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.

Information Sciences 612 (2022) 745–758

Contents lists available at ScienceDirect

Information Sciences

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

An efficient deep neural network framework for COVID-19 lung infection segmentation

Ge Jin [a], Chuancai Liu [a,b,*], Xu Chen [a]

[a] School of Computer Science and Engineering, Nanjing University of Science and Technology, Nanjing 210094, China
[b] Collaborative Innovation Center of IoT Technology and Intelligent Systems, Minjiang University, Fuzhou 350108, China

ARTICLE INFO

Article history:
Received 29 October 2021
Received in revised form 11 August 2022
Accepted 13 August 2022
Available online 2 September 2022

Keywords:
COVID-19
Infection segmentation
VQ-VAE
Proportions loss
Semi-supervised learning
Adversarial network

ABSTRACT

Since the outbreak of Coronavirus Disease 2019 (COVID-19) in 2020, it has significantly affected the global health system. The use of deep learning technology to automatically segment pneumonia lesions from Computed Tomography (CT) images can greatly reduce the workload of physicians and expand traditional diagnostic methods. However, there are still some challenges to tackle the task, including obtaining high-quality annotations and subtle differences between classes. In the present study, a novel deep neural network based on Resnet architecture is proposed to automatically segment infected areas from CT images. To reduce the annotation cost, a Vector Quantized Variational AutoEncoder (VQ-VAE) branch is added to reconstruct the input images for purpose of regularizing the shared decoder and the latent maps of the VQ-VAE are utilized to further improve the feature representation. Moreover, a novel proportions loss is presented for mitigating class imbalance and enhance the generalization ability of the model. In addition, a semi-supervised mechanism based on adversarial learning to the network has been proposed, which can utilize the information of the trusted region in unlabeled images to further regularize the network. Extensive experiments on the COVID-SemiSeg are performed to verify the superiority of the proposed method, and the results are in line with expectations.

1. Introduction

Since the first person was diagnosed with Coronavirus disease in 2019 (COVID-19) [1], there have been more than 170 million confirmed cases and more than 3.541 million deaths as of May 31, 2021 [2]. The disease has since spread worldwide, leading to an ongoing pandemic.

Symptoms of COVID-19 are variable, the most common are cough, fever, headache, loss of smell and taste, and breathing difficulties. The virus affects multiple organs in the human body, and the lung is the ground zero for the virus affection. Computed Tomography (CT) imaging has become a critical means to detect lung tissue associated with the virus, where segmentation of the infection regions is important for the subsequent assessment. In the past, studies observed that radiological imaging is effective in the inchoate screening of COVID-19 [3]. The manual segmentation of the lesions from CT images is time-consuming and requires a large expenditure of labor, whereas, manual annotation is a subjective task, and its accuracy is affected by the physicians' clinical experience and personal bias. Thus, the automatic segmentation of the lesions is highly desirable in clinic practice.

* Corresponding author at: School of Computer Science and Engineering, Nanjing University of Science and Technology, Nanjing 210094, China.
E-mail address: chcailiu@163.com (C. Liu).

https://doi.org/10.1016/j.ins.2022.08.059
0020-0255/

Recently, deep learning has achieved great success in the field of medical image analysis [4]. Many robust models have been applied to segment COVID-19 infected area from CT images, including classic U-Net [5], Unet++ [6], and DenseUnet [7]. To obtain a robust segmentation network, the model requires sufficient annotated labels. Shan et al. [8] proposed a method that involved interactivity with experts in the training process of the network. Moreover, Zheng et al. [9] proposed an unsupervised framework, which can generate pseudo labels to improve the performance of the model. Although many outstanding works have been proposed to tackle the COVID-19 pneumonia lesion segmentation task, there are still many challenges that need to be solved. First, it is difficult to segment the corresponding lesion area correctly due to the variety of the shape, texture, and position of the lesion. Second, the infection lesions have different complex characteristics, for example, Ground-Glass Opacity (GGO), consolidation, etc. The inter-class gap between them is small, the boundary is blurred, and the contrast of the lesion compared with the normal area is low. These challenges make it difficult for even experts to segment the area of the lesions accurately, let alone automatic segmentation without manual intervention. Finally, for the above reasons, it is difficult to obtain sufficient high-quality labels in practice and the acquisition of annotations is time-consuming and expensive.

To address the aforementioned issues, the VQ-VAE [10] branch is introduced to reconstruct the input images. The branch can regularize the shared decoder. The generated latent maps with rich semantic information are fed into the decoder to further enhance the feature representation. Moreover, the proportion loss is proposed to mitigate class imbalance, which encourages the label marginal to match target class proportions. Furthermore, a novel semi-supervised framework is proposed which is based on adversarial learning to alleviate the shortage of high-quality labels. Specifically, the training process includes two phases: fully-supervised learning and semi-supervised learning. The first phase leverages labeled images to train the model, and the second part utilizes unlabeled data to provide supervised signals. The two phases are performed simultaneously. The whole training process will be detailed in Section 3. In a nutshell, the contributions and novelty of the present study are highlighted as follows:

(1) A novel network for COVID-19 infected area segmentation from CT slices is presented. The framework integrates a VQ-VAE branch to reconstructs the input image into itself, and the shared encoder is regularized by the module. It can not only improve the performance but also help to consistently obtain reliable performance for any random initialization.
(2) The proportion loss is proposed to mitigate the class imbalance, but without losing generality. It can effectively induce the proportions of the label to match the target. Also, the training processing can be stabilized by the proposed loss.
(3) A novel semi-supervised framework based on adversarial learning is proposed. It can not only make full use of the labeled images, but also utilized the supervision signals generated by unlabeled images to guide the training process.

The sections of this paper are organized as follows. The second section introduces the related work. Furthermore, Section 3 provides a detailed explanation of the proposed method. The experimental results and performance analysis are introduced in Section 4. Finally, we present the conclusion in Section 5.

2. Related work

Some works related to our research will be discussed in this section, including medical image segmentation, loss function for semantic segmentation, and artificial intelligence techniques for COVID-19.

2.1. Medical image segmentation

In computer-aided diagnosis, medical image segmentation is a fundamental and challenging task, aiming to accurately recognize the target regions (e.g., organs, tissues, lesions). In recent years, convolutional neural network have dominated the field of image segmentation, and have been widely employed in medical image processing.

Fully supervised learning is the most widely used technology for medical image segmentation tasks, which requires adequate high-quality annotations. One of the most prominent works is U-Net, which is proposed by Ronneberger et al. [11]. The model includes an encoder for feature extraction and an asymmetrical decoder for restoring spatial resolution and generating segmentation results. Features at different levels are merged through skip-connections, which improves segmentation accuracy. Furthermore, Zhou et al. [12] improved the multi-layer features fusion method and proposed U-Net++, which adds a nested convolutional structure between the encoder and decoder. The methods mentioned above are all for 2D data, but most medical image data type are 3D. To solve this problem, Çiçek et al. [13] present the 3D U-Net that utilizes the inter-slice knowledge by replacing the layers with a 3D version. Meanwhile, the attention mechanism is introduced into some works [14] to re-weight the features to strengthen the effective characteristic and suppress the ineffective features.

To train a robust segmentation network, sufficient annotations are essential. As manual delineation for lesions is label-intensive and time-consuming, adequate high-quality annotations are often unprocurable for medical image segmentation. To mitigate the dependence on annotations and reduce costs, semi-supervised learning algorithms have gained extensive attention and research. Research scholars in Ref. [15]16 applied self-taught algorithm to spread useful information obtained by labeled data through manifold assumptions to generate pseudo-labels and perform iterative optimization. The drawback of this type of method is that the performance highly depends on the quality of the generated pseudo-labels. If the model

learns the failing label, it may continuously amplify it and affect the final prediction. Bai et al. [16] combined the post-processing algorithm with the self-training to segment the left ventricular MRI image. This method first learns the labeled data and segments the unlabeled image. Then, the conditional random field is introduced to refine the predictions, which are leveraged to guide the next iteration. Rajchl et al. [15] also applied the self-taught and additionally used the bounding-box level annotations to assist the supervision process. To reduce the limitations of single model prediction, the co-training algorithm is introduced to use multiple pre-trained models to comprehensively predict pseudo-labels. Peng et al. [17] used the mean values predicted by multiple models as pseudo labels, and introduced adversarial samples to capture the differences between different models to make the models learn more complementary knowledge. The various alternative methods can adjust the process of learning pseudo labels by introducing additional constraints to improve the utilization efficiency of the pseudo labels. Andriy [18] proposed a model that integrates variational auto-encoder(VAE) to mitigate insufficient training data, and won 1st place in the BraTS. 2018 challenge.

