Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Pattern Recognition 113 (2021) 107828

Contents lists available at ScienceDirect

Pattern Recognition

journal homepage: www.elsevier.com/locate/patcog

Synergistic learning of lung lobe segmentation and hierarchical multi-instance classification for automated severity assessment of COVID-19 in CT images

Kelei He[a,b,1], Wei Zhao[c,1], Xingzhi Xie[c], Wen Ji[b,d], Mingxia Liu[f], Zhenyu Tang[g], Yinghuan Shi[b,d], Feng Shi[i], Yang Gao[b,d], Jun Liu[c,e,*], Junfeng Zhang[a,b,*], Dinggang Shen[h,i,j,*]

[a] Medical School of Nanjing University, Nanjing, China
[b] National Institute of Healthcare Data Science at Nanjing University, China
[c] Department of Radiology, the Second Xiangya Hospital, Central South University, Changsha,Hunan, China
[d] State Key Laboratory for Novel Software Technology, Nanjing University, Nanjing, China
[e] Department of Radiology Quality Control Center, Changsha, China
[f] Biomedical Research Imaging Center and the Department of Radiology, University of North Carolina, Chapel Hill, NC, U.S.
[g] Beijing Advanced Innovation Center for Big Data and Brain Computing, Beihang University, Beijing, China
[h] School of Biomedical Engineering, ShanghaiTech University, Shanghai, China
[i] Department of Research and Development, Shanghai United Imaging Intelligence Co., Ltd., Shanghai, China
[j] Department of Artificial Intelligence, Korea University, Seoul, Republic of Korea

ARTICLE INFO

Article history:
Received 16 June 2020
Revised 10 December 2020
Accepted 22 December 2020
Available online 16 January 2021

Keywords:
COVID-19
CT
Severity assessment
Lung lobe segmentation
Multi-instance learning

ABSTRACT

Understanding chest CT imaging of the coronavirus disease 2019 (COVID-19) will help detect infections early and assess the disease progression. Especially, automated severity assessment of COVID-19 in CT images plays an essential role in identifying cases that are in great need of intensive clinical care. However, it is often challenging to accurately assess the severity of this disease in CT images, due to variable infection regions in the lungs, similar imaging biomarkers, and large inter-case variations. To this end, we propose a synergistic learning framework for automated severity assessment of COVID-19 in 3D CT images, by jointly performing lung lobe segmentation and multi-instance classification. Considering that only a few infection regions in a CT image are related to the severity assessment, we first represent each input image by a bag that contains a set of 2D image patches (with each cropped from a specific slice). A multi-task multi-instance deep network (called M^2UNet) is then developed to assess the severity of COVID-19 patients and also segment the lung lobe simultaneously. Our M^2UNet consists of a patch-level encoder, a segmentation sub-network for lung lobe segmentation, and a classification sub-network for severity assessment (with a unique hierarchical multi-instance learning strategy). Here, the context information provided by segmentation can be implicitly employed to improve the performance of severity assessment. Extensive experiments were performed on a real COVID-19 CT image dataset consisting of 666 chest CT images, with results suggesting the effectiveness of our proposed method compared to several state-of-the-art methods.

1. Introduction

The coronavirus disease 2019 (COVID-19) is spreading fast worldwide since the end of 2019. Until October 5, about 37.10 million patients are confirmed with this infectious disease, among which 1.07 million were died, reported by.[3] This raises a Public Health Emergency of International Concern (PHEIC) of WHO. In the field of medical image analysis, many imaging-based artificial intelligence methods have been developed to help fight against this disease, including automated diagnosis [1–3], segmentation [4–6], and follow-up and prognosis [7].

Previous imaging-based studies mainly focus on identifying COVID-19 patients from non-COVID-19 subjects. As the golden

* Corresponding authors.
E-mail addresses: junliu123@csu.edu.cn (J. Liu), jfzhang@nju.edu.cn (J. Zhang), Dinggang.Shen@gmail.com (D. Shen).
[1] These authors contributed equally to this work.

[3] https://en.wikipedia.org/wiki/Template:COVID-19_pandemic_data.

https://doi.org/10.1016/j.patcog.2021.107828
0031-3203/

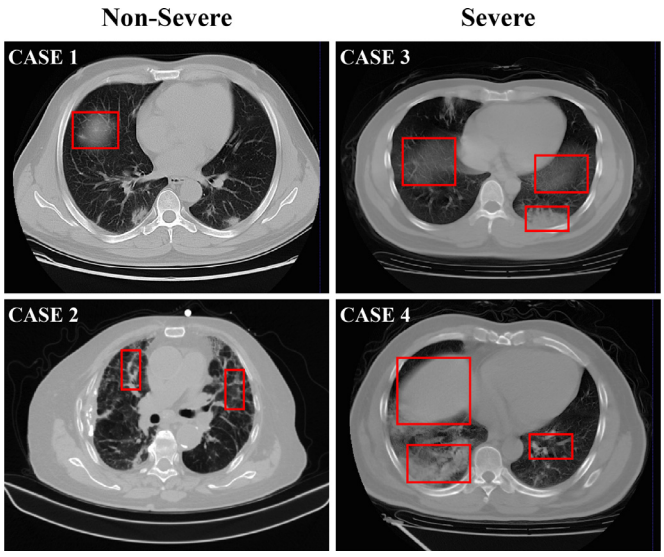


Fig. 1. Typical cases of two non-severe (left) and two severe (right) patients with COVID-19, where infections often occur in small regions of the lungs in CT images. The similar imaging biomarkers (e.g., ground glass opacities, mosaic sign, air bronchogram and interlobular septal thickening) of both cases (denoted by red boxes) make the non-severe and severe images difficult to distinguish. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

standard for COVID-19 is the Reverse Transcription-Polymerase Chain Reaction (RT-PCR) test, the effectiveness of those imaging-base applications is limited.

Moreover, approximately 80% of patients with COVID-19 have only mild to moderate symptoms [8], while the remaining patients have severe symptoms. Based on previous studies [4,9], the imaging-based characters of COVID-19 patients are distinct to related diseases, e.g., viral pneumonia. Therefore, the severity assessment of the disease is of high clinical value, which helps effectively allocate medical resources such as ventilator. Among various radiological examinations, chest CT imaging plays an essential role in fighting this infectious disease by helping early identify lung infections and assess the severity of the disease. Previous studies show that computed tomography (CT) has the ability to provide valuable information in the screening and diagnosis [9]. In this work, CT would help the clinicians to evaluate the condition of the patients in advance, by which necessary measures or treatments could better proceed, especially for severe patients in time.

However, an automatic assessment of the severity of COVID-19 in CT images is a very challenging task. First, infections caused by COVID-19 often occur in small regions of the lungs and are difficult to identify in CT images, as shown in Fig. 1. Second, imaging biomarkers of COVID-19 patients caused by an infection are similar in some severe and non-severe cases are similar, including ground-glass opacities (GGO), mosaic sign, air bronchogram, and interlobular septal thickening (Fig. 1). In addition, there are large inter-case variations in CT images of COVID-19 patients (Fig. 2), because these images are usually acquired by multiple imaging centers with different scanners and different scanning parameters.

Several recent methods have been proposed for the diagnosis of COVID-19 [1–3,10,11], with only some specifically designed for severity assessment of the disease. In several studies [4–6], segmentation of lung or lung lobe is used as a prerequisite procedure for diagnosis purposes.

However, most of these methods treat the lung lobe segmentation and disease diagnosis as two separate tasks, ignoring their underlying correlation. Note that the segmentation of lung lobe can provide rich information regarding spatial locations and tissue types in CT images. Therefore, it is intuitively reasonable to jointly perform lung lobe segmentation and severity assessment/prediction. The reason is that the context information provided by segmentation results can be used to improve the prediction performance. The joint learning scheme is obviously faster than the two-stage framework, since detecting and cropping the lung field are not needed. Besides, the classification task raises high signal responses in lung lobe area, as demonstrated by the related works of class activation maps (CAMs) [12]. Therefore, the infection patterns of lung lobe in disease progression could also provide useful guidance for the segmentation of lung lobes.

Moreover, most of the previous works are based on 2D image slices [1–3]. However, the annotation of 2D CT slices is a heavy workload for radiologists. It is interesting to directly employ 3D CT images for automated severity assessment of COVID-19, which is desired for real-world clinical applications.

