PDAtt-Unet: Pyramid Dual-Decoder Attention Unet for Covid-19 infection segmentation from CT-scans

Fares Bougourzi, Cosimo Distante, Fadi Dornaika, Abdelmalik Taleb-Ahmed

Medical Image Analysis 86 (2023) 102797

Contents lists available at ScienceDirect

Medical Image Analysis

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

PDAtt-Unet: Pyramid Dual-Decoder Attention Unet for Covid-19 infection segmentation from CT-scans

Fares Bougourzi [a,b], Cosimo Distante [a], Fadi Dornaika [c,d,*], Abdelmalik Taleb-Ahmed [e]

[a] Institute of Applied Sciences and Intelligent Systems, National Research Council of Italy, 73100 Lecce, Italy
[b] University Paris-Est Cretéil, Laboratoire LISSI, 94400, Vitry sur Seine, Paris, France
[c] University of the Basque Country UPV/EHU, San Sebastian, Spain
[d] Ho Chi Minh City Open University, 97 Vo Van Tan, Ward Vo Thi Sau, District 3, Ho Chi Minh City, 70000, Viet Nam
[e] Université Polytechnique Hauts-de-France, Université de Lille, CNRS, Valenciennes, 59313, Hauts-de-France, France

ARTICLE INFO

Keywords:
Covid-19
Convolutional neural network
Deep learning
Segmentation
Unet

ABSTRACT

Since the emergence of the Covid-19 pandemic in late 2019, medical imaging has been widely used to analyze this disease. Indeed, CT-scans of the lungs can help diagnose, detect, and quantify Covid-19 infection. In this paper, we address the segmentation of Covid-19 infection from CT-scans. To improve the performance of the Att-Unet architecture and maximize the use of the Attention Gate, we propose the PAtt-Unet and DAtt-Unet architectures. PAtt-Unet aims to exploit the input pyramids to preserve the spatial awareness in all of the encoder layers. On the other hand, DAtt-Unet is designed to guide the segmentation of Covid-19 infection inside the lung lobes. We also propose to combine these two architectures into a single one, which we refer to as PDAtt-Unet. To overcome the blurry boundary pixels segmentation of Covid-19 infection, we propose a hybrid loss function. The proposed architectures were tested on four datasets with two evaluation scenarios (intra and cross datasets). Experimental results showed that both PAtt-Unet and DAtt-Unet improve the performance of Att-Unet in segmenting Covid-19 infections. Moreover, the combination architecture PDAtt-Unet led to further improvement. To Compare with other methods, three baseline segmentation architectures (Unet, Unet++, and Att-Unet) and three state-of-the-art architectures (InfNet, SCOATNet, and nCoVSegNet) were tested. The comparison showed the superiority of the proposed PDAtt-Unet trained with the proposed hybrid loss (PDEAtt-Unet) over all other methods. Moreover, PDEAtt-Unet is able to overcome various challenges in segmenting Covid-19 infections in four datasets and two evaluation scenarios.

1. Introduction

Since the outbreak of the Covid-19 pandemic (late 2019, Wuhan, China), the world has been in a global crisis. Detection of infected individuals plays a critical role in the fight against this disease. Medical imaging is often used as a primary or complementary tool to RT-PCR (which has a significant false-negative rate) to detect infected individuals (Kucirka et al., 2020). Medical imaging modalities used include radiographs (Vantaggiato et al., 2021) and the CT-scans (Bougourzi et al., 2021a). Indeed, CT-scans are strongly recommended for both initial and follow-up examinations when Covid-19 is suspected (Jin et al., 2020; Wu et al., 2021; Bougourzi et al., 2023). In addition, CT scans have demonstrated their efficiency for Covid-19 diagnosis at early stages, whereas CT findings may be present before the onset of symptoms (Kim et al., 2020; Pan et al., 2020b). On the other hand, CT-scans can be used for more purposes. CT scans have been used to segment the infected parts and can be used directly to estimate the severity of infection and percentage (Bougourzi et al., 2021c).

Indeed, there are different clinical scenarios for patients with Covid-19 infection, including early versus advanced stages, asymptomatic versus symptomatic patients, and severe versus non-severe situations (Sun et al., 2020). Therefore, the CT imaging findings associated with Covid-19 correlate strongly with these clinical scenarios. Typically, the time course of Covid-19 infection consists of three main stages: early, mid, and late (Salehi et al., 2020). In addition, the infection findings in CT-scans vary from case to case. Initial infection presents as bilateral multi-lobar Ground-Glass Opacification (GGO), especially in the lower lobes, with peripheral or posterior distribution (Salehi et al., 2020; Sun et al., 2020). In the intermediate stage, the number and size of GGOs increase, and there is progressive transformation of GGO into multi-focal consolidating opacities, thickening of the septum, and development of

* Corresponding author at: University of the Basque Country UPV/EHU, San Sebastian, Spain.
E-mail addresses: fares.bougourzi@isasi.cnr.it (F. Bougourzi), cosimo.distante@cnr.it (C. Distante), fadi.dornaika@ehu.eus (F. Dornaika), Abdelmalik.Taleb-Ahmed@uphf.fr (A. Taleb-Ahmed).

https://doi.org/10.1016/j.media.2023.102797
Received 16 May 2022; Received in revised form 10 January 2023; Accepted 8 March 2023
Available online 21 March 2023

a crazy-patch pattern, especially in the severe cases. The study by Pan et al. (2020a) showed that imaging findings had improved in 75% of patients after day 14, including a reduced number of affected lobes and resolution of the crazy paving pattern and consolidating opacities.