2.2. Loss function for semantic segmentation

The loss function is highly important for designing comprehensive image segmentation-based frameworks. Many researchers have experimented with diversified domain-specific loss functions to improve performance for specific datasets. The most widely applied loss function is cross-entropy [19], which is defined as measuring the difference between two probability distributions for a given random variable or set of events. To address the issue of unbalanced data categories, the weighted binary cross-entropy [20] was proposed to weight specific classes. The Focal Loss [21] is also applicable to imbalanced class scenarios, which can induce the model to focus more on hard examples and down-weights the contribution of easy examples. The dice coefficient is a commonly applied metric for evaluating the performance of medical image segmentation. The V-Net [22] proposed the Dice Loss based on the Dice coefficient. Tversky index [23] is a generalization of the Dice coefficient. It weights the false positives and false negatives with a hyper-parameter. There are some distance-based losses [24] that have also been proposed recently, and they are generally sensitive to boundary and contour information, and are suitable for medical image segmentation. Some other loss functions attempt to leverage structural priors such as CRF, and Generative Adversarial Networks(GANs) to supplement the information obtained by the model. Zhao et al. [25] proposed a Structural Similarity Loss (SSL) to get a high positive linear correlation between the labels and the predictions. In a nutshell, each type of loss has its merit and disadvantage. Therefore, some researchers [26] combined the different types of loss so that the compound loss can maximize its strengths and avoid weaknesses.

2.3. Artificial intelligence techniques for COVID-19

During the outbreak of COVID-19, many artificial intelligence(AI) technologies was proposed against the virus. Compared to the traditional imaging workflow that heavily relies on human labors, AI enables more safe, accurate, and efficient imaging solutions. Medical imaging such as X-ray and computed tomography (CT) [27] plays an essential role in this war without smoke. There are several essential contactless imaging workflows proposed in [28]. These techniques are more flexible installation and more accessible to the patient. Scientists in [29] proposed a self-contained mobile based on artificial intelligence for pre-scanning and diagnosis systems.

The applications for COVID-19 can be divided into three categories, including patient scale (e.g., medical imaging for diagnosis [30]), molecular scale (e.g., protein structure prediction [31], and societal scale (e.g., epidemiology [32]). The patient scales are mainly focused on the present study [33]. Whereas, automatic segmentation is the most challenging task, which is extremely important for the assessment and quantification of COVID-19. Fan et al. [34] proposed a two-step training scheme for multiclass COVID-19 infection segmentation. The authors first applied a convolutional neural network(CNN) to generate the binary lesion segmentation, and the second network utilizes the results to classify the lesion voxels into ground-glass opacity (GGO) and consolidation. They also introduced a novel semi-supervised learning method, which leverages a few labeled images to generate pseudo labels. CovidENet [35] is an ensemble of 2D and 3D CNNs based on AtlasNet [36] for total lesion segmentation, and the performance of the model is in line with the expectation. The U-shape types are the most common architectures applied in medical segmentation, which have achieved reasonable segmentation results in COVID-19 applications. Furthermore, Shan et al. [37] proposed the VB-Net, which utilizes the bottleneck blocks to obtain a more efficient segmentation. Recently, attention mechanisms have been widely applied in various applications. Oktay et al. [38] combined attention and UNet to capture fine structures of the medical images. Whereas, authors in [39] applied the technology to the segmentation of COVID-19 infected areas and achieved considerable performance.

3. Method

In the following, the overall structure of the whole network would be firstly described. Then, the details of the proposed methods are introduced, including VQ-VAE module, proportions loss and semi-supervised learning.

3.1. Network architectures overview

The proposed model is based on adversarial learning, which consists of two parts, ie., a segmentation network, and a discriminator network. The segmentation network comprises a backbone network and a VQ-VAE branch. The workflow of the proposed approach is displayed in Fig. 1.

There are two separated branches in the segmentation network processing the input images. The backbone is a modified Resnet-based network, and it could be any robust model designed for semantic segmentation, e.g., FCN, Unet, and PSPNet. The feature maps of the intermediate layer are fed into the VQ-VAE branch to generate reconstructions of input images, which can regularize the shared network portions. Moreover, the top and bottom levels of the latent maps generated by VQ-VAE are fused and concatenated to the intermediate feature maps generated by the backbone network, and it can further enhance the feature representation of semantics. The model is trained on the COVID-SemiSeg dataset [34], which is composed of 50 multi-class labels by doctors and 1600 unlabeled images. A novel semi-supervised framework is introduced to make full use of limited labeled images and large amounts of unlabeled images. In the initial stage of training, only the labeled images are utilized to update the model. When the loss tends to be saturated, the unlabeled images are introduced and the coarse pseudo labels of the images are generated. Then, the discriminator is applied to filter out the untrusted region of the coarse pseudo labels and generate refined pseudo labels to further enhance the performance of the model. The entire self-taught process is completed simultaneously without human intervention.

The discriminator is designed in a fully convolutional manner, which indicates the untrusted regions of the prediction and discards them. It takes a prediction generated from the segmentation network and yields a confidence map of the same size as the prediction. Every pixel of the confidence map indicates the probability of input data sampled from the ground truth distribution. From Fig. 1, for the output of the discriminator, the brighter part is, the more trusted.

The training phase can be split into two stages, the first stage is fully supervised learning, and then semi-supervised learning is executed. For the first phase, the segmentation network is supervised by L_{comp}, which consists of two parts, namely, standard cross-entropy loss L_{ce} and L_{prop} (proportion loss). The L_{prop} is proposed to encourage label marginals to match target