To this end, in this paper, we propose a synergistic learning framework for automated severity assessment of COVID-19 in the raw 3D CT images, by jointly performing severity assessment and lung lobe segmentation. Considering that only a few slices in CT images are related to severity assessment, each input CT image is represented by a bag of 2D image patches which are randomly cropped from image slices. Furthermore, each patch is represented by a bag of infection regions represented by intermediate embedding features. Then, with each bag as input, a multi-task multi-instance deep neural network (called M^2UNet) is developed, including 1) a shared patch-level encoder, 2) a classification sub-network for severity assessment of COVID-19 patients (i.e., severe or non-severe) using a hierarchical multi-instance learning strategy, and 3) a segmentation sub-network for lung lobe segmentation. Extensive experiments have been performed on a real-world COVID-19 dataset with 666 chest CT images, with the results demonstrating the effectiveness of the proposed method compared to several state-of-the-art methods. The implementation of the proposed method is available at.[4]

The contributions of this work are three-fold:

- A multi-task multi-instance learning framework is proposed to jointly assess the severity of COVID-19 patients and segment lung lobes in chest CT images, where the segmentation task provides context information to aid the task of severity assessment in chest CT image.
- A unique hierarchical multi-instance learning strategy is developed to predict the severity of patients in a weakly supervised manner.
- We evaluate the proposed method on a real clinical dataset with 666 3D CT images of COVID-19 patients, achieved promising results in severity assessment compared to several state-of-the-art methods.

The rest of the paper is organized as follows. In Section 2, we introduce the related works for the segmentation and diagnosis of CT images of COVID-19 patients, as well as related studies on deep multi-instance learning. Then, we introduce the proposed method in Section 3. In Section 4, we present the materials, experimental setup, and experimental results. Finally, we conclude this paper and present several future research directions in Section 5.

2. Related work

In this section, we briefly review the most relevant studies from the following three aspects: 1) lung segmentation of CT images with COVID-19, 2) automated diagnosis of COVID-19 patients, and 3) deep multi-instance learning.

[4] https://github.com/KeleiHe/M2UNet.

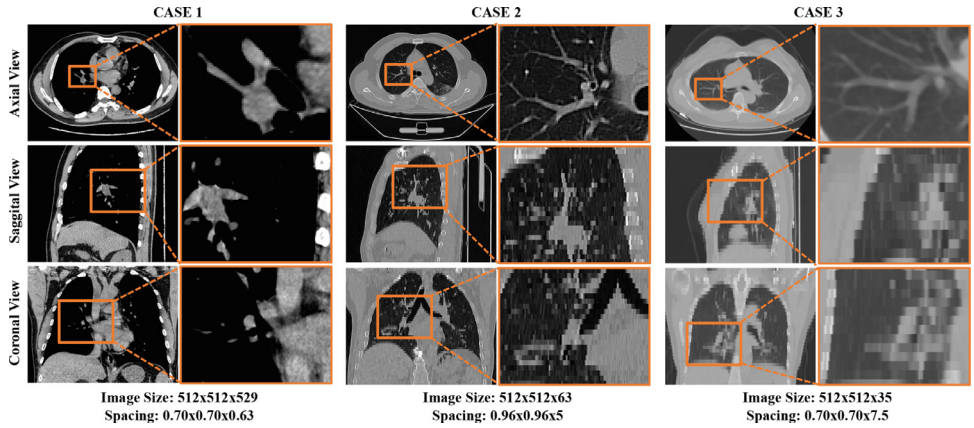


Fig. 2. Visualization of three typical cases in the COVID-19 CT image dataset from three different views. As shown in this figure, large inter-case variations (e.g., image size and spatial resolution) exist in CT images of COVID-19 patients.

2.1. Lung segmentation of CT images with COVID-19

Segmentation of lung or lung lobe has been used as a common pre-requisite procedure for automatic diagnosis of COVID-19 based on chest CT images. Several deep learning methods have been proposed for the segmentation of lung in CT images with COVID-19. For instance, U-Net et al.[13] has been widely used for segmentation of both lung regions and lung lesions in COVID-19 applications [6,14–16]. Qi et al. [6] use U-Net to delineate the lesions in the lung and extract radiometric features of COVID-19 patients with the initial seeds given by a radiologist for predicting hospital stay. Also, several variants of U-Net have been applied to the diagnosis or severity assessment of COVID-19. Jin et al. [5] design a two-stage pipeline to screen COVID-19 in CT images, and they utilize U-Net++ [17] to detect the whole lung region and to separate lesions from lung regions. Shan et al. [4] integrates human-in-the-loop strategy into the training process of VB-Net (a variant of V-Net). The human-aided strategy is an intuitive way to address the issue of lacking manual labels during segmentation in CT images.

2.2. Automated diagnosis of COVID-19

Both X-rays [18] and CT images [9] can provide effective information for the computer-assisted diagnosis of COVID-19. Compared with X-rays, chest CT imaging contains hundreds of slices, which is clearer and more precise but has to take more time for specialists to diagnose. Therefore, there is a great demand to use CT images for automated diagnosis of COVID-19. In general, the existing methods for COVID-19 diagnosis based on CT images can be roughly divided into two categories: 1) classification; 2) severity assessment. In the former category, many studies have been conducted to determine whether patients are infected with COVID-19 disease. For example, Chen et al. [1] exploits a UNet++ based segmentation model to segment COVID-19 related lesions in chest CT images of 51 COVID-19 patients and 55 patients with other diseases, and finally determine the label (COVID-19 or non-COVID-19) of each image based on the segmented lesions. Ying et al. [2] propose a CT diagnosis system, namely DeepPneumonia, which is based on the ResNet50 model to identify patients with COVID-19 from bacteria pneumonia patients and healthy people. In the second category, Tang et al. [3] proposed to first adopt VB-Net to separate the lung into anatomical sub-regions, and then use these sub-regions to compute quantitative features for training a random forest (RF) model for COVID-19 severity assessment (with labels of being non-severe or severe).

2.3. Deep multi-instance learning

The scenario of multi-instance learning (MIL) [19–21] or learning from weakly annotated data [22] arises when only a general statement of the category is given, but multiple instances can be observed. MIL aims to learn a model that can predict the label of a bag accurately, and many recent studies have focused on implementing MIL via deep neural networks. For instance, Oquab et al. [22] train a deep model with multiple image patches of multiple scales as input, and aggregate the prediction results of multiple inputs by using a max-pooling operation. Besides, many studies [23–25] propose to formulate image classification as a MIL problem so as to address the weakly supervised problem. Moreover, MIL is particularly suitable for problems with only a limited number (e.g., tens or hundreds) of training samples in various medical image-based applications, such as computer-assisted disease diagnosis [26–29]. For instance, Yan et al. [28] propose a two-stage deep MIL method to find discriminative local anatomies, where the first-stage convolutional neural network (CNN) is learned in a MIL fashion to locate discriminative image patches and the second-stage CNN is boosted using those selected patches. More recently, a landmark-based deep MIL framework [29] is developed to learn both local and global representations of MRI for automated brain disease diagnosis, leading to a new direction for handling limited training samples in the domain of medical image analysis. Since there are only a limited number of cases at hand, it is desirable to employ the multi-instance learning strategy for severity assessment of COVID-19 patients in chest CT images.

3. Proposed method

3.1. Framework

The framework of the proposed method is illustrated in Fig. 3, where the input is the raw 3D CT image and the output is the lung segmentation and severity assessment of COVID-19 patients (i.e., severe or non-severe). Specifically, each 3D CT image is processed via several image pre-processing steps. Then, a set of 2D image patches is randomly cropped from the processed image to construct an instance bag, and each bag represents a specific input CT image. This bag is regarded as the input of the proposed multi-task multi-instance U-Net (M^2UNet). The M^2UNet is designed to learn two tasks jointly, i.e., severity assessment of a COVID-19 patient and segmentation of the lung lobe.