Although CT-scanners are available in most hospitals, even in less developed countries, an experienced radiologist is needed to detect and follow the progression of Covid-19 infection. To address this problem, many machine learning approaches have been proposed to provide automated solutions that reduce the need for radiologists and help them (Vantaggiato et al., 2021; Muhammad et al., 2022; Bougourzi et al., 2021b; Santa Cruz et al., 2021).

Because Covid-19 infection findings exhibit high variability in shape, size, type, and position during disease progression, the segmentation task is very challenging. On the other hand, the segmentation task with artificial intelligence comes with additional challenges (Kumar Singh et al., 2021; Laradji et al., 2021). One of the most difficult challenges is the low intensity contrast between the infection and normal tissue, especially for the ground glass opacity (GGO) feature. In addition, it is difficult to distinguish between consolidation and crazy paving and non-lung tissue, especially when the infection has a peripheral or posterior distribution with the presence of consolidation or crazy paving (Kumar Singh et al., 2021; Laradji et al., 2021).

In the last decade, Deep Learning approaches, especially CNNs, have achieved high performance on most computer vision tasks compared to traditional methods (Bougourzi et al., 2020, 2022). However, to train an efficient Deep Learning architecture, huge labeled data is needed (Bougourzi et al., 2020, 2022). When Covid-19 began to proliferate, such adequate data were not available due to many factors. These factors include: the privacy restrictions on the use of patient data, the large amount of time and effort required in the labeling process, and the need for many experienced radiologists to segment the infection regions in the CT-scans (Zhao et al., 2021b; Fung et al., 2021; Laradji et al., 2021). In this pandemic, it was difficult to find radiologists who could perform this task.

Due to data limitation, most existing works have used small data to train their models (Fan et al., 2020; Abdel-Basset et al., 2021; Liu et al., 2021; Elharrouss et al., 2021). On the other hand, these works have only used infected slices to train and test their models (Fan et al., 2020; Wang et al., 2020; Abdel-Basset et al., 2021; Liu et al., 2021; Kumar Singh et al., 2021; Elharrouss et al., 2021; Wang et al., 2022). However, this does not correspond to the real scenario where the infections in CT-scan usually occur only in a few slices. Therefore, it is interesting to investigate the performance of the segmentation approaches for both infected and uninfected slices corresponding to a realistic configuration.

In the primary datasets used for Covid-19 segmentation, only the middle slices of the CT scans were used. However, infection can occur in any area of the lung. In recent months, additional datasets for Covid-19 segmentation have been made publicly available (Radiologists, 2019; Ma et al., 2021; Zhang et al., 2020). The aim of this work is to segment Covid-19 infections in a real scenario (all slices of 3D CTs are used for training and testing). To this end, inspired by the Att-Unet architecture, we propose three different architectures. Moreover, we compare our results with three CNN-based segmentation architectures and three published architectures for Covid-19 segmentation whose implementation codes have been made publicly available.

Training segmentation models with the Cross-Entropy (CE) loss function usually have problems in predicting object boundaries, resulting in fuzzy boundaries (Qin et al., 2019). In addition, a Covid-19 infection may have multiple small infection regions. To address these issues, we proposed an additional BCE loss function that focuses on the infection boundaries. The small infection regions have more boundary pixels (relative to the area of the infection) than the larger ones. Therefore, the proposed hybrid loss function weighs the small infection regions more than the large infections.

For pulmonary and Covid-19 infections, there may be high variability in screening and changes in intensity and illumination from one CT-scanner to another and from one imaging environment to another. Patient age, ethnicity, and severity of Covid-19 infection may also affect the appearance of the CT scan. Therefore, it is interesting to investigate the effectiveness of segmentation methods in cross-datasets experiments. The cross-datasets experiments investigate the generalization ability of the proposed machine learning solutions and can provide a clear idea of their applicability in the real world.

In order to provide an efficient Covid-19 infection segmentation solution that can overcome the above challenges from different aspects, our main contributions can be summarized as follows:

- Inspired by the Att-Unet architecture, we propose three different architectures for segmenting Covid-19 infections from CT-scans. The first variant, Pyramid Att-Unet (PAtt-Unet), uses image pyramids to preserve the spatial awareness in all of the encoder layers. Unlike most attention-based segmentation architectures, our proposed PAtt-Unet uses the attention gates not only in the decoder but also in the encoder.
- In addition to the PAtt-Unet architecture, we propose DAtt-Unet, designed to segment Covid-19 infections and lungs simultaneously. In DAtt-Unet, the two segmentation tasks share the encoder and the intermediate blocks of Att-Unet but each has its own decoder.
- Based on PAtt-Unet and DAtt-Unet, we propose a Pyramid Dual-Decoder Att-Unet (PDAtt-Unet) architecture using the pyramid and attention gates to preserve the global spatial awareness in all of the encoder layers. In the decoding phase, PDAtt-Unet has two independent decoders that use the Attention Gates to segment infection and lung simultaneously.
- To address the shortcomings of the binary cross entropy loss function in distinguishing the infection boundaries and the small infection regions, we propose the BCE_{Edge} loss that focuses on the edges of the infection regions.
- To evaluate the performance of our proposed architectures, we use four public datasets with two evaluation scenarios (intra and cross datasets), all slices from CT scans are used for the training and testing phases.
- To compare the performance of our approach with other CNN-based segmentation architectures, we use three baseline architectures (Unet Ronneberger et al., 2015, Att-Unet Oktay et al., 2018 and Unet++ Zhou et al., 2018) and three state-of-the-art architectures for Covid-19 segmentation (InfNet Fan et al., 2020, SCOATNet Zhao et al., 2021a, and nCoVSegNet Liu et al., 2021). The experimental results show the superiority of our proposed architecture compared to the basic segmentation architectures as well as to the three state-of-the-art architectures in both intra-database and inter-database evaluation scenarios. The codes of the proposed architectures and the loss function have been made publicly available at. https://github.com/faresbougourzi/PDAtt-Unet. (Last accessed April, 20^{th} 2022).

The remainder of this paper is organized as follows: Section 2 presents some related work on segmentation of Covid-19 infections from CT-scans. In Section 3, we describe our proposed approach. Section 4 consists of the description of the datasets used and the evaluation scenarios and metrics. Section 5 presents and discusses the experiments and results. Section 6 shows some segmentation examples. Finally, Section 7 concludes the paper.

2. Related work

Over the past decade, Deep Learning has enabled us to master many tasks that seemed impossible for machines (Bougourzi et al., 2020, 2022). In particular, medical imaging tasks that are very complicated

and require specialists and radiologists (Prevedello et al., 2019; Roth et al., 2022). Machine learning can provide an efficient solution to medical imaging tasks to assist and guide radiologists and physicians to reduce subjectivity and bias in decision-making. It could also significantly reduce the time and effort required (Wang et al., 2020; Zhao et al., 2021b; Fung et al., 2021; Laradji et al., 2021).

Since the emergence of the Covid-19 pandemic, many approaches to segmenting Covid-19 infection have been proposed In Zhao et al. (2021b), Xiangyu Zhao et al. proposed dilated dual attention U-Net (D2 A U-Net) to automatically segment lung infection in Covid-19 CT slices. To evaluate the performance of their approach, they took only the infected slices from Segmentation dataset nr.2 (Radiologists, 2019) and the COVID-19-CT-Seg dataset (Ma et al., 2021) and used them as training data. As test data, they used the 100 slices of the dataset presented in Radiologists (2019).

Liu et al. (2021) proposed a two-stage cross-domain transfer learning system. Their framework consists of two main components. First, they proposed nCoVSegNet, a deep learning-based approach that uses attention-driven feature fusion and large receptive fields. Second, they trained nCoVSegNet with a cross-domain transfer learning strategy that leverages knowledge from natural images (i.e., ImageNet) and medical images (i.e., LIDC-IDRI) to reinforce final training on CT images with Covid-19 infections. They used MosMedData dataset (Morozov et al., 2020) for the final stage of transfer learning (40 slices for training and 10 slices for testing), then they tested the obtained model on the COVID-19-CT-Seg dataset (Ma et al., 2021).

In Wang et al. (2022), Ruxin Wang et al. proposed an encoder–decoder CNN-based system for segmenting Covid-19 infections from the CT scans. Their approach is based on aggregation of peer and cross-level contextual learning. To capture the complex structure, they used an autofocus module to detect and incorporate multiscale contextual information at the peer level. They also proposed a panoramic module to capture complementary fine details and semantic information. To evaluate the performance of their approach, Ruxin Wang et al. combined different datasets, MosMedData dataset (Morozov et al., 2020; Ma et al., 2021), COVID-19 CT segmentation (Radiologists, 2019), and Segmentation dataset nr. 2 (Radiologists, 2019), and removed the slices that have no infection. Then, the obtained dataset was randomly divided into a training dataset and a test dataset (it is not clear whether the patient-independent protocol was followed or not).

Deng-Ping Fan et al. proposed a deep network (Inf-Net) for segmentation of Covid-19 lung infections (Fan et al., 2020). Their Inf-Net approach uses parallel partial decoders to aggregate high-level features and create a global map. They also used implicit reverse attention and explicit edge attention to model the boundaries and improve the representations.

Since Covid-19 infection labeling is a tedious and time-consuming task, the semi-supervised method has been extensively studied (Yang et al., 2021; Abdel-Basset et al., 2021; Wang et al., 2020; Mu et al., 2021). Fan et al. (2020) investigated a semi-supervised segmentation strategy that exploits the framework of randomly selected reproduction. The experimental results showed that using the semi-supervised framework can improve learning ability and achieve better performance.