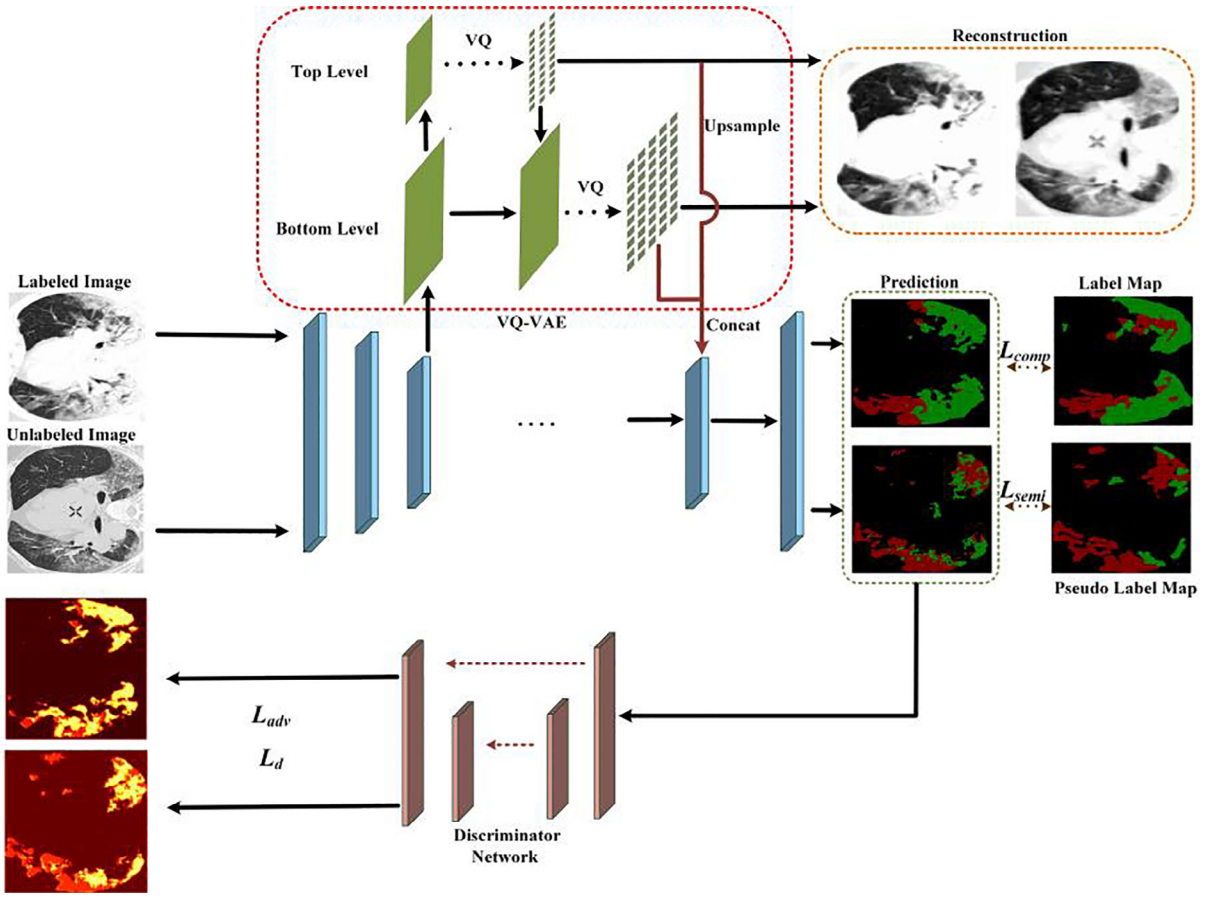

Fig. 1. Overview of the proposed model.

class proportions. Then after several training iterations, the segmentation network takes unlabeled images to generate pseudo labels, which are applied to further improve the performance of the model. Furthermore, adversarial learning is introduced in the training scheme. In this phase, the segmentation network is supervised by $L_{\text{semi}}, L_{ce}, L_{\text{adv}}$, and L_{prop}, where L_{semi} is the cross entropy between the unlabeled images and the pseudo label, and L_{ce} is the CE between the labeled data and the labels. The labeled and unlabeled images are all fed into segmentation to generate predictions, and then the discriminator takes the predictions to generate confidence maps, which are used to refine the coarse pseudo labels. Moreover, the L_{adv} is applied to fool the discriminator by maximizing the probability of the predicted results being generated from the ground truth distribution. Note that, the discriminator is trained with the labeled data by L_{d}. The detail will be described in Section 3.4.

3.2. VQ-VAE module

A novel VQ-VAE module is proposed to improve the performance of the model, which is inspired by [10]. The segmentation network follows encoder-decoder architecture with an asymmetrically Resnet-based encoder and a smaller decoder. As the training dataset size is limited, the VQ-VAE branch is added to the endpoint of the encoder to reconstruct the input images. It can provide additional guidance to the training process and impose regularization on the shared encoder, which induces the encoder to extract more representative features. In particular, a hierarchy of vector quantized codes is applied to reconstruct the input images. The branch generates a top-level latent code that contains global information such as the shape and geometry of objects, and a bottom-level latent code representing the local details such as texture. Furthermore, the two levels of representation are integrated into a comprehensive feature by up-sampling and concatenate operations. Then, the fused features are fed into the subsequent network to further improve the performance of the model.

The VQ-VAE can be classified as the VAE family, which has an encoder, code, and decoder. The difference is that the code is not directly generated by the encoder, but is obtained through vector quantization. The module can reconstruct more coherence and fidelity images with less resource consumption. Whereas, the model is based on likelihood, in principle, it covers all modes of the data and can capture the diversity of the true distribution. Thus, the aforementioned issues inspired us to apply the VAE-based branch to improve the expression of features and the robustness of the model.

The branch consists of an encoder and a decoder, and a shared codebook. The encoder maps the observations to a series of discrete latent variables, and the decoder reconstructs the observations from these discrete variables. The input x is transformed into a vector $E(x)$ by the nonlinear mapping of the encoder, and the vector is quantized according to its distance from the prototype vectors in the codebook $\mathrm{e}_k, k \in 1 \ldots K$ such that each vector $E(x)$ is replaced by the index of the nearest prototype vector in the codebook. Finally, the indices are fed into the decoder and mapped back to their corresponding vectors in the codebook, from which it reconstructs the data via another non-linear function.

$$Quantize(E(x)) = e_k \text{ where } k = \arg\min_j \|E(x) - e_j\| \tag{1}$$

Following [10], the objective function of the VQ-VAE contains three terms:

$$L(x, D(e)) = \|x - D(e)\|_2^2 + \|sg[E(x)] - e\|_2^2 + \beta\|sg[e] - E(x)\|_2^2 \tag{2}$$

where e is the quantized code for the training example x. The first part is the data fidelity term, which makes the reconstruction error as small as possible. The other two additional terms are ingeniously designed to align the vector space of the codebook with the output of the encoder. The second part is applied to the codebook, where $sg[\cdot]$ means stop gradient, that is, backward gradient transfer is not executed. It makes the selected codebook e close $E(\mathrm{x})$, which denotes the output of the encoder. The last term is effective for the encoder, which induces the $E(\mathrm{x})$ to stay close to the chosen codebook vector and prevents the parameters of the encoder from fluctuating too frequently. In this paper, the $\beta$is set to 0.25.

3.3. Proportions loss

Semantic segmentation is a pixel-wise classification task, and choosing a suitable loss function is extremely important for the task. Most work applies variants of the Cross-Entropy(*CE*) or Dice loss as their objective function. The medical image segmentation task generally faces the problem of class imbalance, i.e., the proportion of segmented regions varies greatly and may cause large-region terms in the objective to completely dominate small-region ones. To tackle this issue, some work focus on designing a novel architecture or training schemes, yet, the loss function to be minimized during training plays a critical role. [26] demonstrate that *CE* and Dice share a very deep connection, and Dice is intrinsically more preferring small regions, while *CE* implicitly encourages the ground-truth region proportions. The difference between the two types of losses mentioned above is called label-marginal biases. In this paper, a novel way is proposed to solve the problem.

We first present the analysis of *CE*, and then introduce the proposed method. Let F and K denote the random variables associated with the learned features and the labels, and $H(F|K)$ is the conditional entropy of learned features given the labels. Note that, the F is continuous, while the K is discrete with a random variable sampled from a finite set $\{1, \ldots, K\}$. Following [40], the marginal distribution of the labels could be denoted as follow:

$$\mathrm{P}(K=k) \approx y_k \wedge = \frac{|\Omega_k|}{|\Omega|} \tag{3}$$

where the y_k denotes the GT proportion of region k, the Ω indicates the spatial image domain, and Ω_k denotes the GT proportion of the k_{th} class. Therefore, the conditional entropy of the learned features could be expressed as follows:

$$H(F|K) = \sum_{k}^{K} P(K=k) H(F|K=k) \approx \frac{1}{|\Omega|} \sum_{k}^{K} |\Omega_k| H(F|K=k) \tag{4}$$

with each $\mathrm{H}(F|K=k)$ given by:

$$H(F|K=k) = -\int_{f^\theta} P\left(f^\theta|K=k\right) \log P\left(f^\theta|K=k\right) df^\theta \tag{5}$$

where f^θ denotes feature embedding. Then, the conditional entropy in Eq. (5) is estimated by well-known Monte-Carlo estimation, which is shown as follow [41].

$$\int_{\mathrm{f}} g(f) P(f|S) \approx \frac{1}{|S|} \sum_{\mathrm{i} \in S} g(f_i) \tag{6}$$

where $g(f_i)$ denotes a feature vector at point i, and the S is a discrete set of points which belongs to Ω. The $g(\cdot)$ is an arbitrary function, and $P(f|S)$ expresses the density of $g(f_i)$.