As shown in Fig. 3, in M^2UNet, an encoding module is first used for patch-level feature extraction of 2D patches in each input

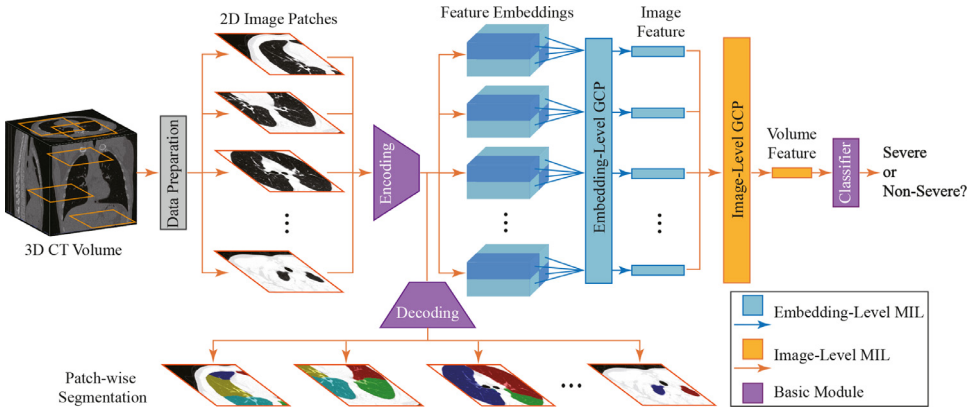


Fig. 3. Illustration of the proposed framework for joint lung lobe segmentation and severity assessment of COVID-19 in 3D CT images. Each raw 3D CT image is first pre-processed, and multiple 2D image patches (with each patch from a specific slice) are then extracted to construct an instance bag for representing each input CT scan. This bag is then fed into the proposed multi-task multi-instance UNet (M^2UNet) for joint lung lobe segmentation and severity assessment of COVID-19, consisting of a shared patch-level encoder, a segmentation sub-network, and a classification sub-network for severity assessment. Here, the segmentation task can provide location and tissue guidance for the task of severity assessment that employs a hierarchical multi-instance learning strategy.

bag, followed by two sub-networks for joint severity assessment and lung lobe segmentation. Specifically, in the classification sub-network, these extracted patch-level features are fed into a feature embedding module and an image-level feature learning module to capture the local-to-global volume representation of the input CT image. With the learned volume features, a classification layer is finally used to assess the severity of each COVID-19 patient (i.e., severe or non-severe). In the segmentation sub-network, those patch-level features are fed into a decoding module to perform lung lobe segmentation for each patch in the input bag. Since these two sub-networks are trained jointly with a shared patch-level encoder, the context information provided by the segmentation results can be implicitly employed to improve the performance of severity assessment.

3.2. Data preparation

To eliminate the effect of the background noise in each raw 3D CT image, we crop each scan to only keep the region containing the human body, by using a threshold-based processing method. Specifically, we first binarize the image using the threshold of zero, through which the human tissues and the gas regions will be separated. Then, the human body region is cropped according to the binary mask. Each body region image has a size of at least 256×256 for the axial plane in this work.

While image resampling is commonly used in many deep learning methods for segmentation and classification [30–32], we do not resample the raw CT images in order to preserve their original data distributions. Moreover, image registration methods [33,34] are not included in our pipeline. Since our method is clinical-oriented with inconsistent imaging qualities, CT images used in this study are not as clean as those in benchmark datasets [35]. For example, the physical spacing of our data has large variation, e.g., from 0.6 mm to 10 mm between slices, because of the use of different CT scanners and scanning parameters. Using a common interpolation method (e.g., trilinear interpolation) to resample a CT image into 1 mm, one will introduce heavy artifacts to the image. Besides, only a few infection regions in each CT image are related to severity assessment. To this end, we employ the weakly supervised multi-instance learning (MIL) strategy for handing these inconsistent CT images. Specifically, for each pre-processed CT image, we randomly crop a set of 2D patches sampled from 2D slices (with each patch from a specific slice) in each image to construct an instance bag, and each bag is used to represent a specific CT image and treated as the input of the subsequent M^2UNet. In this way, the inter-slice/patch relationships can be implicitly captured by our M^2UNet. In addition, this MIL strategy represents each 3D image through a set of 2D image patches rather than sequential slices. This can partially alleviate the problem of data inconsistency, so our method has high practical value in real-world applications.

3.3. Network architecture

As shown in Fig. 3, using each bag (consisting of a set of 2D image patches) as the input data, the proposed M^2UNet first employs an encoding module for patch-level feature extraction. Based on these features, the classification and segmentation sub-networks are then used to jointly perform two tasks, respectively, i.e., 1) severity assessment of the patients, and 2) segmentation of lung lobes in each patch. Specifically, the classification sub-network uses a unique hierarchical MIL strategy to extract the local-to-global representation of each input image, with an embedding-level MIL module, an image-level MIL module, and a classification module. The segmentation sub-network contains a decoding module to segment lung lobes of 2D image patches in each bag.

The detailed network architecture is listed in Table 1. The combination of the encoder and decoder is U-Net like, with four down-sampling blocks in the encoder and four up-sampling blocks in the decoder. The outputs of the same level blocks in the encoder and decoder are concatenated and fed into the next block of the decoder. Limited by computational resources, all the convolutional layers in the encoder and decoder have the same number (i.e., 64) of kernels, except the last block in the encoder. The last block of encoder outputs 512 dimensional features to help build a more robust classification for severity assessment. The decoder outputs the corresponding segmentation mask of five types of lung lobes for each image patch.

3.4. Hierarchical multi-instance learning

While infection regions of the lung, related to COVID-19 (e.g., nodule and GGO) are usually located in regions of the CT image, the category of each CT image is labeled at the entire image level, rather than the region-level. That is, many regions are actually unrelated to the classification task for severity assessment.

Multi-instance learning (MIL) provides a useful tool to solve such a weakly supervised problem. Conventional MIL represents a 2D image as a bag, and each bag consists of multiple regions of the input image (i.e., instances). Their overall prediction is made at the

Table 1
Network architecture of the proposed M^{2}UNet. The network has three main components: 1) a encoding module containing five encoding blocks; 2) a classification sub-network containing the embedding-level MIL and image-level MIL, and a classifier; and 3) a segmentation sub-network consisting of a decoding module with five decoding blocks. MIL: multi-instance learning; Num.: Number of layers, K: kernel size; PAD: padding size; STR: stride; #: Number of learnable parameters; cov: convolution; GCP: global contrast pooling; concat: concatenation.

Block Name	Num.	Layers	Parameter Setting	Input	#
Encoding block 1	2	{conv, batchnorm, ReLU}	K: $\{3\times3\times64\}$, PAD:1, STR:1	2D image patches	37*K*
Pool 1	1	max-pooling	K: $\{2\times2\}$, STR:2	Encoding block 1	-
Encoding block 2	2	{conv, batchnorm, ReLU}	K: $\{3\times3\times64\}$, PAD:1, STR:1	Pool 1	72*K*
Pool 2	1	max-pooling	K: $\{2\times2\}$, STR:2	Encoding block 2	-
Encoding block 3	2	{conv, batchnorm, ReLU}	K: $\{3\times3\times64\}$, PAD:1, STR:1	Pool 2	72*K*
Pool 3	1	max-pooling	K: $\{2\times2\}$, STR:2	Encoding block 3	-
Encoding block 4	2	{conv, batchnorm, ReLU}	K: $\{3\times3\times64\}$, PAD:1, STR:1	Pool 3	72*K*
Pool 4	1	max-pooling	K: $\{2\times2\}$, STR:2	Encoding block 4	-
Encoding block 5	1	{conv, batchnorm, ReLU}	K: $\{3\times3\times64\}$, PAD:1, STR:1	Pool 4	2595*K*
	1	{conv, batchnorm, ReLU}	K: $\{3\times3\times512\}$, PAD:1, STR:1		
Embedding-Level	1	GCP	Num. Concepts: 256	Encoding block 5	193*K*
MIL	1	conv	K: $\{1\times1\times256\}$, PAD:0, STR:1		
Image-Level MIL	1	GCP	Num. Concepts: 128	Embedding-Level MIL	48*K*
	1	conv	K: $\{1\times1\times128\}$, PAD:0, STR:1		
Classifier	1	conv	K: $\{1\times1\times128\}$, PAD:0, STR:1	Image-Level MIL	0.3*K*
Decoding block 5	1	{up-sample, conv, batchnorm, ReLU, concat}	K: $\{3\times3\times512\}$, PAD:1, STR:1	Encoding block 5	397*K*
Decoding block 4	1	{up-sample, conv, batchnorm, ReLU, concat}	K: $\{3\times3\times64\}$, PAD:1, STR:1	Decoding block 5	145*K*
	2	{conv, batchnorm, ReLU}	K: $\{3\times3\times128\}$, PAD:1, STR:1	Encoding block 3	
Decoding block 3	1	{up-sample, conv, batchnorm, ReLU, concat}	K: $\{3\times3\times64\}$, PAD:1, STR:1	Decoding block 4	145*K*
	2	{conv, batchnorm, ReLU}	K: $\{3\times3\times128\}$, PAD:1, STR:1	Encoding block 2	
Decoding block 2	1	{up-sample, conv, batchnorm, ReLU, concat}	K: $\{3\times3\times64\}$, PAD:1, STR:1	Decoding block 3	145*K*
	2	{conv, batchnorm, ReLU}	K: $\{3\times3\times128\}$, PAD:1, STR:1	Encoding block 1	
Decoding block 1	1	conv	K: $\{1\times1\times64\}$, PAD:0, STR:1	Decoding block 2	0.5*K*

bag-level by roughly two kinds of methods, i.e., the embedding-level methods and the instance-level methods. The former learns the relationship among instances by projecting them into a new embedding space. The latter directly generates the bag-level predictions by performing voting on the instance predictions. However, both methods are inapplicable for the classification of 3D images, as 3D images contain multiple 2D image slices, and the class labels are also related to some local regions of the slices.