At the advent of the Covid-19 pandemic, the available labeled data for training and evaluating machine learning approaches was very limited, especially for the segmentation task. For this reason, the first state-of-the-art approaches were evaluated with very limited data, which limits the study population (Fan et al., 2020; Abdel-Basset et al., 2021; Liu et al., 2021; Elharrouss et al., 2021). This reduces the reliability of these approaches in real-world scenarios, where Covid-19 infection exhibits high case-to-case variability in shape, type, location, and intensity (Sun et al., 2020; Kumar Singh et al., 2021; Laradji et al., 2021). On the other hand, many state-of-the-art approaches were trained and evaluated using only the infected slices (Fan et al., 2020; Wang et al., 2020; Abdel-Basset et al., 2021; Liu et al., 2021; Kumar Singh et al., 2021; Elharrouss et al., 2021; Wang et al., 2022). However, this does not correspond to the real scenario, where the infections in CT-scan usually occur only in a few slices with different degrees. In addition to the above two evaluation limitations, most of the state-of-the-art approaches were evaluated only for the intra-dataset scenario. However, the cross-dataset scenario investigates the generalization capability for real-world application scenarios. This is a very important evaluation scenario, especially in the case of CT-scanning data where the intensity and illumination vary greatly from one CT-scanner to another and from one acquisition setting to another. The goal of this work is to develop an efficient CNN-based approach for segmenting Covid-19 infections from CT-scans in real clinical situations, using all slices of 3D CT-scans for training and testing. In addition, extensive evaluations are performed on four datasets for intra-datasets and cross-datasets to investigate the generalization ability of the proposed approach in different real-world scenarios.

3. The proposed approach

In this section, we first provide details of our proposed Pyramid Att-Unet (PAtt-Unet), Dual-Decoder Att-Unet (DAtt-Unet) and Pyramid Dual-Decoder Att-Unet (PDAtt-Unet) architectures in terms of network architecture and core network components. Then, we present the proposed Hybrid loss function.

3.1. Pyramid Att-Unet (PAtt-Unet)

During the encoding phase of Unet-like segmentation architectures, high-level features are learned through successive convolutional layers and pooling layers. This shrinks the spatial dimensionality and expands the depth of the feature maps. However, this learning process leads to a loss of spatial information, which is very important for identifying the salient parts of the infection. To preserve global contextual awareness in all of the encoder layers, we propose a Pyramid Att-Unet architecture (PAtt-Unet). The encoder of the proposed PAtt-Unet uses four pyramid levels (P_1, P_2, P_3, and P_4) of the input slice image (I), as shown in Fig. 1b. Specifically, the PAtt-Unet architecture encoder is constructed using Basic ConvBlocks, Attention Gates (AGs), and the four pyramid paths. First, the Basic ConvBlock consists of two 3×3 convolution kernels, each followed by a batch normalization and a ReLU activation function, as shown in Fig. 2(a). Second, the Attention Gate (AG) is identical to the one proposed in the Att-Unet architecture (Oktay et al., 2018). Unlike the Att-Unet architecture, the AGs are not only used in the decoding phase, but are also used to preserve spatial awareness during the encoding phase. Fig. 2(b) shows the details of the AG block. As can be seen in this figure, the AG receives an input feature x ($[h, w, c_x]$) and a gating signal g ($[h, w, c_g]$), where h and w represent the height and width and c_x and c_g represent the number of channels of the input feature x and gating signal g, respectively. Each of the input features x and the gating signal g is fed into a 1-by-1 convolutional layer to unify the number of channels c_x and c_g to c_{int}. The obtained G and X features are summed to construct GX features, which are fed to a ReLU activation function and then to a 1-by-1 convolutional layer that shrinks the number of channels to 1, followed by a Sigmoid activation function. This produces a spatial attention filter A, which is of the shape $[h, w]$. Finally, the input feature x is element-wise multiplied by the attention filter A, as shown in the following equation:

$$x_{att} = A \otimes x \tag{1}$$

As shown in Fig. 1.(b), the encoder of the proposed PAtt-Unet uses four pyramid levels (P_1, P_2, P_3, and P_4) of the input slice image (I). Each pyramid image is fed into Basic ConvBlocks to match the feature maps of the corresponding encoder layers' output. Each output of the pyramid convolutional feature maps is used as a gating signal to control the attention of the corresponding encoder layer. On the other hand, the decoder of the PAtt-Unet architecture is identical to the decoder of the Att-Unet architecture (Oktay et al., 2018). The detailed structure of the proposed PAtt-Unet architecture can be found in Fig. A.5 in Appendix.