Then, Eq. (6) is applied to $H(F|K=k)$, and Eq. (4) could be denoted as follows:

$$H(F|K) \approx -\frac{1}{|\Omega|} \sum_{k}^{K} \sum_{i \in \Omega_k} \log(P(f_i^\theta|k)) \tag{7}$$

Furthermore, following the Bayes rule $P(\mathrm{f}_i^\theta|k) \propto \frac{p_{ik}}{p_k \wedge}$. $p \wedge_k = \frac{1}{|\Omega|} \sum_{i \in \Omega} p_{ik}$ denotes the predicted probability of class k, and $p_{ik} = P(k|f_i^\theta)$ is the softmax predictions at the pixel. Thus, the $H(F|K)$ could be re-written as follow:

$$\begin{aligned} H(F|K) \quad &\approx -\tfrac{1}{\Omega} \sum_{k}^{K} \sum_{i \in \Omega_k} \log(\frac{p_{ik}}{p_k \wedge}) \\ &= CE + \sum_{k}^{K} y_k \wedge \log(p_k \wedge) \end{aligned} \tag{8}$$

Due to the definition of the label-marginal KL divergence, we have:

$$D_{KL}(y||p) = \sum_{k=1}^{K} y_k \wedge \log(\frac{y_k \wedge}{p_k \wedge}) \stackrel{c}{=} -\sum_{k}^{K} y_k \wedge \log(p_k \wedge) \tag{9}$$

where $\stackrel{c}{=}$stands for equality up to an additive and/or nonnegative multiplicative constant, and y denotes GT label-marginal probability, yet, the probability of the predicted label-marginal is denoted as p

Finally, the CE is denoted as:

$$CE \stackrel{c}{=} H(F|K) + D_{KL}(y||p) \tag{10}$$

The first part is the matching of GT, and the second part is the label-marginal bias. From this formula, the label-marginal bias is a hidden term, and it can induce the proportions of the predicted segmentation regions to match the ground-truth proportions. The KL term also can be viewed as a regularization term, which encourages low uncertainty within each ground-truth segmentation region. If CE only contains the entropy term, it may lead to trivial imbalanced solutions. However, the two competing terms are implicit so that the contribution of the related parts cannot be controlled. It is evident that the contribution is significant in imbalanced problems. In addition, the label-marginal can mitigate the difficulty that the ground-truth matching terms differ by several orders of magnitude across regions. Appropriate label-marginal terms can effectively avoid large-region terms dominating small-region ones. Thus, the proportions loss is proposed to control explicitly the label-marginal bias. We applied CE, proportion loss L_{prop} to form a novel compound loss to update the segmentation network.

$$L_{comp} = L_{\mathrm{ce}} + \lambda(y-p)^2 \tag{11}$$

where $\lambda$is a hyper-parameter, which is set to 0.5 in the present study. From the function, the second part is L_{prop}, which can be viewed as a regularization term to encourage the prediction proportions to match the ground-truth proportions. The solution is simple but effective, especially in the scenario of class imbalance in medical image segmentation. Fig. 2 shows visual comparisons of the segmentation results, which demonstrate the effectiveness of the proposed method.

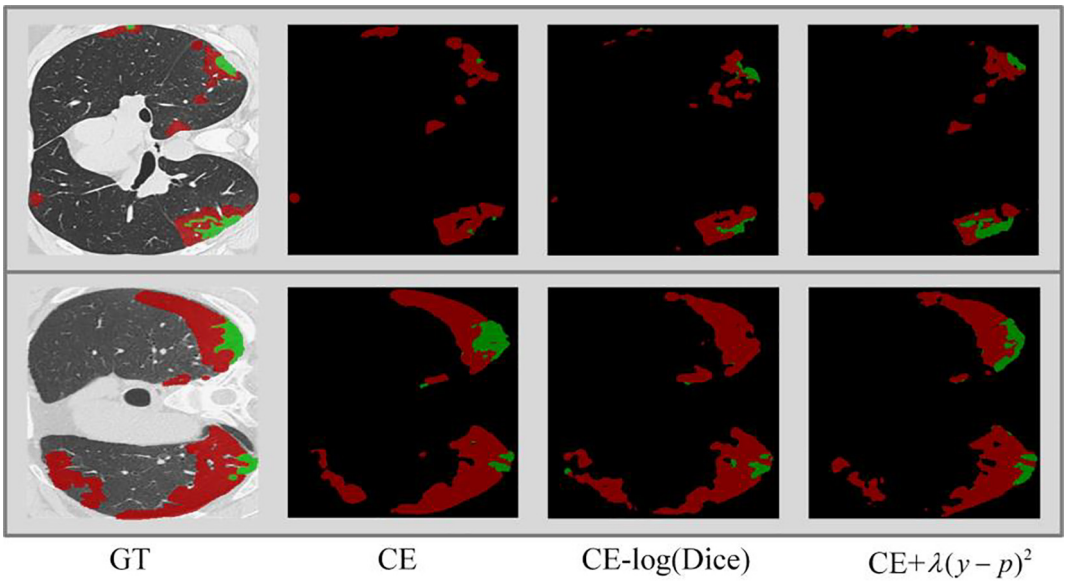

Fig. 2. The visual comparison between the proposed method and various losses. The color masks denote the COVID-19 infected regions in CT axial slice, where the red and green denote the GGO and consolidation respectively. The images in the left-most column are the ground-truth. The examples show that our method can achieve the best performance. The proposed method is susceptible to the segmentation regions of both classes. However, for the first example, CE basically did not identify the consolidation region. Moreover, the combinations of the CE and Dice also lost some details.

3.4. Semi-supervised learning

To alleviate the shortage of annotations, a semi-supervised learning scheme is introduced, which only needs a few annotations and make full use of the rich unlabeled images. Furthermore, we design an adversarial framework to enhance the fitting and generalization ability of the model. The whole model contains two branches, which are used for fully supervised learning and semi-supervised learning.

In the early stages of training, the labeled data is only utilized as the training source data, which can be viewed as the fully supervised learning phase. Moreover, the segmentation network is supervised by L_{comp} in this stage, while adversarial learning has not been introduced into the training scheme. Then, after the model is trained for several rounds, we introduce semi-supervised adversarial learning, which not only applies the L_{comp} as the objective function but also introduces an adversarial loss to further improve the performance of the model. In this phase, the dual-branches are executed simultaneously. specifically, the L_{semi}, L_{prop}, and L_{adv} are applied to the semi-supervised branch, meantime, L_{adv} is also introduce in fully supervised learning. The adversarial loss is displayed as follows:

$$L_{\text{adv}} = -\sum_{h,w} \log(D(S(I_n))^{(h,w)}) \tag{12}$$

where I_{n} denotes input images, $S(\cdot)$ represents the segmentation network, and $D(\cdot)$ is the discriminator network. In practice, the label of the discriminator is a mask filled with 1 and the same size as the original images. For the discriminator, if the input data is more likely to be sampled from ground truth, the corresponding pixel of the confidence map, ie., the output of the discriminator would tend to be 1. The probability of the prediction would be maximized by this loss to confuse the discriminator, which can encourage the model to generate predictions close to the ground truth distribution.

When the segmentation network is trained for several rounds, and the objective function tends to saturate. The unlabeled images are introduced to regularize the segmentation network. The model takes the labeled and unlabeled images simultaneously, and the branch with labeled images continues to perform fully supervised learning. For the semi-supervised learning branch, the segmentation network takes the unlabeled images to generate predictions, that is, coarse pseudo labels, which are then fed into the discriminator to obtain the confidence maps. Moreover, we set a threshold T for the elements in the confidence map and the regions greater than T are assumed to be the confidence area. In practice, we found that by setting T to [0.1, 0.3], the model can get reasonable results. On the contrary, the region smaller than the threshold would be zeroed, which would not participate in the gradient transfer process. Finally, the confidence areas of coarse pseudo labels are retained and obtained the refined pseudo labels, ie., the label of the unlabeled images.

$$Pseudo - Label = I(D(S(U_n))^{(i)} > T) \cdot \arg\max_{c}(S(U_n)) \tag{13}$$

where $I(\cdot)$ denotes the indicator function. The semi-supervised learning is a self-taught process, which is fully automatic without manual intervention. The pseudo labels of the unlabeled images are gradually refined and rationalized, whereby progressively imposing beneficial regularization to the model. It should be noted that, the discriminator is supervised by binary spatial cross-entropy loss.

$$L_D = -\sum_{i \in X_n} (1 - y_n) \log(1 - D(S(X_n))^{(i)}) + y_n \log(D(Y_n)^{(i)}) \tag{14}$$

where y_n takes two values, 0 and 1. If the data is sampled from the ground truth distribution, $y_n = 1$, otherwise $y_n = 0$. The certain pixel of the input image is denoted as i.