In this work, based on the previous study on pathological images [36], we propose a hierarchical MIL strategy in the classification sub-network of our M^{2}UNet to perform severity assessment of COVID-19 in 3D CT images, as shown in Fig. 3. As mentioned in Section 3.2, we represent each input 3D volumetric image as a bag consisting of a set of 2D image patches, and these patches are regarded as the instances in the MIL problem settings. Formally, we first construct a bag with n 2D patches cropped from the regional slices to represent each input CT image. Denote the ith and the jth 3D CT image as $\mathcal{X}_i$ and $\mathcal{X}_j$, respectively, where $\mathcal{X}_i = \{\phi_{i1}^{ins}, \phi_{i2}^{ins}, \ldots, \phi_{in_i}^{ins}\}$ and $\mathcal{X}_j = \{\phi_{j1}^{ins}, \phi_{j2}^{ins}, \ldots, \phi_{jn_j}^{ins}\}$. Here, $\phi_{kl}^{ins} \in \mathrm{R}^d$ $(k = 1, 2, \ldots, n_k)$ indicates the lth instance of the kth image. Then, these 2D patches (size: $height \times width$) are fed into the encoding module for patch/instance-level feature extraction. These instance-level features are further fed into an embedding-level MIL module, which will be introduced later. After obtaining the instance-level features, the bag/image-level feature Φ_i are then generated by our proposed global contrast pooling (GCP) layer in the image-level MIL module.

As illustrated in Fig. 4, the proposed GCP layer aims to make the instance features closer to the relevant concepts, and also push those irrelevant instance features and concepts away from each other. In this work, the term "concept" denotes the to-be-learned feature of GCP layer that is discriminative for severity assessment. Theoretically, the concept is a normalized weight to map features in instance feature space to an ordered embedding space. Specifically, in the GCP layer, we assume the bag-level feature Φ_i is represented by the relationship between instance features and p concepts. Here, these concepts are learned to reveal the data structure in a global perspective. The bag-level feature is then denoted as a p dimensional feature vector, with each dimension denoting the maximum similarity between one concept and all instance features. We use the cosine function to measure such relationships.

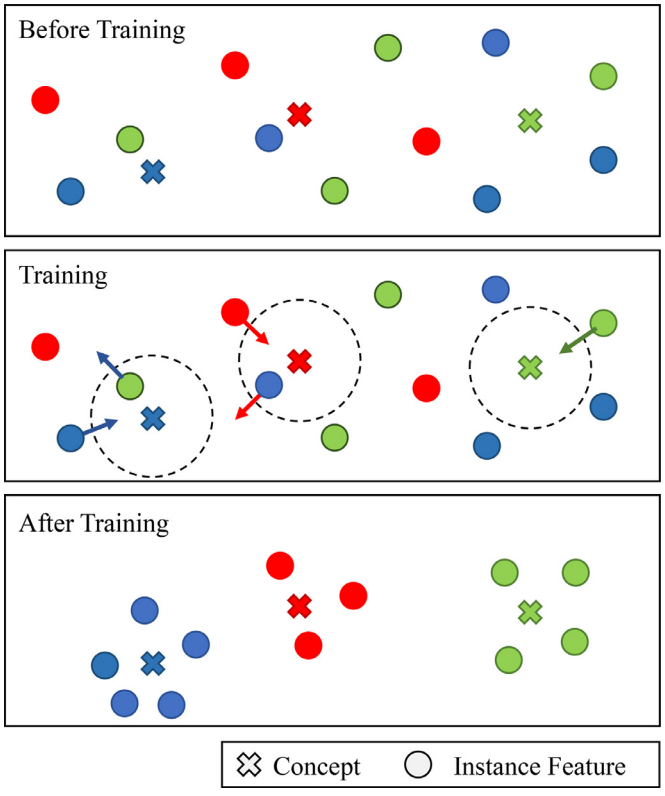


Fig. 4. A brief illustration of the learning principle for the proposed global contrast pooling (GCP) layer. Here, the concepts denote to-be-learned features that are discriminative for severity assessment. The GCP layer is designed to pull the relevant instance features and concepts closer, and push the irrelevant instance features and concepts away from each other.

Thus, the bag feature and the similarity can be written as

$$\mathbf{\Phi}_i = [s_{i1}, s_{i2}, \ldots, s_{im}, \ldots, s_{ip}], \tag{1}$$

$$s_{im} = \max_{k=1}^{n_i} \mathbf{w}_m^\top \phi_{ik} + \mathbf{R}(\mathbf{w}_m), \tag{2}$$

where s_{im} $(m = 1, \ldots, p)$ is the maximum similarity between the instance features of the ith bag and the mth image-level concept w_m. $R(\cdot)$ denotes the commonly used regularization term used in deep networks. With Eqs. (1)-(2), one can observe that the proposed GCP layer can automatically learn the concepts that are related to those discriminative instances, thus reducing the influence of those irrelevant instances. Note such a GCP layer can be also used in other weakly supervised problems, where only a small portion of regions in an image are related to the task at hand (such as MRI-based brain disorder diagnosis [29]).

We further use an embedding-level MIL module (with a GCP layer) to learn embedding-level representations, by regarding each image patch as a bag and the intermediate patch-level features produced by the encoder as instances. In this way, the relationships among small regions in each patch can be modeled by our method. Based on the embedding-level features, an image-level MIL module (with a GCP layer) is further used to generate the volume features. Based on the volume features, we use a fully-connected layer followed by a cross-entropy loss to predict the severity score (i.e., severe or non-severe) of each input CT image. The final loss function in the proposed hierarchical MIL network for severity assessment can be formulated as

$$\mathcal{L}_{MIL} = -log(fc(\Phi_i), y), \tag{3}$$

where $fc(\cdot)$ denotes the mapping function of the fully-connected layer, and y denotes the severity type confirmed by clinicians.

3.5. Multi-task learning for joint lung segmentation and severity assessment

The segmentation task is supervised by the aggregation of cross-entropy loss and Dice loss as follows

$$\mathcal{L}_{seg} = -\frac{1}{N}\sum_{n=1}^{N}\sum_{c=1}^{C} log(\hat{p}_n^c, l_n^c) - 2 \times \frac{\sum(\hat{p}_n^c \cap l_n^c)}{\sum \hat{p}_n^c + \sum l_n^c} + 1, \tag{4}$$

where $\hat{p}_n^c$ and l_n^c denote the predicted and ground-truth segmentation masks for the nth patch in the cth category. In this work, we segment $c = 7$ categories, including five parts of lung lobes and the background. It is worth noting that most of the cases in our dataset do not have ground-truth segmentation masks. For these cases, we simply avoid calculating the segmentation loss for them.

Finally, the losses in Eqs. (3) and (4) are trained simultaneously in a multi-task learning manner, and the overall loss of the proposed method is written as

$$\mathcal{L} = \lambda \mathcal{L}_{MIL} + \mathcal{L}_{Seg}, \tag{5}$$

where λ is the trade-off parameter used to balance the contributions of these two tasks. In this work, λ is empirically set to 0.01.

3.6. Implementation

The proposed method is implemented based on the open-source deep learning library *Pytorch*. The training of the network is accelerated by four NVidia Tesla V100 GPUs (each with 32 GB memory). For feasible learning of the lung region images, we clamp the intensities of the image into $[-1200, 0]$, which indicates that we use the width of 1200 and the level of -600 for the pulmonary window. Then, the data is normalized to the value of [0,255], as used by other deep learning methods. The dataset is highly imbalanced as the number of severe patients is much fewer than the non-severe patient. The ratio of the severe patient is less than 20% in our dataset. Therefore, we augmented the data by directly duplicated the severe cases in the training set. This can be done because the proposed method uses a random cropping strategy to construct the inputs. This makes the duplicated cases not the same to each other for the training of the network. We also use the random cropping strategy in the testing stage, by assuming that the data distribution is already well learned in training. Other cropping strategies, e.g., center cropping, may not be suitable here, as the center of pulmonary is dominated by the trachea and other tissues.

In both training and testing stage, we randomly crop 200 image patches from each input 3D CT image to construct the image-level bag (i.e., with the bag size of $n = 200$). The concept number p for multiple instance learning is set to 256 for embedding-level MIL module, and 128 for image-level MIL module. And we use the output of the encoder to construct the embedding-level bag that contains 8×8 feature maps. We train M^{2}UNet using the learning rate of 0.01 with a decay strategy of "Poly" (with the power of 0.75). The network is optimized by a standard Stochastic Gradient Descent (SGD) algorithm with 100 epochs. And the weights are decayed by a rate of 1×10^{-4} with the momentum of 0.9.

4. Experiments

In this section, we first introduce the materials, competing methods, and experimental setup. We then present the experimental results achieved by our method and several state-of-the-art methods. We finally investigate the influence of the parameters and two major strategies used in our method.