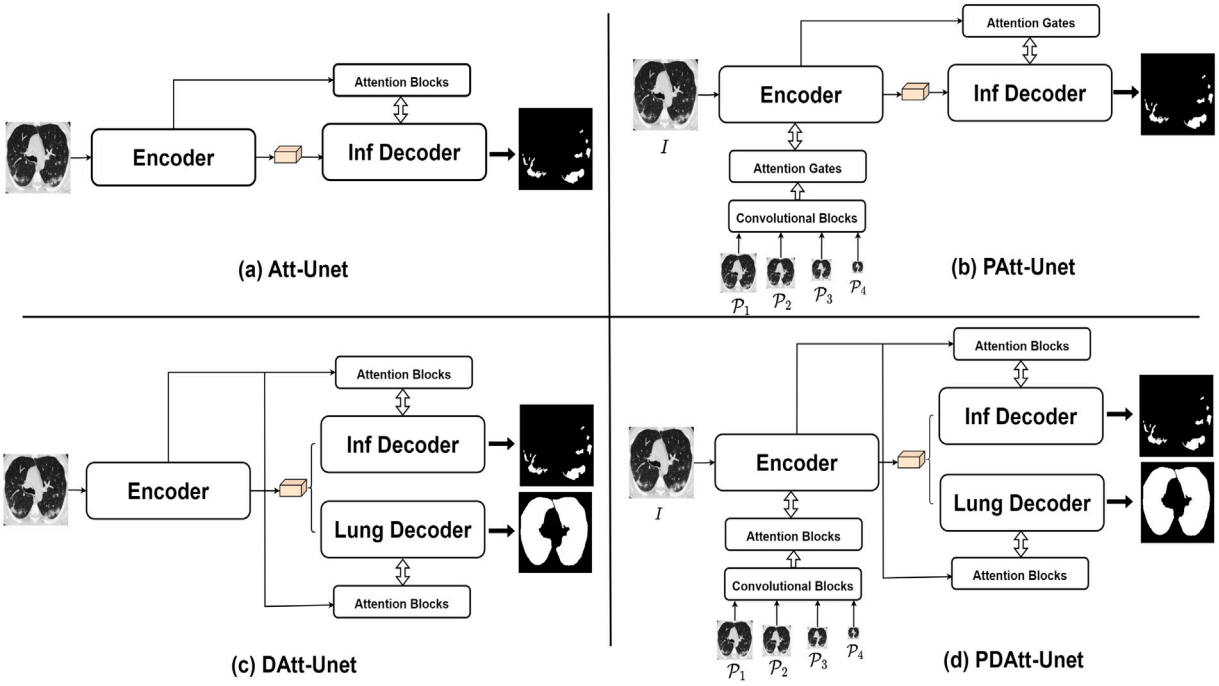

Fig. 1. Comparison of Att-Unet (a), our proposed PAtt-Unet (b), our proposed DAtt-Unet (c), and PDAtt-Unet (d).

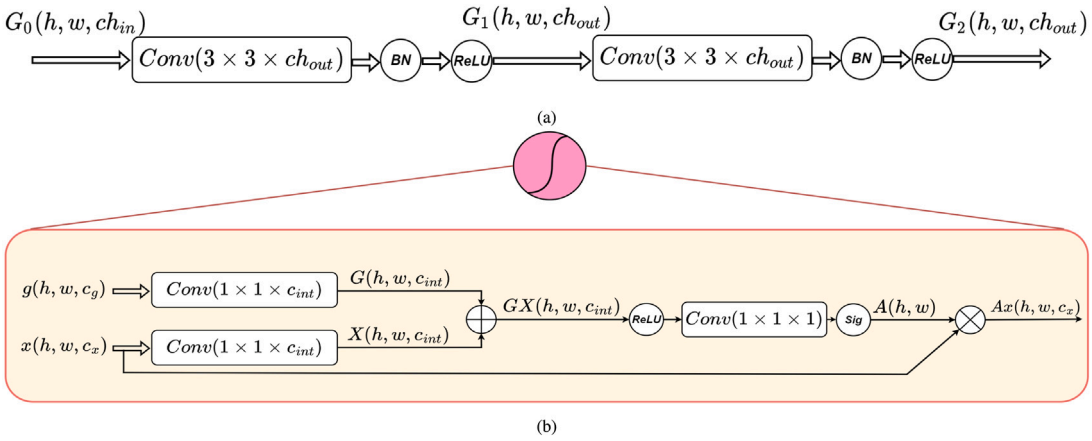

Fig. 2. Schematic description of (a) Encoder and Decoder Basic ConvBlock. (b) Attention Gate block, where g is the gating signal and the x is the input feature maps. $A(h, w)$ is the obtained spatial attention, which is applied for all channels of the input feature maps (x).

3.2. Dual-decoder Att-Unet (DAtt-Unet)

To segment Covid-19 infection and exclude confusion with the other lung tissues, we propose a DAtt-Unet architecture. Our DAtt-Unet is designed to segment both infections and lung regions, simultaneously. The goal is to guide the training process to look inside the lung regions and distinguish between the infection tissue (especially when consolidation and crazy paving are peripherally or posteriorly distributed) and non-lung tissue. Figs. 1(a) and (c) show the difference between our proposed DAtt-Unet and Att-Unet architectures. Both DAtt-Unet and Att-Unet use the attention gates (AG). As shown in Fig. 2(b), AG learns the important spatial regions from the skip features and reduces redundant features by using the up-sampling features as the gating signal. As shown in Fig. 1(c), DAtt-Unet has the same encoder for the infection and lung segmentation tasks that is used by the skip connections, while each task has its own decoder. The goal is to use the encoder to learn high-level features for both tasks. Then, the decoders use these features to segment the infection and lung regions independently. The detailed structure of the proposed PAtt-Unet architecture can be found in Fig. A.6 in Appendix.

3.3. Pyramid Dual-Decoder Att-Unet (PDAtt-Unet)

Our proposed Pyramid Dual-Decoder Att-Unet (PDAtt-Unet) combines the two architectures PAtt-Unet and DAtt-Unet, as shown in Fig. 1, (b), (c), and (d). Fig. 3 illustrates the details of the proposed PDAtt-Unet architecture, which consists of one encoder with pyramid paths and two decoders.