4. Experimental Results

In this section, we will discuss the details of the proposed method for the comprehensive experiments on the COVID-19 infection segmentation dataset (COVID-SemiSeg). First, a brief introduction of the dataset, evaluation metrics and implementation details are presented. Moreover, an ablation study is performed to verify the predominance of the proposed method. The experimental results show that our proposed method can reach and surpass the state-of-the-art models.

4.1. Evaluation datasets and metric

The performance of the proposed methods is verified on the COVID-SemiSeg [34], which contains 100 labeled CT images collected from [42], and 45 CT images are used as training data, 5 CT images for validation, meanwhile, the rest part for testing. Furthermore, the dataset also leverages large-scale unlabeled CT images as a supplement to training samples, whereby applied in semi-supervised learning. There are 1600 unlabeled CT images extracted from the COVID-19 CT Collection [43], which comprises 20 CT volumes from different COVID-19 patients. In conclusion, this dataset is highly suitable for applying the proposed method.

To evaluate our model more comprehensively, six metrics are introduced for evaluation. In addition to using the three commonly adopted metrics, i.e., the Dice similarity coefficient, Sensitivity (Sen.), and Specificity (Spec.), we also introduce another three golden metrics from the object detection field, i.e., Structure Measure(S_α) [44], Enhance-alignment Measure (E_ϕ^{mean}) [45], and Mean Absolute Error(MAE). The S_α can measure the structural similarity between a prediction and the label, and the E_ϕ^{mean} can evaluate the local and global similarity between two binary masks. The last one is a measure of the pixel-wise error between the prediction and ground-truth mask.

4.2. Implementation details

We implemented our network in Pytorch framework. To train the segmentation network, we leverage the stochastic gradient descent(SGD) optimization method, and the learning rate is set to 0.01. The momentum of the SGD optimizer is set to 0.9, and the weight decay is set to 10^{-4}. Meanwhile, to train the discriminator, we use the Adam optimizer with the learning rate set to 10^{-4}. In the fully-supervised phase, the L_{comp} is the only objective function to update the segmentation network, and the model is trained for 1000 iterations in this stage. Then, the model goes into the semi-supervised mode, the dual branches are executed simultaneously. Moreover, L_{semi} and L_{adv} are weighted in the objective function, their weight parameters are 0.1 and 0.01 respectively. In the process of adversarial learning, the segmentation network and the discriminator are updated alternately. Specifically, the parameters of one network are not updated while the other network is training.

4.3. Segmentation results

In this work, our model is experimented on COVID-SemiSeg dataset, and the proposed algorithm is compared against the FCN [46], Deeplabv3+ [47], U-Net [11], U-Net++ [12], Attention UNet [38], R2U-Net [48], PSPNet [49], MA-Net [50], to demonstrate that our method performs comparably with the state-of-the-art model. The results are shown in Table 1.

As can be seen in Table 1, the proposed method outperforms the other models in terms of six metrics. For Dice, our method surpasses the two state-of-the-art models MANet and U-Net++, by 9.6% and 18.5% on average segmentation result respectively. Due to regions of the consolidation are generally small, and the contrast with the surrounding area is low, the boundary is usually blurred. Thus, it is more challenging to segment it correctly. Whereas, the proposed method can achieve the best performance. As presented in Table 1, the proposed method surpasses the Attention UNet by almost 16% in terms of Dice. In short, the proposed method achieves competitive performance in most evaluation metrics.

The visual comparison of COVID-19 infection segmentation results is shown in Fig. 3, indicating that the proposed method's performance surpasses other methods remarkable. It can be seen from Fig. 3 that U-Net cannot obtain a satisfactory result, and there are a large number of mis-segmented tissues exist. The Attention U-Net get better performance, but the results are still not promising. Moreover, the segmentation regions of the aforementioned method are not coherent. Although the segmentation results of FCN are smooth, it also generates a lot of mis-segmentation. The success of the

Table 1
Quantitative results of ground-glass opacities and consolidation on the COVID-SemiSeg dataset. The best and the second results are highlighted in red and blue.

	Ground-Glass Opacity						Consolidation						Average					
Methods	Dice	Sen.	Spec.	S_α	E_ϕ^{mean}	MAE	Dice	Sen.	Spec.	S_α	E_ϕ^{mean}	MAE	Dice	Sen.	Spec.	S_α	E_ϕ^{mean}	MAE
FCN[46]	0.480 ± 0.03	0.450 ± 0.01	0.910 ± 0.01	0.582 ± 0.02	0.754 ± 0.01	0.101 ± 0.03	0.283 ± 0.02	0.268 ± 0.01	0.714 ± 0.02	0.554 ± 0.02	0.561 ± 0.01	0.051 ± 0.01	0.382 ± 0.03	0.359 ± 0.01	0.812 ± 0.02	0.568 ± 0.02	0.658 ± 0.01	0.076 ± 0.02
Deeplabv3+[47]	0.462 ± 0.03	0.591 ± 0.03	0.912 ± 0.03	0.551 ± 0.01	0.690 ± 0.02	0.079 ± 0.02	0.201 ± 0.04	0.289 ± 0.04	0.685 ± 0.05	0.598 ± 0.01	0.610 ± 0.01	0.049 ± 0.01	0.332 ± 0.03	0.440 ± 0.03	0.799 ± 0.04	0.565 ± 0.01	0.650 ± 0.01	0.064 ± 0.01
U-Net[11]	0.473 ± 0.02	0.530 ± 0.03	0.944 ± 0.02	0.577 ± 0.01	0.743 ± 0.01	0.098 ± 0.01	0.274 ± 0.02	0.302 ± 0.01	0.673 ± 0.02	0.640 ± 0.01	0.805 ± 0.02	0.048 ± 0.02	0.374 ± 0.02	0.416 ± 0.02	0.809 ± 0.02	0.609 ± 0.01	0.774 ± 0.01	0.073 ± 0.01
U-Net++[12]	0.488 ± 0.02	0.542 ± 0.03	0.952 ± 0.02	0.580 ± 0.01	0.787 ± 0.02	0.080 ± 0.01	0.280 ± 0.02	0.311 ± 0.01	0.745 ± 0.02	0.615 ± 0.03	0.820 ± 0.01	0.046 ± 0.01	0.384 ± 0.02	0.427 ± 0.02	0.849 ± 0.02	0.598 ± 0.01	0.740 ± 0.02	0.063 ± 0.01
Attention UNet[38]	0.491 ± 0.02	0.539 ± 0.02	0.961 ± 0.04	0.576 ± 0.01	0.792 ± 0.02	0.082 ± 0.01	0.279 ± 0.01	0.320 ± 0.01	0.752 ± 0.04	0.597 ± 0.03	0.811 ± 0.01	0.049 ± 0.01	0.385 ± 0.02	0.430 ± 0.01	0.857 ± 0.04	0.587 ± 0.01	0.802 ± 0.01	0.066 ± 0.01
R2U-Net[48]	0.410 ± 0.04	0.481 ± 0.02	0.875 ± 0.02	0.554 ± 0.03	0.716 ± 0.01	0.097 ± 0.01	0.193 ± 0.02	0.284 ± 0.02	0.693 ± 0.01	0.495 ± 0.04	0.554 ± 0.01	0.049 ± 0.02	0.302 ± 0.03	0.383 ± 0.02	0.784 ± 0.01	0.525 ± 0.03	0.635 ± 0.01	0.070 ± 0.01
PSPNet[49]	0.500 ± 0.01	0.445 ± 0.03	0.945 ± 0.02	0.594 ± 0.03	0.820 ± 0.02	0.081 ± 0.01	0.241 ± 0.02	0.279 ± 0.02	0.687 ± 0.03	0.655 ± 0.02	0.794 ± 0.03	0.044 ± 0.02	0.371 ± 0.02	0.343 ± 0.02	0.816 ± 0.03	0.625 ± 0.03	0.807 ± 0.01	0.063 ± 0.01
MANet[50]	0.537 ± 0.02	0.519 ± 0.02	0.969 ± 0.03	0.633 ± 0.03	0.880 ± 0.02	0.078 ± 0.01	0.292 ± 0.03	0.285 ± 0.02	0.743 ± 0.04	0.651 ± 0.02	0.797 ± 0.02	0.044 ± 0.01	0.415 ± 0.02	0.402 ± 0.02	0.849 ± 0.03	0.642 ± 0.02	0.839 ± 0.02	0.061 ± 0.01
Ours	0.587 ± 0.02	0.606 ± 0.03	0.963 ± 0.02	0.644 ± 0.01	0.877 ± 0.03	0.073 ± 0.01	0.323 ± 0.03	0.333 ± 0.02	0.759 ± 0.02	0.678 ± 0.02	0.855 ± 0.01	0.039 ± 0.01	0.455 ± 0.02	0.470 ± 0.03	0.861 ± 0.01	0.671 ± 0.02	0.866 ± 0.02	0.056 ± 0.01