4.1. Materials

The real COVID-19 dataset contains a total of 666 3D chest CT scans acquired from 242 patients who are confirmed with COVID-19 (i.e., RT-PCR Test Positive). These CT images are collected from seven hospitals with a variety of CT scanners, including Philips (Ingenuity CT iDOSE4), GE (Bright speed S), Siemens (Somatom perspective), Hitachi (ECLOS), and Anke (ANATOM 16HD). The images are of large variation in terms of the image size of $512 \times (512 \sim 666) \times (23 \sim 732)$, and the spatial resolution of $0.586 \sim 0.984$ mm, $0.586 \sim 0.984$ mm and $0.399 \sim 10$ mm. Obviously, diagnosis based on these images is a very challenging task. The severity of the patient is confirmed by clinicians, following the guideline of 2019-nCoV (trail version 7) issued by the China National Health Commission. The severity of the patient is categorized into four types, i.e., mild, moderate, severe, and critical. In clinical practice, patients are often divided into two groups with different treatment regimens, i.e., severe and non-severe. The segmentation of 152 out of 666 images were delineated by an AI-based software and confirmed by experienced radiologists. In this work, we employ this partitioning strategy. That is, mild and moderate are treated as non-severe, while severe and critical are regarded as severe. Therefore, the task of severe assessment is formulated into a binary classification problem. Therefore, the dataset is partitioned into 51 severe and 191 non-severe patients.

4.2. Competing methods

We first compare the proposed M^{2}UNet with five state-of-the-art methods in [3,37–39] for severity assessment of COVID-19 patients. The first two methods [3,37] are both based on hand-crafted features of CT images, while the last two [38,39] are deep learning-based methods that can learn imaging features automatically from data. Specifically, Tang et al. [3] first segment the lung, lung lobe

and lesions in CT images. Then, the quantitative features of COVID-19 patients, e.g., the infection volume and the ratio of the whole lung, are calculated based on the segmentation results. The prediction is done by a random forest method. Yang et al. [37] proposed to aggregate infection scores calculated on 20 lung regions for severity assessment. ResNet50-3d is a non-MIL method using ResNet50 [38] with 3D convolutional layers to directly classify 3D CT images. In this case, all CT images are resampled to a fixed size of $128 \times 128 \times 64$. The ResNet50+Max method is compared for patch-wise classification. ResNet50+Max is an instance-level MIL method, which has a 2d version of ResNet-50 backbone to extract patch features, and performs image-level classification through max-voting of the patch features. In addition, we apply the Gated Att. MIL method proposed in [39] on our dataset, which is a embedding-level MIL method with an gated attention mechanism. For fair comparison, this method shares the same multi-instance pool as our M^{2}UNet.

We further compare our method with two state-of-the-art methods for lung lobe segmentation, including 1) UNet [13], and 2) UNet++ [17]. The parameter settings for these five competing methods are the same as those in their respective papers.

To evaluate the influence of the proposed multi-task learning and hierarchical MIL strategies used in M^{2}UNet, we further compare M^{2}UNet with its two variants: 1) M^{2}UNet with only the classification sub-network (denoted as Cls. Only), 2) M^{2}UNet with only the segmentation sub-network (denoted as Seg. Only).

4.3. Experimental setup

A five-fold cross-validation (CV) strategy is used in the experiments for performance evaluation. Specifically, the whole dataset is first randomly partitioned into five subsets (with approximately equal sample size of subjects). We treat one subset as the testing set (20%), while the remaining four subsets are combined to construct the training set (70%) and validation set (10%). The validation set here is used for selecting the hyper-parameters. This process is iterated until each subsets serve as a testing set once. The final results are reported on the test set.

Two tasks are included in the proposed method, i.e., 1) classification of severity assessment, and 2) segmentation of the lung lobe. For performance evaluation, two sets of metrics are used in these two tasks, with the details given below.

4.3.1. Metrics for classification

We use five commonly used metrics to evaluate the classification performance achieved by different methods in the severity assessment task, i.e., Accuracy, Precision, Recall, F1 Score, and the area under the receiver operating characteristic curve (AUC).

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}, \tag{6}$$

$$Precision = \frac{TP}{TP + FP}, \quad Recall = \frac{TP}{TP + FN}, \tag{7}$$

$$F1\ Score = \frac{2 \times Precision \times Recall}{Precision + Recall}. \tag{8}$$

where TP, TN, FP and FN denote true positive, true negative, false positive and false negative, respectively. In this work, the threshold of 0.5 is used to calculate these metrics.

4.3.2. Metrics for segmentation

We use three metrics, i.e., Dice Similarity Coefficient (DSC), Positive Predict Value (PPV) and Sensitivity (SEN), to evaluate the segmentation performance of different methods, with the definitions given below.

$$DSC = \frac{2\|V_{gt} \cap V_{seg}\|}{\|V_{gt}\| + \|V_{seg}\|}; \tag{9}$$

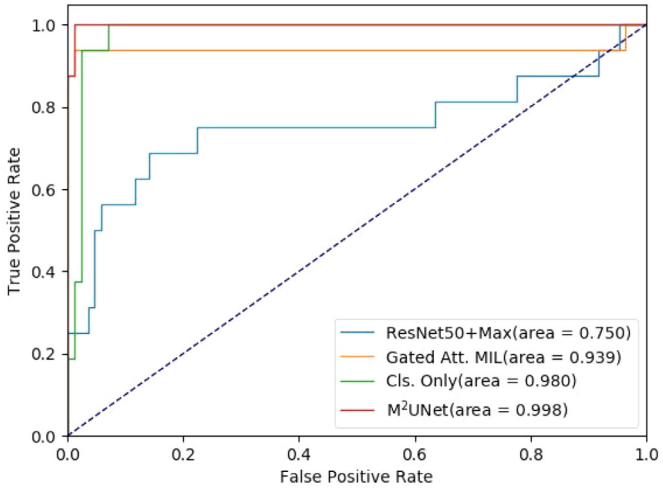


Fig. 5. The receiver operating characteristic (ROC) curves achieved by four different methods in the task of severity assessment.

$$PPV = \frac{\|V_{gt} \cap V_{seg}\|}{\|V_{seg}\|}; \qquad SEN = \frac{\|V_{gt} \cap V_{seg}\|}{\|V_{gt}\|}. \tag{10}$$

where V_{gt} and V_{seg} represent the ground-truth and predicted segmentation maps for each scan, respectively.

4.4. Comparison with state-of-the-art methods

4.4.1. Results of severity assessment

We first report the performance of seven different methods in the task of severity assessment for COVID-19 patients, with the results shown in Table 2. Note that the results from the competing methods are directly referred from the respective papers. As can be seen, ResNet50-3d performs worse than the other competing methods. This is mainly because the data spacing and sizes are inconsistent across the dataset, making the patterns vary greatly when resampling all the images into the same size. The other four deep learning-based methods usually outperform two hand-crafted feature-based methods in most cases. For some specific metrics, the method in [3] achieves the Recall of 0.933, which is significantly better than ResNet50+Max. The conventional embedding-level MIL-based method in [39] gets a performance improvement in terms of accuracy by 8%. Three MIL methods (i.e., [39], Cls. Only, and M^{2}UNet) yield satisfying performance, and the proposed M^{2}UNet achieves the best results (e.g.., the accuracy of 98.5% and F1 Score of 99.1%). However, the proposed method with multiple instances in MIL achieves the accuracy of 98.5% and F1 Score of 99.1%. The receiver operating characteristic (ROC) curves of four competing methods are illustrated in Fig. 5. Note that this ROC curve is plotted based on the results on one fold testing data, which is slightly different from the average performance on five-folds in Table 2. Table 2 and Fig. 5 clearly suggest that our M^{2}UNet generates the overall best performance in the task of severity assessment of COVID-19 patients based on chest CT images. The Precision-Recall curve of four competing methods are illustrated in Fig. 6. The Precision-Recall curve suggests that the proposed method is also the overall best method among these four methods. Also, the performance of the Gated Att. MIL method and Cls. Only method drop with threshold higher than 0.8.

4.4.2. Results of lung lobe segmentation

We then report the results of lung lobe segmentation achieved by four different methods in Table 3. Comparing Seg. Only and the

Table 2
Quantitative comparison for severity assessment tasks with the state-of-the-art methods.