The PDAtt-Unet encoder consists of two main components: the main encoder path, which consists of five Basic ConvBlocks layers and four maxpooling layers. The outputs of the five Encoder layers are denoted by E_1, E_2, E_3, and E_4, respectively. The second component of the encoder is the pyramid paths. In total there are four pyramid paths, which has pyramid inputs (P_1, P_2, P_3, and P_4) of the input slice image (I), respectively. The pyramid paths generate the pyramid feature maps, denoted by P_{f_1}, P_{f_2}, P_{f_3} and P_{f_4}, by using Basic ConvBlocks. The pyramid feature maps are used to preserve the spatial attention

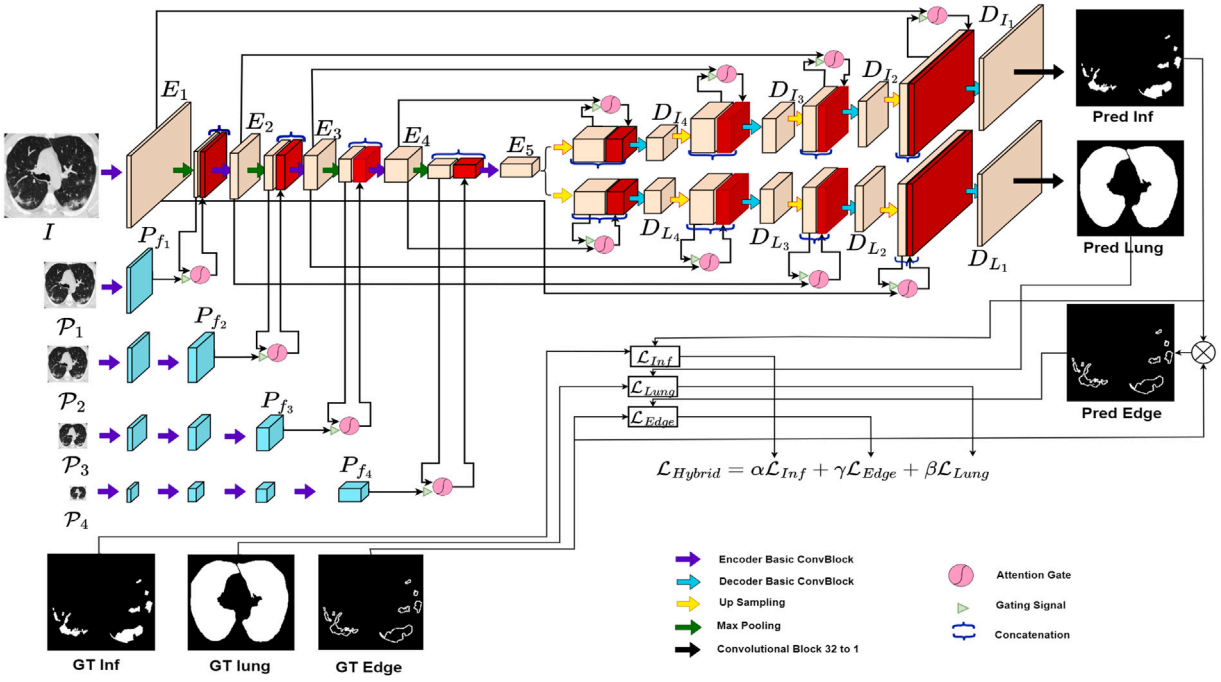

Fig. 3. Our proposed Pyramid PDEAtt-Unet. Both PDAtt-Unet and PDEAtt-Unet have identical architecture. However, PDEAtt-Unet exploits loss function denoted as Edge loss.

awareness in all encoder layers, where they are used as gating signals of the Encoder AGs. More specifically, E_1 is obtained using a Basic ConvBlock that uses 3 input channels (input slice images) and generates 32 feature maps. The obtained E_1 is passed to the maxpooling layer, which reduces the spatial dimensions to half. On the other hand, the first input to the pyramid path is the pyramid image P_1 (which is half the size of the input image I). P_1 is fed into the Basic ConvBlock, which is similar to the first Encoder Basic ConvBlock (3–32), and the output is denoted P_{f_1}. P_{f_1} and the maxpooling of E_1 are the input of the first encoder attention gate, where P_{f_1} is the gating signal. The output of the AG block is concatenated with the maxpooled E_1. The concatenated feature maps are fed into the second Basic ConvBlock of the main encoder path, which learns higher feature representations from the concatenated features ([112, 112, 64]). The features of the second encoder layer are denoted by E_2 [112, 112, 64]. Similarly, the input of the second pyramid path is P_2, which is fed into two Basic ConvBlocks (3–32 and 32–64), the output is denoted by P_{f_2}. P_{f_2} and the maxpooled E_2 [56, 56, 64] are the inputs of the second encoder attention gate, where P_{f_2} is the gating signal. The output of the second AG block is concatenated to the maxpooled E_2. The concatenated feature maps are fed into the third Basic ConvBlock of the main encoder path, which learns higher feature representations from the concatenated features ([56, 56, 128]). The features of the third encoder layer are denoted by E_3 [56, 56, 128]. Similarly, E_4 [28, 28, 256] and E_5 [14, 14, 512] are obtained by using pyramid paths (P_{f_3} and P_{f_4}), attention gates (third and fourth), and Basic ConvBlocks, as shown in Fig. 3.

