMS-Net: An attention-guided feature selection network for rib segmentation in chest X-rays, Comput. Biol. Med. 156 (2023) 106702.

[13] Y. Liu, X. Zhang, G. Cai, Y. Chen, Z. Yun, Q. Feng, W. Yang, Automatic delineation of ribs and clavicles in chest radiographs using fully convolutional DenseNets, Comput. Methods Programs Biomed. 180 (2019) 105014.

[14] H. Oliveira, V. Mota, A.M. Machado, J.A. dos Santos, From 3D to 2D: Transferring knowledge for rib segmentation in chest X-rays, Pattern Recognit. Lett. 140 (2020) 10–17.

[15] M. Tan, Q. Le, Efficientnet: Rethinking model scaling for convolutional neural networks, in: International Conference on Machine Learning, PMLR, 2019, pp. 6105–6114.

[16] J. Wang, Z. Li, R. Jiang, Z. Xie, Instance segmentation of anatomical structures in chest radiographs, in: 2019 IEEE 32nd International Symposium on Computer-Based Medical Systems (CBMS), IEEE, 2019, pp. 441–446.

[17] K. He, G. Gkioxari, P. Dollár, R. Girshick, Mask r-cnn, in: Proceedings of the IEEE International Conference on Computer Vision, 2017, pp. 2961–2969.

[18] M. Kholiavchenko, I. Sirazitdinov, K. Kubrak, R. Badrutdinova, R. Kuleev, Y. Yuan, T. Vrtovec, B. Ibragimov, Contour-aware multi-label chest X-ray organ segmentation, Int. J. Comput. Assist. Radiol. Surg. 15 (2020) 425–436.

[19] T. Peng, T.C. Xu, Y. Wang, F. Li, Deep belief network and closed polygonal line for lung segmentation in chest radiographs, Comput. J. 65 (5) (2022) 1107–1128.

[20] D. Pal, P.B. Reddy, S. Roy, Attention UW-Net: A fully connected model for automatic segmentation and annotation of chest X-ray, Comput. Biol. Med. 150 (2022) 106083.

[21] O. Ronneberger, P. Fischer, T. Brox, U-net: Convolutional networks for biomedical image segmentation, in: Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part III 18, Springer, 2015, pp. 234–241.

[22] I. Ullah, F. Ali, B. Shah, S. El-Sappagh, T. Abuhmed, S.H. Park, A deep learning based dual encoder–decoder framework for anatomical structure segmentation in chest X-ray images, Sci. Rep. 13 (1) (2023) 791.

[23] O. Petit, N. Thome, L. Soler, Iterative confidence relabeling with deep ConvNets for organ segmentation with partial labels, Comput. Med. Imaging Graph. 91 (2021) 101938.

[24] X. Fang, P. Yan, Multi-organ segmentation over partially labeled datasets with multi-scale feature abstraction, IEEE Trans. Med. Imaging 39 (11) (2020) 3619–3629.
[25] G. Shi, L. Xiao, Y. Chen, S.K. Zhou, Marginal loss and exclusion loss for partially supervised multi-organ segmentation, Med. Image Anal. 70 (2021) 101979.
[26] J. Zhang, Y. Xie, Y. Xia, C. Shen, Dodnet: Learning to segment multi-organ and tumors from multiple partially labeled datasets, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021, pp. 1195–1204.
[27] G. Zhang, Z. Yang, B. Huo, S. Chai, S. Jiang, Multiorgan segmentation from partially labeled datasets with conditional nnU-Net, Comput. Biol. Med. 136 (2021) 104658.
[28] G. Huang, Z. Liu, L. Van Der Maaten, K.Q. Weinberger, Densely connected convolutional networks, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017, pp. 4700–4708.
[29] J. Hu, L. Shen, G. Sun, Squeeze-and-excitation networks, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018, pp. 7132–7141.
[30] L.-C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, A.L. Yuille, Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs, IEEE Trans. Pattern Anal. Mach. Intell. 40 (4) (2017) 834–848.
[31] O. Oktay, J. Schlemper, L.L. Folgoc, M. Lee, M. Heinrich, K. Misawa, K. Mori, S. McDonagh, N.Y. Hammerla, B. Kainz, et al., Attention u-net: Learning where to look for the pancreas, 2018, arXiv preprint arXiv:1804.03999.
[32] Z. Zhou, M.M. Rahman Siddiquee, N. Tajbakhsh, J. Liang, Unet++: A nested u-net architecture for medical image segmentation, in: Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support: 4th International Workshop, DLMIA 2018, and 8th International Workshop, ML-CDS 2018, Held in Conjunction with MICCAI 2018, Granada, Spain, September 20, 2018, Proceedings 4, Springer, 2018, pp. 3–11.
[33] F. Isensee, P.F. Jaeger, S.A. Kohl, J. Petersen, K.H. Maier-Hein, nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation, Nat. Methods 18 (2) (2021) 203–211.
[34] Z. Gu, J. Cheng, H. Fu, K. Zhou, H. Hao, Y. Zhao, T. Zhang, S. Gao, J. Liu, Ce-net: Context encoder network for 2d medical image segmentation, IEEE Trans. Med. Imaging 38 (10) (2019) 2281–2292.
[35] S. Feng, H. Zhao, F. Shi, X. Cheng, M. Wang, Y. Ma, D. Xiang, W. Zhu, X. Chen, CPFNet: Context pyramid fusion network for medical image segmentation, IEEE Trans. Med. Imaging 39 (10) (2020) 3008–3018.
[36] C. Yu, C. Gao, J. Wang, G. Yu, C. Shen, N. Sang, Bisenet v2: Bilateral network with guided aggregation for real-time semantic segmentation, Int. J. Comput. Vis. 129 (2021) 3051–3068.
[37] H. Pan, Y. Hong, W. Sun, Y. Jia, Deep dual-resolution networks for real-time and accurate semantic segmentation of traffic scenes, IEEE Trans. Intell. Transp. Syst. (2022).
[38] O. Gómez, P. Mesejo, O. Ibáñez, A. Valsecchi, O. Cordón, Deep architectures for high-resolution multi-organ chest X-ray image segmentation, Neural Comput. Appl. 32 (20) (2020) 15949–15963.