Method	Accuracy	Precision	Recall	F1 Score	AUC
ResNet50-3d	0.913±0.059	0.902±0.118	0.756±0.106	0.786±0.112	0.759±0.102
ResNet50+Max	0.924±0.497	0.856 ± 0.154	0.793±0.106	0.816± 0.120	0.803±0.090
Gated Att. MIL [39]	0.955±0.015	0.879±0.054	0.946±0.019	0.906± 0.037	0.973±0.024
Tang et al. [3]*	0.875	-	0.933	-	0.910
Yang et al. [37]	-	-	0.750	-	0.892
Cls. Only (Ours)	0.969±0.023	0.928±0.073	**0.958±0.031**	0.938±0.045	0.980±0.013
M^{2}UNet (Ours)	**0.985±0.005**	**0.975±0.022**	0.952±0.011	**0.963±0.011**	**0.991±0.010**

Table 3
Quantitative comparison for the performance of lung lobe segmentation with the state-of-the-art methods.

Method	# (MB)	DSC	SEN	PPV
U-Net	131.71	0.776 ± 0.050	0.759 ± 0.037	**0.834 ± 0.033**
U-Net+	34.97	0.784 ± 0.035	0.773 ± 0.038	0.821 ± 0.018
Seg. Only (Ours)	14.37	0.759 ± 0.055	0.756 ± 0.064	0.785 ± 0.045
M^{2}UNet (Ours)	15.32	**0.785 ± 0.058**	**0.783 ± 0.059**	0.799 ± 0.051

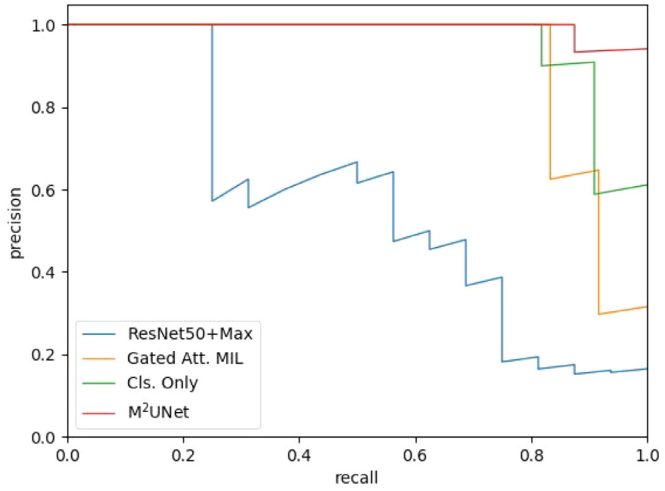


Fig. 6. The Precision-Recall (PR) curve achieved by four different methods in the task of severity assessment.

conventional U-Net, the former dramatically reduces the parameter from 131.71MB to 14.37MB. As a consequence, the performance in terms of DSC and PPV is also decreased by 1.7% and 4.9%, respectively. By using multi-task learning, M^{2}Net improves the performance, from 0.759 to 0.785 in terms of DSC, which also outperform the performance of conventional U-Net, with a decreasing of parameters, from 131.71 to 15.32. The proposed M^{2}UNet also achieves a slightly higher performance compared with U-Net++.

The visualization of segmentation results achieved by three different methods on two subjects is shown in Fig. 7. From this figure, we can see that M^{2}UNet generates the overall best segmentation masks, while U-Net and U-Net++ usually yield under-segmentation results on these cases. These results further show the advantage of our M^{2}UNet.

4.5. Ablation study

We further evaluate the influence of two major strategies used in our M^{2}UNet, i.e., 1) the hierarchical MIL strategy for classification, and 2) the multi-task learning strategy for joint severity assessment and lung lobe segmentation.

4.5.1. Influence of hierarchical MIL strategy

To evaluate the effectiveness of the hierarchical MIL strategy, we compare the variant of the proposed M^{2}UNet (i.e., Cls. Only without the segmentation sub-network) with a non-MIL method (i.e., ResNet50-3d) and two one-stage MIL methods (i.e., ResNet50+Max, Gated Att. MIL [39]). The classification results of these four methods in the task of severity assessment are reported in Table 2. As shown in Table 2, three MIL methods (i.e., ResNet50+Max, Gated Att. MIL and Cls. Only) can generate more accurate decisions under the weakly supervised setting, compared with the non-MIL method ResNet50+Max. Besides, our hierarchical MIL strategy can further boost the classification performance compared to the conventional one-stage MIL strategy. For instance, our Cls. Only method achieves an F1 Score of 0.938, which is higher than that (i.e., 0.906) yielded by Gated Att. MIL with a one-stage MIL strategy. These results suggest the effectiveness of the proposed hierarchical MIL strategy.

4.5.2. Influence of multi-task learning strategy

Our M^{2}UNet can jointly learn the segmentation task and the classification task in a multi-task learning manner. Here, we also investigate the influence of such a multi-task learning paradigm, by comparing M^{2}UNet with its two single-task variants, i.e., "Cls. Only" for classification and "Seg. Only" for segmentation. The performance comparison in two tasks for severity assessment and lung lobe segmentation are reported in Tables 2 and 3, respectively. Table 2 suggests that, compared with Cls. Only, the multi-task learning paradigm used in M^{2}UNet helps to improve the classification accuracy by 1.6%, while increasing the precision score by over 5% and the F1 Score by 2.5%. Notably, the F1 and precision of the Cls. Only method are already higher than 90%, which are hard to be improved. This is more valuable in this classification scheme, as the F1 score is more representative in evaluating such an imbalanced classification task.

As can be observed from Table 3, although M^{2}UNet is not specifically designed for lung lobe segmentation, it still improves the overall segmentation performance in terms of three metrics, compared with its single-task variant (i.e., Seg. Only). This implies that the proposed multi-task learning strategy is useful in boosting the learning performance of both tasks of severity assessment and lung lobe segmentation.

4.6. Influence of bag size

We further investigate the performance of our method using different bag sizes, and the results are shown

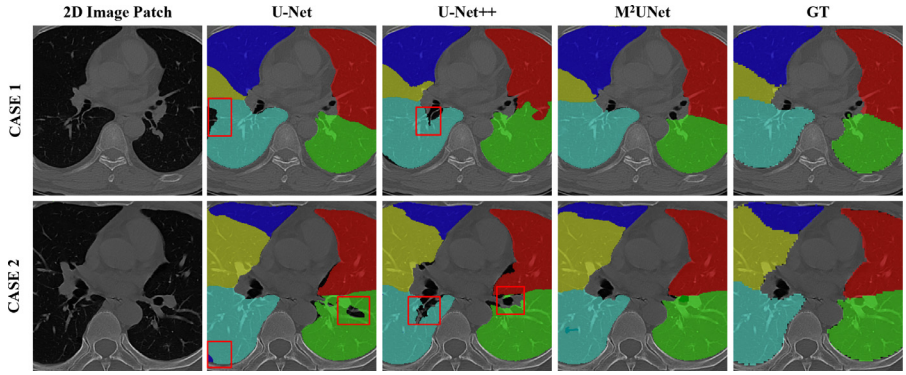


Fig. 7. The visualization of lung lobe segmentation results by three different methods on two typical cases. GT denotes the ground-truth masks. The under-segmentation regions are denoted by red boxes. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

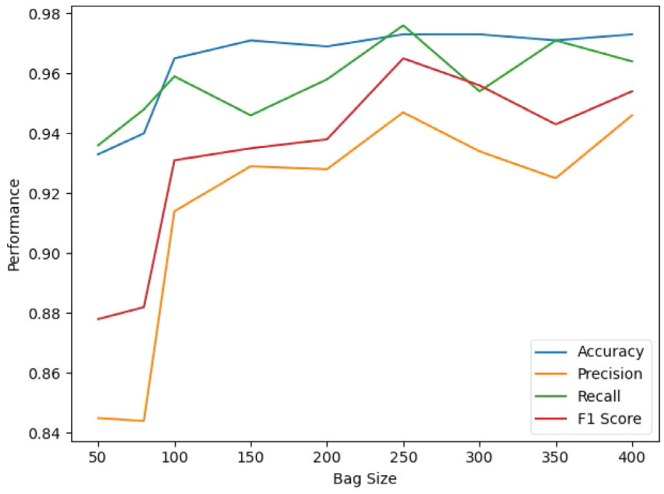


Fig. 8. Performance comparison for severity prediction with respect to different bag sizes.