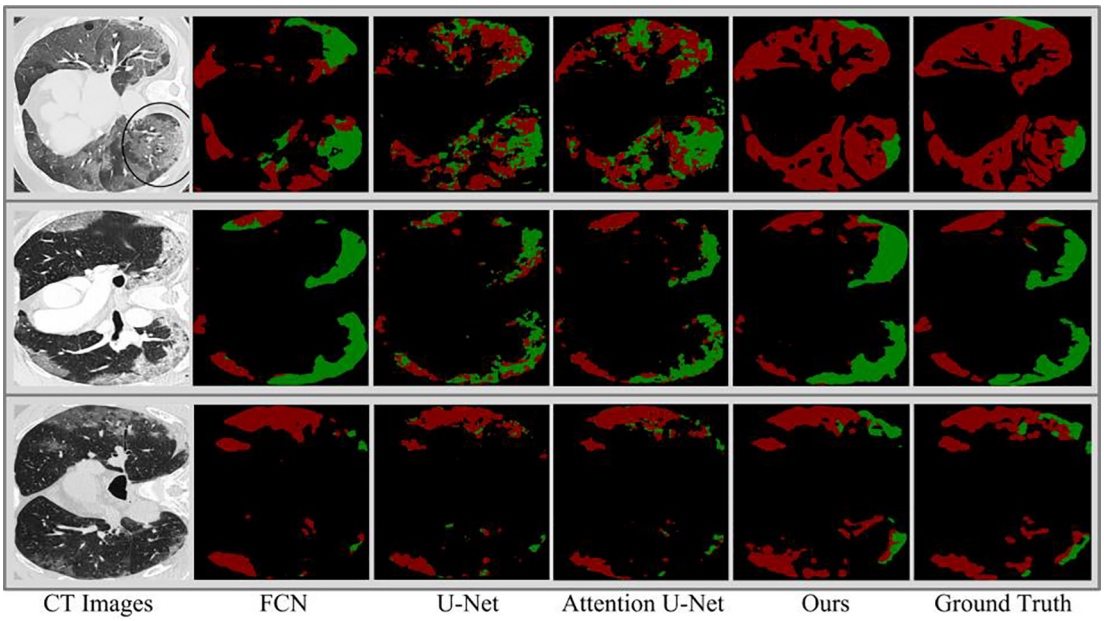

Fig. 3. Visual comparison of lung infection segmentation results, where the red and green regions indicate the GGO and consolidation, respectively.

proposed method is owed to the VQ-VAE module, proportion loss and semi-supervised learning. As can be observed, the proposed method is more sensitive to the small target, and the segmentation boundaries are more accurate and smoother.

4.4. Ablation study

In this section, the proposed VQ-VAE module, proportion loss, and semi-supervised adversarial learning are analyzed detailly. The VQ-VAE module is applied to regularize the shared network, which can be viewed as an encoder. It makes the encoder more robust and can generate features that are more representative. Furthermore, the high and low-level latent features are fused with the intermediate features of the backbone network to enhance the feature representation. Moreover, the proportion loss induces label marginal to match target class proportions, thereby alleviating class imbalance without losing generality. We also introduce semi-supervised adversarial learning to the framework, which makes the best use of unlabeled data. Meantime, the discriminator is applied to discover the trustworthy regions in the prediction of the unlabeled images, thereby providing additional supervisory signals. The semi-supervised adversarial learning is flexible enough to detect mismatches in a wide range of higher-order statistics between the predictions and the ground-truth without defining these manually, and the adversarial process encourages the segmentation network to generate predictions closer to the ground truth. The ablation study validates the proposed methods, and the results are presented in Table 2. It can be observed from Table 2 that all of the proposed methods improve the performance of the model. The VQ-VAE module can regularize the shared parameters of the network, and it can further provide supplementary features to enhance the feature representations. Moreover, the low and high-level features generated by this module can reconstruct the original image. The low-level features represent the global information of the images, such as the shape, color, and geometry of objects. On the contrary, the high-level latent code can provide local details, such as texture. As can be observed, VVM and L_{prop} have similar improvements in almost all metrics, but L_{prop} improves the results by 34.2% in terms of Dice on Consolidation areas. It reveals that the proportion loss encourages the model to focus on the small target. As presented in Table 2, L_{prop} brings considerable improvement in terms of S_α and E_ϕ^{mean}. Due to the S_α is presented to measure the structural similarity between prediction and ground-truth mask, and the E_ϕ^{mean} is a metric for evaluating local and global similarity. Thus, the gain demonstrates that the L_{prop} encourages the proportions of the predicted segmentation regions to match the ground-truth proportions, thereby increasing the coherence and structure of the segmentation. The *SSA* brings the greatest improvement, which shows that the model benefits from the supervision signal provided by unlabeled images. It leverages the self-taught process to refine the coarse pseudo labels of the unlabeled images, which are further applied to guide the training. Fig. 4.

4.5. Visualization results

To get a deeper understanding of the proposed method, the reconstructed images generated by VQ-VAE and the confidence maps generated by the discriminator are visualized in Fig. 4. As can be observed, the reconstructed images generated by the VQ-VAE module can restore the original images to a great extent, yet there are still some noises that failed to reconstruct the original

Table 2
Ablation study of the proposed methods. *VVM* stands for VQ-VAE module, *SSA* denotes semi-supervised adversarial learning. The best two results are shown in red and blue fonts.

	Ground-Glass Opacity						Consolidation					
Methods	Dice	Sen.	Spec.	S_α	E_ϕ^{mean}	MAE	Dice	Sen.	Spec.	S_α	E_ϕ^{mean}	MAE
Backbone	0.473 ± 0.05	0.575 ± 0.03	0.891 ± 0.02	0.591 ± 0.01	0.574 ± 0.03	0.140 ± 0.02	0.187 ± 0.03	0.261 ± 0.02	0.613 ± 0.02	0.573 ± 0.03	0.759 ± 0.02	0.049 ± 0.02
Backbone+VVM	0.520 ± 0.02	0.579 ± 0.02	0.912 ± 0.02	0.599 ± 0.01	0.659 ± 0.03	0.136 ± 0.02	0.213 ± 0.02	0.279 ± 0.02	0.628 ± 0.01	0.621 ± 0.01	0.775 ± 0.01	0.048 ± 0.01
Backbone+L_{prop}	0.514 ± 0.03	0.587 ± 0.03	0.935 ± 0.01	0.510 ± 0.01	0.731 ± 0.01	0.105 ± 0.01	0.251 ± 0.02	0.284 ± 0.03	0.662 ± 0.02	0.588 ± 0.00	0.802 ± 0.02	0.045 ± 0.01
Backbone+SSA	0.539 ± 0.04	0.599 ± 0.05	0.944 ± 0.02	0.625 ± 0.01	0.854 ± 0.02	0.128 ± 0.02	0.249 ± 0.02	0.278 ± 0.02	0.690 ± 0.01	0.597 ± 0.03	0.783 ± 0.01	0.043 ± 0.00
Backbone+*VVM*+L_{prop}	0.557 ± 0.03	0.590 ± 0.01	0.938 ± 0.04	0.618 ± 0.03	0.810 ± 0.02	0.095 ± 0.01	0.279 ± 0.01	0.301 ± 0.01	0.702 ± 0.01	0.630 ± 0.01	0.791 ± 0.02	0.043 ± 0.01
Backbone+SSA+L_{prop}	0.571 ± 0.04	0.609 ± 0.02	0.956 ± 0.02	0.641 ± 0.03	0.838 ± 0.01	0.091 ± 0.01	0.294 ± 0.02	0.315 ± 0.03	0.713 ± 0.01	0.628 ± 0.01	0.811 ± 0.03	0.041 ± 0.00
Backbone+VVM+SSA+L_{prop}	0.587 ± 0.02	0.606 ± 0.03	0.963 ± 0.02	0.644 ± 0.01	0.877 ± 0.03	0.073 ± 0.01	0.323 ± 0.03	0.333 ± 0.02	0.759 ± 0.02	0.678 ± 0.02	0.855 ± 0.01	0.039 ± 0.01