The decoder of PDAtt-Unet has two identical paths, the first path is for segmenting of Covid-19 infection and the second path is for segmenting of lungs. Each path consists of four decoder Basic ConvBlocks, four up-sampling blocks (bi-linear up-sampling) and four attention gates. More specifically, the last encoder feature map (E_5) is up-sampled to twice the spatial dimension to obtain feature maps of [28, 28, 512]. The up-sampled feature maps and the feature maps from the fourth skip encoder E_4 are fed into the attention gate, with the up-sampled feature maps representing the gating signal. The output of the attention gate is concatenated with the up-sampled feature maps of E_5, then the decoder Basic ConvBlock is applied to obtain D_{I_4} [28, 28, 256] for the infection segmentation path. Similar is applied to obtain D_{L_4} for the lung segmentation path. Similarly, D_{I_3}, D_{I_2} and D_{I_1} and D_{L_3}, D_{L_2} and D_{L_1} are obtained for the infection and lung segmentation paths, respectively. For this purpose, the encoder feature maps (E_3, E_2, and E_1), attention gates, Decoder Basic ConvBlocks, and up-sampling blocks are used, as shown in Fig. 3. The last infection path feature map D_{I_1} [224, 224, 32] is mapped to a channel output [224, 224, 1] using the (1 × 1 × 1) convolutional block. Similarly, the last lung path feature map D_{L_1} [224, 224, 32] is mapped to a channel output [224, 224, 1] using the (1 × 1 × 1) convolution block. Details of the dimensions of the PDAtt-Unet layers can be found in Table 1.

Table 1
Our proposed PDAtt-Unet architecture in detail. E2conv is an encoder two convolutional block. D2conv is a decoder two convolutional block.

Block	Layer	Input	Output
Encoder	$E2conv_1$	224 × 224 × 3	224 × 224 × 32 = E_1
	$MaxPooling_1$	224 × 224 × 32	112 × 112 × 32
	$E2conv_2$	112 × 112 × 64	112 × 112 × 64 = E_2
	$MaxPooling_2$	112 × 112 × 64	56 × 56 × 64
	$E2conv_3$	56 × 56 × 128	56 × 56 × 128 = E_3
	$MaxPooling_3$	56 × 56 × 128	28 × 28 × 128
	$E2conv_4$	28 × 28 × 256	28 × 28 × 256 = E_4
	$MaxPooling_4$	28 × 28 × 256	14 × 14 × 256
	$E2conv_5$	14 × 14 × 512	14 × 14 × 512 = E_5
Decoder	$Up - Sampling_1$	14 × 14 × 512	28 × 28 × 512
	$D2conv_1$	28 × 28 × 768	28 × 28 × 256 = $D_{I_4} = D_{L_4}$
	$Up - Sampling_2$	28 × 28 × 256	56 × 56 × 256
	$D2conv_2$	56 × 56 × 384	56 × 56 × 128 = $D_{I_3} = D_{L_3}$
	$Up - Sampling_3$	56 × 56 × 128	112 × 112 × 128
	$D2conv_3$	112 × 112 × 192	112 × 112 × 64 = $D_{I_2} = D_{L_2}$
	$Up - Sampling_4$	112 × 112 × 64	224 × 224 × 64
	$D2conv_4$	224 × 224 × 96	224 × 224 × 32 = $D_{I_1} = D_{L_1}$

3.4. Hybrid loss function

In our experiments, we consider the standard BCE loss function, where we use it to train Unet (Ronneberger et al., 2015), Att-Unet (Oktay et al., 2018), Unet++ (Zhou et al., 2018), InfNet (Fan et al., 2020), SCOATNet (Zhao et al., 2021a), nCoVSegNet (Liu et al., 2021) and the proposed PAtt-Unet.

For the proposed DAtt-Unet and PDAtt-Unet architectures, which have two segmentation tasks simultaneously, the loss function (L_{Global}) is composed as follows:

$$\mathcal{L}_{Global} = \alpha\,\mathcal{L}_{Inf} + \beta\,\mathcal{L}_{Lung} \tag{2}$$

where $\mathcal{L}_{Inf}$ and $\mathcal{L}_{Lung}$ are the infection and lung task losses, respectively. The weights α and β are set to 0.7 and 0.3, respectively. The goal is to pay more attention to the infection segmentation task than to the lung segmentation one, since it is the main segmentation task.