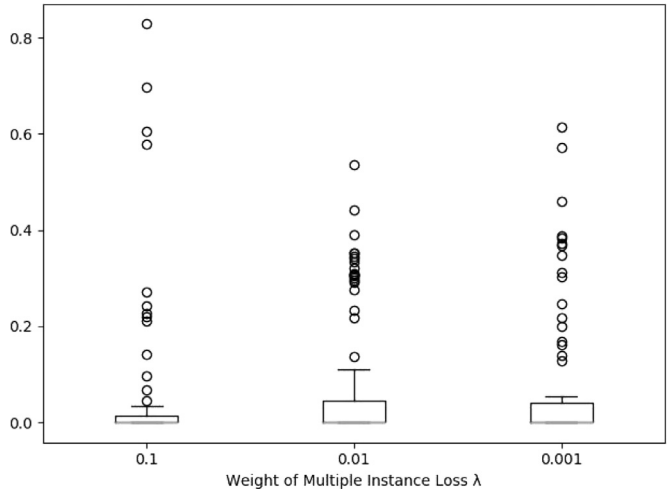


Fig. 9. Boxplot for the margin of prediction and label with respect to different weight λ for multiple instance loss. The margin larger than 0.5 indicates the wrong prediction.

in Fig. 8. Specifically, we set the bag size within {50, 80, 100, 150, 200, 250, 300, 350, 400}. As shown in the table, the classification performance of M^{2}UNet gradually improves along with the bag size. The proposed method achieves the best performance with the bag size of 250. Another observation is that the table suggests the performance of the proposed AI-based severity assessment model is insensitive to the bag size when larger than 100, indicating that at least 100 patches are required for the proposed method to achieve an acceptable result.

4.7. Influence of weigh parameter

We evaluate the influence of the hyper-parameter λ. The boxplot for the margin of prediction and ground-truth labels is shown in Fig. 9, where the margin is calculated by $|p - l|$, p denotes the prediction score and l denotes the label. The margin below 0.5 indicates that the prediction is correct. As shown in this figure, the task weight λ affects the classification performance of the model, and the proposed M^{2}UNet performs best with $\lambda = 0.01$.

4.8. Influence of learning rate

We further investigate the influence of the learning rate on the performance of M^{2}UNet, with the results given in Tables 4-5. As suggested by the table, the performance of M^{2}UNet for both classification and segmentation tasks are affected by different learning rates. The proposed method achieves the best performance for both classification and segmentation tasks with the learning rate of 0.01. Another observation is that the performance of the network is unstable with a large learning rate (i.e., 0.1), and the gap between precision and recall is large, with 0.833 in terms of precision and 0.970 in terms of recall. The performance of the network is stable with a small learning rate (i.e., 0.001), with precision of 0.881 and recall of 0.922. However, it cannot achieve the best performance, compared with the network trained by the learning rate of 0.01.

Table 4
Performance comparison for segmentation with respect to different learning rate.

Learning rate	DSC	SEN	PPV
0.1	0.783	0.772	0.789
0.01	0.785	0.783	0.799
0.001	0.734	0.705	0.754

Table 5
Performance comparison for severity prediction with respect to different learning rate.

Learning rate	Accurate	Precision	Recall	F1 Score
0.1	0.946	0.833	0.970	0.885
0.01	0.985	0.975	0.952	0.963
0.001	0.960	0.881	0.922	0.900

5. Conclusion and future work

In this paper, we propose a synergistic learning framework for automated severity assessment and lung segmentation of COVID-19 in 3D chest CT images. In this framework, we first represent each input image by a bag to deal with the challenging problem that the severity is related to local infected regions in the CT image. We further develop a multi-task multi-instance deep network (called M^2UNet) to assess the severity of COVID-19 patients and segment the lung lobe simultaneously, where the context information provided by segmentation can be employed to boost the performance of severity assessment. A hierarchical multi-instance learning strategy is also proposed in M^2UNet for severity assessment. Experimental results on a real COVID-19 CT image dataset demonstrate that our method achieves promising results in severity assessment of COVID-19 patients, compared with several state-of-the-art methods.

In the current work, our dataset is a multi-center dataset that is collecting in a short period. Therefore, constructing a new independent testing data can further evaluate the generalization ability of the proposed method. Also, the severity assessment of COVID-19 only relies on one time-point data, without considering longitudinal imaging biomarkers. It is interesting to perform a longitudinal study to investigate the progression of the disease, which is one of our future work. Since annotations for lung lobe in 3D CT images are usually tedious and error-prone, only a small number of subjects in our dataset have ground-truth segmentation. Therefore, it is highly desired to develop automated or even semi-automated image annotation methods, which will also be studied in the future. The methodology of this work is general. We believe it can be flexibly applied to the prediction and segmentation of other diseases, e.g., brain diseases. This will also be our future work.

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.

Acknowledgement

This work is supported in part by Key Emergency Project of Pneumonia Epidemic of novel coronavirus infection under grant 2020sk3006, Emergency Project of Prevention and Control for COVID-19 of Central South University under grant 160260005, Foundation of Changsha Scientific and Technical Bureau under grant kq2001001, and National Key Research and Development Program of China under grant 2018YFC0116400.