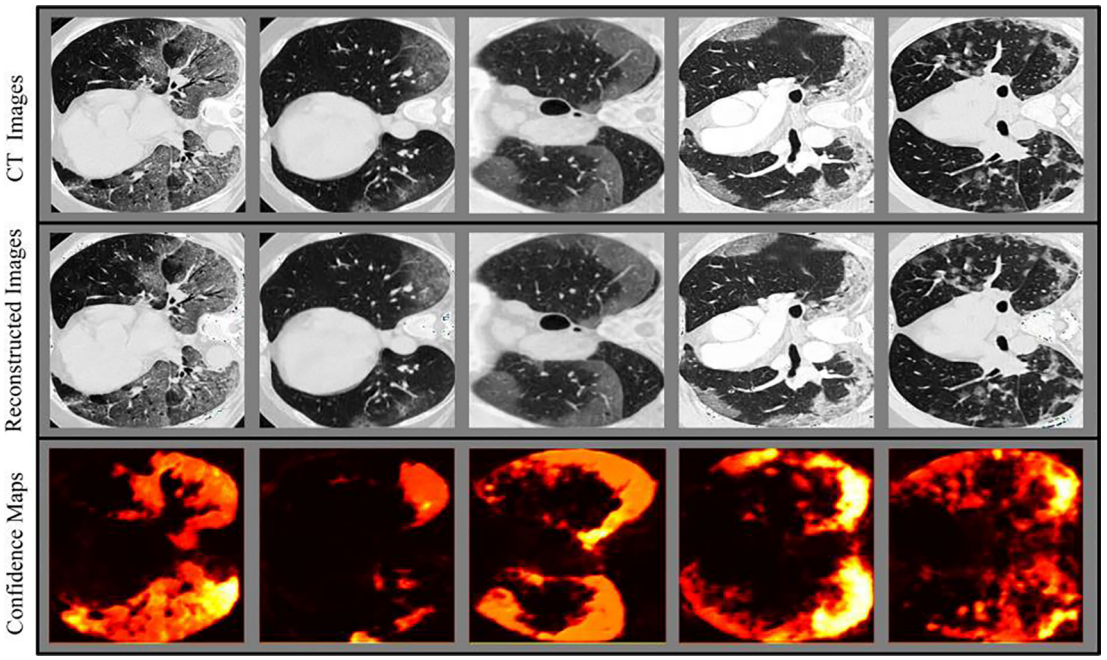

Fig. 4. Visualization results of the reconstructed images and confidence maps.

images' structure. It is worth mentioning that the high and low-level latent features of the module are applied to reconstruct the original images. Due to the features being another high-order representation of the images, they are fused with the intermediate features of the backbone network to improve the performance. The third line of Fig. 4 indicates the confidence maps generated by the discriminator. The brighter regions denote which part of the input data is closer to the ground truth distribution. The model discovers the trustworthy regions to guide the self-taught process and provide supplementary supervision signals, thereby promoting segmentation accuracy. According to the confidence maps, the untrusted regions are discarded in the coarse pseudo labels to obtain the refined pseudo labels. This self-taught process can further improve the performance by using the supervision signal provided by unlabeled data and regularizing the segmentation network.

5. Conclusions

In this work, a novel adversarial network has been proposed that consists of two sub-networks: segmentation network and discriminator network of which the segmentation network includes two parts: the backbone network, and the VQ-VAE module. The VQ-VAE module can regularize the shared parameters of the network. Meantime, the latent features generated by it are fused with the intermediate features of the backbone network to further enhance the performance. Moreover, the cross-entropy loss is analyzed which consists of two parts, ie., the conditional entropy part and label-marginal bias. It inspired us to propose a novel proportion loss to encourage the target class proportions to match the ground truth proportions. Furthermore, a semi-supervised scheme is presented that the unlabeled data are fed into the adversarial network to generate refined pseudo labels, which provide the supervised signals to improve the performance of the model. Extensive experiments on the COVID-SemiSeg dataset demonstrated that the proposed model outperforms the state-of-the-art model. The proposed model has great potential to be applied in assessing the diagnosis of COVID-19, e.g., quantifying the infected regions, monitoring the longitudinal disease changes, and mass screening processing. In the future, model accuracy can be further improved by reducing the computational complexity and enhancing robustness. In addition, unsupervised learning has recently gained a lot of attention from scholars, and our future work also considers extending semi-supervised learning to unsupervised learning, among which contrastive learning is particularly prominent. Contrastive learning can make full use of a large number of unannotated images to learn the prior knowledge distribution of the data. Through fine-tuning, the knowledge acquired during the pre-training process is transferred to downstream tasks and improves the performance of the model. Our future work will apply contrastive learning to existing tasks to further alleviate the reliance on labeled data.

CRediT authorship contribution statement

Ge Jin: Conceptualization, Methodology, Software, Investigation, Writing - original draft. **Chuancai Liu:** Supervision, Resources, Writing - review & editing, Funding acquisition. **Xu Chen:** Formal analysis, Writing - review & editing, Writing - review & editing.

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.

Acknowledgments

This work was supported by the National Natural Science Fund of China [Grant Nos. 61473155, 61872188, 62172225].