References

[1] J. Chen, L. Wu, J. Zhang, L. Zhang, D. Gong, Y. Zhao, S. Hu, Y. Wang, X. Hu, B. Zheng, et al., Deep learning-based model for detecting 2019 novel coronavirus pneumonia on high-resolution computed tomography: a prospective study, medRxiv (2020).
[2] Y. Song, S. Zheng, L. Li, X. Zhang, X. Zhang, Z. Huang, J. Chen, H. Zhao, Y. Jie, R. Wang, et al., Deep learning enables accurate diagnosis of novel coronavirus (COVID-19) with CT images, medRxiv (2020).
[3] Z. Tang, W. Zhao, X. Xie, Z. Zhong, F. Shi, J. Liu, D. Shen, Severity assessment of coronavirus disease 2019 (COVID-19) using quantitative features from chest CT images, arXiv preprint arXiv:2003.11988(2020).
[4] F. Shan, Y. Gao, J. Wang, W. Shi, N. Shi, M. Han, Z. Xue, D. Shen, Y. Shi, Lung infection quantification of COVID-19 in CT images with deep learning, arXiv preprint arXiv:2003.04655(2020).
[5] S. Jin, B. Wang, H. Xu, C. Luo, L. Wei, W. Zhao, X. Hou, W. Ma, Z. Xu, Z. Zheng, et al., Ai-assisted CT imaging analysis for COVID-19 screening: building and deploying a medical AI system in four weeks, medRxiv (2020).
[6] X. Qi, Z. Jiang, Q. Yu, C. Shao, H. Zhang, H. Yue, B. Ma, Y. Wang, C. Liu, X. Meng, et al., Machine learning-based CT radiomics model for predicting hospital stay in patients with pneumonia associated with SARS-CoV-2 infection: a multicenter study, medRxiv (2020).
[7] F. Shi, J. Wang, J. Shi, Z. Wu, Q. Wang, Z. Tang, K. He, Y. Shi, D. Shen, Review of artificial intelligence techniques in imaging data acquisition, segmentation and diagnosis for COVID-19, CoRR(2020). arXiv preprint:2004.02731.
[8] R. Verity, L.C. Okell, I. Dorigatti, P. Winskill, C. Whittaker, N. Imai, G. Cuomo–Dannenburg, H. Thompson, P.G.T. Walker, H. Fu, et al., Estimates of the severity of coronavirus disease 2019: a model-based analysis, Lancet Infect. Dis. (2020).
[9] X. Xie, Z. Zhong, W. Zhao, C. Zheng, F. Wang, J. Liu, Chest CT for typical 2019-nCoV pneumonia: relationship to negative RT-PCR testing, Radiology (2020) 200343.
[10] A. Abbas, M.M. Abdelsamea, M.M. Gaber, Classification of COVID-19 in chest X-ray images using DeTraC deep convolutional neural network, 2020. arXiv: 2003.13815.
[11] A. Oulefki, S. Agaian, T. Trongtirakul, A.K. Laouar, Automatic COVID-19 lung infected region segmentation and measurement using CT-scans images, Pattern Recognit. (2020) 107747.
[12] B. Zhou, A. Khosla, A. Lapedriza, A. Oliva, A. Torralba, Learning deep features for discriminative localization, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016, pp. 2921–2929.
[13] O. Ronneberger, P. Fischer, T. Brox, U-Net: convolutional networks for biomedical image segmentation, in: International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer, 2015, pp. 234–241.
[14] C. Zheng, X. Deng, Q. Fu, Q. Zhou, J. Feng, H. Ma, W. Liu, X. Wang, Deep learning-based detection for COVID-19 from chest CT using weak label, medRxiv (2020).
[15] Y. Cao, Z. Xu, J. Feng, C. Jin, X. Han, H. Wu, H. Shi, Longitudinal assessment of COVID-19 using a deep learning–based quantitative CT pipeline: illustration of two cases, Radiology 2 (2) (2020) e200082.
[16] L. Huang, R. Han, T. Ai, P. Yu, H. Kang, Q. Tao, L. Xia, Serial quantitative chest CT assessment of COVID-19: deep-learning approach, Radiology 2 (2) (2020) e200075.
[17] Z. Zhou, M.M.R. 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, Springer, 2018, pp. 3–11.
[18] H.Y.F. Wong, H.Y.S. Lam, A.H.-T. Fong, S.T. Leung, T.W.-Y. Chin, C.S.Y. Lo, M.M.-S. Lui, J.C.Y. Lee, K.W.-H. Chiu, T. Chung, et al., Frequency and distribution of chest radiographic findings in COVID-19 positive patients, Radiology (2020) 201160.
[19] T.G. Dietterich, R.H. Lathrop, T. Lozano-Pérez, Solving the multiple instance problem with axis-parallel rectangles, Artif. Intell. 89 (1–2) (1997) 31–71.
[20] K. Xu, J. Ba, R. Kiros, K. Cho, A. Courville, R. Salakhudinov, R. Zemel, Y. Bengio, Show, attend and tell: neural image caption generation with visual attention, in: International Conference on Machine Learning, 2015, pp. 2048–2057.
[21] M. Liu, J. Zhang, E. Adeli, D. Shen, Joint classification and regression via deep multi-task multi-channel learning for Alzheimer's disease diagnosis, IEEE Trans. Biomed. Eng. 66 (5) (2018) 1195–1206.
[22] M. Oquab, L. Bottou, I. Laptev, J. Sivic, et al., Weakly supervised object recognition with convolutional neural networks, in: NIPS, 2014, pp. 1545–5963.
[23] J. Feng, Z.-H. Zhou, Deep MIML network, in: Thirty-First AAAI Conference on Artificial Intelligence, 2017.
[24] M. Sun, T.X. Han, M.-C. Liu, A. Khodayari-Rostamabad, Multiple instance learning convolutional neural networks for object recognition, in: 2016 23rd International Conference on Pattern Recognition (ICPR), IEEE, 2016, pp. 3270–3275.
[25] J. Wu, Y. Yu, C. Huang, K. Yu, Deep multiple instance learning for image classification and auto-annotation, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2015, pp. 3460–3469.
[26] T. Tong, R. Wolz, Q. Gao, R. Guerrero, J.V. Hajnal, D. Rueckert, Multiple instance learning for classification of dementia in brain MRI, Med. Image Anal. 18 (5) (2014) 808–818.
[27] Y. Xu, J.-Y. Zhu, I. Eric, C. Chang, M. Lai, Z. Tu, Weakly supervised histopathology cancer image segmentation and classification, Med. Image Anal. 18 (3) (2014) 591–604.
[28] Z. Yan, Y. Zhan, Z. Peng, S. Liao, Y. Shinagawa, S. Zhang, D.N. Metaxas, X.S. Zhou, Multi-instance deep learning: discover discriminative local anatomies for bodypart recognition, IEEE Trans. Med. Imaging 35 (5) (2016) 1332–1343.
[29] M. Liu, J. Zhang, E. Adeli, D. Shen, Landmark-based deep multi-instance learning for brain disease diagnosis, Med. Image Anal. 43 (2018) 157–168.
[30] K. He, X. Cao, Y. Shi, D. Nie, Y. Gao, D. Shen, Pelvic organ segmentation using distinctive curve guided fully convolutional networks, IEEE Trans. Med. Imaging 38 (2) (2018) 585–595.
[31] C. Lian, M. Liu, J. Zhang, D. Shen, Hierarchical fully convolutional network for joint atrophy localization and Alzheimer's disease diagnosis using structural MRI, IEEE Trans. Pattern Anal. Mach.Intell. (2018).
[32] Y. Wang, B. Yu, L. Wang, C. Zu, D.S. Lalush, W. Lin, X. Wu, J. Zhou, D. Shen, L. Zhou, 3D conditional generative adversarial networks for high-quality PET image estimation at low dose, Neuroimage 174 (2018) 550–562.
[33] Y. Zhan, Y. Ou, M. Feldman, J. Tomaszeweski, C. Davatzikos, D. Shen, Registering histologic and MR images of prostate for image-based cancer detection, Acad. Radiol. 14 (11) (2007) 1367–1381.
[34] A. Mohamed, E.I. Zacharaki, D. Shen, C. Davatzikos, Deformable registration of brain tumor images via a statistical model of tumor-induced deformation, Med. Image Anal. 10 (5) (2006) 752–763.
[35] S. Bakas, M. Reyes, A. Jakab, S. Bauer, M. Rempfler, A. Crimi, R.T. Shinohara, C. Berger, S.M. Ha, M. Rozycki, et al., Identifying the best machine learning algorithms for brain tumor segmentation, progression assessment, and overall survival prediction in the BRATS challenge, arXiv preprint arXiv:1811.02629(2018).

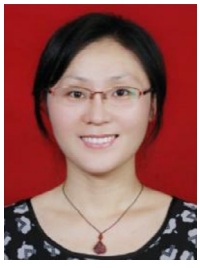
[36] K. He, J. Huo, Y. Shi, Y. Gao, D. Shen, MIDCN: a multiple instance deep convolutional network for image classification, in: Pacific Rim International Conference on Artificial Intelligence, Springer, 2019, pp. 230–243.
[37] R. Yang, X. Li, H. Liu, Y. Zhen, X. Zhang, Q. Xiong, Y. Luo, C. Gao, W. Zeng, Chest CT severity score: an imaging tool for assessing severe COVID-19, Radiology 2 (2) (2020) e200047, doi:10.1148/ryct.2020200047.
[38] K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition, 2015, arXiv:1512.03385.
[39] M. Ilse, J.M. Tomczak, M. Welling, Attention-based deep multiple instance learning, (2018). arXiv preprint:1802.04712.

Kelei He received the Ph.D. degree in computer science and technology from Nanjing University, China. He is currently the assistant dean of National Institute of Healthcare Data Science at Nanjing University. He is also an assistant researcher of Medical School at Nanjing University, China. His research interests include medical image analysis, computer vision and deep learning.

Wei Zhao received the Ph.D. degree in imaging and nuclear medicine from Fudan University, China. He is a radiologist of The Second Xiangya Hospital. His research interests include chest CT imaging, radiomics and deep learning.

Xingzhi Xie, is a graduate student and resident of radiology department of the Second Xiangya Hospital. Her research interests include brain functional imaging, radiomics and deep learning.

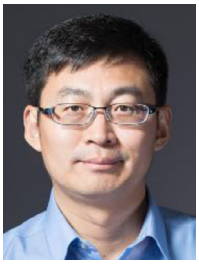
Wen Ji, is a master candidate of the Department of Computer Science and Technology, Nanjing University. Her research interest include medical image analysis, computer vision and deep learning.



Mingxia Liu received the Ph.D. degree from the Nanjing University of Aeronautics and Astronautics, Nanjing, China, in 2015. She is a Senior Member of IEEE (SM'19). Her current research interests include machine learning, pattern recognition, and medical image analysis.

Zhenyu Tang is an associate professor in Beijing Advanced Innovation Center for Big Data and Brain Computing, Beihang University. He received his Ph.D. degree of Computer Engineering in the University of Duisburg-Essen in 2011, and he worked as post-doc in the University of North Carolina at Chapel Hill and Automation institute of Chinese Academy of Science, respectively. His research interests include medical image analysis, computer vision, and pattern recognition.

Yinghuan Shi is an Associate Professor in the Department of Computer Science and Technology of Nanjing University, China. His research interests include computer vision and medical image analysis.



Feng Shi received his Ph.D. degree from Institute of Automation, Chinese Academy of Sciences. He has been an Assistant Professor at the University of North Carolina at Chapel Hill, NC, and Cedar Sinai Medical Center at Los Angeles, CA. He is currently a Research Scientist in Shanghai United Imaging Intelligence, China. His research interests are the projects that involve image processing and artificial intelligence techniques to develop computer-assisted clinical decision support systems.

Yang Gao received the Ph.D. degree from the Department of Computer Science and Technology, Nanjing University, Nanjing, China, in 2000. He is currently a Professor with the Department of Computer Science and Technology, Nanjing University. He has authored more than 100 papers in top conferences and journals in and outside of China. His current research interests include artificial intelligence and machine learning.

Jun Liu, Professor, director of radiology department of the Second Xiangya Hospital. He is also the leader of 225 subjects in Hunan Province, National member of the Neurology Group of the Chinese Society of Radiology, National Committee of the Neurology Group of the Radiological Branch of the Chinese Medical Association. His research interests include brain functional imaging, radiomics and deep learning.

Junfeng Zhang, Professor, the dean of Medical School of Nanjing University, and the dean of National Institute of Healthcare Data Science at Nanjing University. His research interests including medical image analysis, biomedical engineering and medical chemistry.

Dinggang Shen, Professor, IEEE Fellow, AIMBE Fellow, IAPR Fellow. His research interests include medical image analysis, computer vision, and pattern recognition. He has published more than 1000 papers in the international journals and conference proceedings, with h-index of 105. He serves as an editorial board member for eight international journals, and was General Chair for MICCAI 2019.