References

[1] M.B. Page J, Hinshaw D, In hunt for covid-19 origin, patient zero points to second wuhan market, in: The wall street journal, 2021.
[2] Covid-19 dashboard by the center for systems science and engineering (csse) at johns hopkins university (jhu),http://publichealthupdate.com/jhu/.
[3] A. Dixit, A. Mani, R. Bansal, Cov2-detect-net: Design of covid-19 prediction model based on hybrid de-pso with svm using chest x-ray images, Inform. Sci. 571 (2021) 676–692.
[4] M. Abdar, M.A. Fahami, S. Chakrabarti, A. Khosravi, P. Pławiak, U.R. Acharya, R. Tadeusiewicz, S. Nahavandi, Barf: A new direct and cross-based binary residual feature fusion with uncertainty-aware module for medical image classification, Inf. Sci. 577 (2021) 353–378.
[5] X. Qi, Z. Jiang, Y.U. Qian, C. Shao, S. Ju, Machine learning-based ct radiomics model for predicting hospital stay in patients with pneumonia associated with sars-cov-2 infection: A multicenter study.
[6] J. Chen, L. Wu, J. Zhang, L. Zhang, H. Yu, Deep learning-based model for detecting 2019 novel coronavirus pneumonia on high-resolution computed tomography, Sci. Rep. 10 (1).
[7] S. Chaganti, A. Balachandran, G. Chabin, S. Cohen, T. Flohr, B. Georgescu, P. Grenier, S. Grbic, S. Liu, F. Mellot, Quantification of tomographic patterns associated with covid-19 from chest ct, arXiv.
[8] 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.
[9] C. Zheng, X. Deng, Q. Fu, Q. Zhou, X. Wang, Deep learning-based detection for covid-19 from chest ct using weak label.
[10] A. Razavi, A. van den Oord, O. Vinyals, Generating Diverse High-Fidelity Images with VQ-VAE-2, arXiv e-prints (2019) arXiv:1906.00446 arXiv:1906.00446.
[11] O. Ronneberger, P. Fischer, T. Brox, U-net: Convolutional networks for biomedical image segmentation, Springer, Cham.
[12] Z. Zhou, M. Siddiquee, N. Tajbakhsh, J. Liang, Unet++: A nested u-net architecture for medical image segmentation, 4th Deep Learning in Medical Image Analysis (DLMIA) Workshop.
[13] Z. Iek, A. Abdulkadir, S.S. Lienkamp, T. Brox, O. Ronneberger, 3d u-net: Learning dense volumetric segmentation from sparse annotation, Springer, Cham.
[14] Z. Ullah, M. Usman, M. Jeon, J. Gwak, Cascade multiscale residual attention cnns with adaptive roi for automatic brain tumor segmentation, Inform. Sci.
[15] M. Rajchl, M. Lee, O. Oktay, K. Kamnitsas, J. Passerat-Palmbach, W. Bai, M. Rutherford, J. Hajnal, B. Kainz, D. Rueckert, Deepcut: Object segmentation from bounding box annotations using convolutional neural networks, IEEE Trans. Med. Imaging 36 (2) (2016) 674–683.
[16] W. Bai, O. Oktay, M. Sinclair, H. Suzuki, M. Rajchl, G. Tarroni, B. Glocker, A. King, P.M. Matthews, D. Rueckert, Semi-supervised learning for network-based cardiac mr image segmentation, Springer, Cham.
[17] J. Peng, G. Estradab, M. Pedersoli, C. Desrosiers, Deep co-training for semi-supervised image segmentation, Pattern Recognition.
[18] A. Myronenko, 3d mri brain tumor segmentation using autoencoder regularization, in: International MICCAI Brainlesion Workshop, 2018.
[19] Y.D. Ma, Q. Liu, Z.B. Qian, Automated image segmentation using improved pcnn model based on cross-entropy, in: Proceedings of 2004 International Symposium on Intelligent Multimedia, Video and Speech Processing, 2004, 2005..
[20] V. Pihur, S. Datta, S. Datta, Weighted rank aggregation of cluster validation measures: a monte carlo cross-entropy approach, Bioinformatics 23 (13) (2007) 1607.
[21] T.Y. Lin, P. Goyal, R. Girshick, K. He, P. Dollár, Focal loss for dense object detection, IEEE Trans. Pattern Anal. Mach. Intell. PP (99) (2017) 2999–3007.
[22] F. Milletari, N. Navab, S.A. Ahmadi, V-net: Fully convolutional neural networks for volumetric medical image segmentation, in: 2016 Fourth International Conference on 3D Vision (3DV), 2016.
[23] S. Salehi, D. Erdogmus, A. Gholipour, Tversky loss function for image segmentation using 3d fully convolutional deep networks, International Workshop on Machine Learning in Medical Imaging, in, 2017.
[24] F. Caliva, C. Iriondo, A.M. Martinez, S. Majumdar, V. Pedoia, Distance map loss penalty term for semantic segmentation.
[25] S. Zhao, B. Wu, W. Chu, Y. Hu, D. Cai, Correlation maximized structural similarity loss for semantic segmentation, arXiv preprint arXiv:1910.08711.
[26] B. Liu, J. Dolz, A. Galdran, R. Kobbi, I.B. Ayed, The hidden label-marginal biases of segmentation losses, arXiv preprint arXiv:2104.08717.
[27] M.K. Mahbub, M. Biswas, L. Gaur, F. Alenezi, K. Santosh, Deep features to detect pulmonary abnormalities in chest x-rays due to infectious diseasex: Covid-19, pneumonia, and tuberculosis, Inf. Sci. 592 (2022) 389–401.
[28] Y. Wang, X. Lu, Y. Zhang, X. Zhang, K. Wang, J. Liu, X. Li, R. Hu, X. Meng, S. Dou, et al, Precise pulmonary scanning and reducing medical radiation exposure by developing a clinically applicable intelligent ct system: Toward improving patient care, EBioMedicine 54 (2020) 102724.
[29] U. Imaging, United imaging sends out more than 100 ct scanners and x-ray machines to aid diagnosis of the coronavirus, Accessed: Apr 8 (2020) 2020.
[30] J. Chen, L. Wu, J. Zhang, L. Zhang, D. Gong, Y. Zhao, Q. Chen, S. Huang, M. Yang, X. Yang, et al, Deep learning-based model for detecting 2019 novel coronavirus pneumonia on high-resolution computed tomography, Sci. Rep. 10 (1) (2020) 1–11.
[31] A. Senior, J. Jumper, D. Hassabis, P. Kohli, Alphafold: Using ai for scientific discovery, DeepMind. Recuperado de:https://deepmind.com/blog/alphafold.
[32] Z. Hu, Q. Ge, S. Li, L. Jin, M. Xiong, Artificial intelligence forecasting of covid-19 in china, arXiv preprint arXiv:2002.07112.
[33] J. Hofmanninger, F. Prayer, J. Pan, S. Röhrich, H. Prosch, G. Langs, Automatic lung segmentation in routine imaging is primarily a data diversity problem, not a methodology problem, European Radiology Experimental 4 (1) (2020) 1–13.
[34] D.-P. Fan, T. Zhou, G.-P. Ji, Y. Zhou, G. Chen, H. Fu, J. Shen, L. Shao, Inf-net: Automatic covid-19 lung infection segmentation from ct images, IEEE Trans. Med. Imaging 39 (8) (2020) 2626–2637.
[35] G. Chassagnon, M. Vakalopoulou, E. Battistella, S. Christodoulidis, T.-N. Hoang-Thi, S. Dangeard, E. Deutsch, F. Andre, E. Guillo, N. Halm, et al., Ai-driven ct-based quantification, staging and short-term outcome prediction of covid-19 pneumonia, arXiv preprint arXiv:2004.12852.
[36] M. Vakalopoulou, G. Chassagnon, N. Bus, R. Marini, E.I. Zacharaki, M.-P. Revel, N. Paragios, Atlasnet: multi-atlas non-linear deep networks for medical image segmentation, in: International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer, 2018, pp. 658–666.
[37] 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.
[38] 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, arXiv preprint arXiv:1804.03999.
[39] X. Chen, L. Yao, Y. Zhang, Residual attention u-net for automated multi-class segmentation of covid-19 chest ct images, arXiv preprint arXiv:2004.05645.
[40] J. Kim, J.W. Fisher, A. Yezzi, M. Çetin, A.S. Willsky, A nonparametric statistical method for image segmentation using information theory and curve evolution, IEEE Trans. Image Process. 14 (10) (2005) 1486–1502.

[41] M. Tang, D. Marin, I.B. Ayed, Y. Boykov, Kernel cuts: Kernel and spectral clustering meet regularization, Int. J. Comput. Vision 127 (5) (2019) 477–511.
[42] Covid-19 ct segmentation dataset,https://medicalsegmentation.com/covid19/ (2020).
[43] J.P. Cohen, P. Morrison, L. Dao, K. Roth, T.Q. Duong, M. Ghassemi, Covid-19 image data collection: Prospective predictions are the future, arXiv preprint arXiv:2006.11988.
[44] D.-P. Fan, M.-M. Cheng, Y. Liu, T. Li, A. Borji, Structure-measure: A new way to evaluate foreground maps, in: Proceedings of the IEEE international conference on computer vision, 2017, pp. 4548–4557.
[45] D.-P. Fan, C. Gong, Y. Cao, B. Ren, M.-M. Cheng, A. Borji, Enhanced-alignment measure for binary foreground map evaluation, IJCAI (2018).
[46] J. Long, E. Shelhamer, T. Darrell, Fully convolutional networks for semantic segmentation, in: Proceedings of the IEEE conference on computer vision and pattern recognition, 2015, pp. 3431–3440.
[47] L.C. Chen, Y. Zhu, G. Papandreou, F. Schroff, H. Adam, Encoder-decoder with atrous separable convolution for semantic image segmentation, Springer, Cham.
[48] M.Z. Alom, M. Hasan, C. Yakopcic, T.M. Taha, V.K. Asari, Recurrent residual convolutional neural network based on u-net (r2u-net) for medical image segmentation.
[49] H. Zhao, J. Shi, X. Qi, X. Wang, J. Jia, Pyramid scene parsing network, in: Proceedings of the IEEE conference on computer vision and pattern recognition, 2017, pp. 2881–2890.
[50] T. Fan, G. Wang, Y. Li, H. Wang, Ma-net: A multi-scale attention network for liver and tumor segmentation, IEEE Access 8 (2020) 179656–179665.