Since training segmentation models with the loss function CE usually suffers from predicting the boundaries of the objects, resulting in fuzzy boundaries (Qin et al., 2019). In addition, a Covid-19 infection may have multiple small infection regions, and the segmentation models usually cannot segment these small infections. To address these issues, we propose an additional BCE loss function that focuses on the infection boundaries by giving larger weights to the boundaries, with the small infections having larger weights than the large ones. The proposed hybrid loss function is defined as follows:

$$\mathcal{L}_{Hybrid} = \alpha\,\mathcal{L}_{Inf} + \beta\,\mathcal{L}_{Lung} + \gamma\,\mathcal{L}_{Edge} \tag{3}$$

where $\mathcal{L}_{Inf}$ and $\mathcal{L}_{Lung}$ are the losses associated with the infection and lung segmentation tasks, respectively. The weights α and β are set to 0.7 and 0.3, respectively. $\mathcal{L}_{Edge}$ is the edge loss weighted by γ. The $\mathcal{L}_{Inf}$ and $\mathcal{L}_{Lung}$ are defined by:

$$\mathcal{L}_{Inf} = -\sum_{m=1}^{B}\sum_{i=1}^{W\cdot H} G_{R_i}\,log(p_i) + (1 - G_{R_i})\,log(1 - p_i) \tag{4}$$

$$\mathcal{L}_{lung} = -\sum_{m=1}^{B}\sum_{i=1}^{W\cdot H} G_{L_i}\,log(s_i) + (1 - G_{L_i})\,log(1 - s_i) \tag{5}$$

where B, W and H are the used batch size in training and the width and height of the predicted segmentation mask, respectively. $G_{R_i} \in \{0,1\}$ and $G_{L_i} \in \{0,1\}$ are the ground truth label of the infection and lung of pixel i, respectively. Also, p_i and s_i are the prediction probabilities of the infection and the lung of pixel i obtained from the decoders for segmenting the infection and lung, respectively. On the other hand, for batch size (B) and prediction width (W) and height (H), the edge loss function is defined as follows:

$$\mathcal{L}_{Edge} = -\sum_{m=1}^{B}\sum_{i=1}^{W\cdot H} G_{E_i}\,log(p_i) + (1 - G_{E_i})\,log(1 - p_i) \tag{6}$$

where $G_{R_i} \in \{0,1\}$ and p_i are the ground truth label of the edge infection and the edge prediction probabilities of pixel i, respectively. The ground-truth edge pixels are obtained by applying the morphological gradient to the ground-truth infection regions. After obtaining the ground-truth edge pixels, the BCE loss function is applied between the ground-truth pixels and the corresponding pixels in the mask map. The weight of the edge loss γ is experimentally determined for each dataset. For simplicity, the training of the proposed PDAtt-Unet using the hybrid loss function is referred to as PDEAtt-Unet.

In the proposed edge loss, the pixels at the infection boundary, which cause more confusion in the prediction than the other pixels, are weighted more heavily. Unlike other targets where the edges in the original images have significant structure, the edges of the Covid-19 infection have extremely few pixels, so adding a head or a branch to predict the edges is useless for solving multiple tasks, as found experimentally.

4. Datasets and evaluation metrics

4.1. Datasets

To evaluate the performance of our proposed architectures, we used four publicly available datasets, which are COVID-19 CT segmentation (Radiologists, 2019), Segmentation dataset nr.2 (Radiologists, 2019), COVID-19-CT-Seg dataset (Ma et al., 2021), and CC-CCII Segmentation Dataset (Zhang et al., 2020), which are summarized in Table 2. The COVID-19 CT segmentation dataset (Radiologists, 2019) consists of 100 axial CT images (slices) from more than 40 patients with Covid-19 infection. Segmentation dataset nr.2 (Radiologists, 2019) consists of 9 3D CT scans. In total, it consists of 829 slices, of which 373 slices have evidence of Covid-19 infection assigned by a radiologist.

Table 2
The used datasets summary.

Name	Dataset	#CT-Scans	#Slices
Dataset_1	COVID-19 CT segmentation (Radiologists, 2019)	40	100
Dataset_2	Segmentation dataset nr. 2 (Radiologists, 2019)	9	829
Dataset_3	COVID-19-CT-Seg dataset (Ma et al., 2021)	20	3520
Dataset_4	CC-CCII Segmentation Dataset (Zhang et al., 2020)	150	750

The COVID-19-CT-Seg dataset (Ma et al., 2021) consists of 20 Covid-19 CT scans. All cases contain Covid −19 infections. This dataset has been labeled by many experienced radiologists and the labeling is consistent. The entire lung mask, containing both normal and pathologic regions, was labeled. There are a total of 1844 infected slices out of 3520 slices.

For CC-CCII Segmentation Dataset (Zhang et al., 2020), lesions were segmented from CT slice images from the China Consortium of Chest CT Image Investigation (CC-CCII). A total of 750 CT slices from 150 Covid-19 patients were manually segmented into background, lung field, ground glass opacity (GGO), and consolidation.

The used datasets have long and similar names. To avoid confusion between names, we refer to them as Dataset_1, Dataset_2, Dataset_3, and Dataset_4, as shown in Table 2.

In our experiments, we explored both intra-dataset and cross-dataset scenarios. For intra-dataset experiments, Dataset_2 (Radiologists, 2019), Dataset_3 (Ma et al., 2021) and Dataset_4 (Zhang et al., 2020) are randomly split into 70%–30% as training and test splits, respecting the patient-independent splitting scenario in which all slices of a patient belong to only one split (training or validation). For the cross-dataset evaluation scenario, we used the trained models on Dataset_2 (trained on the training data 70%) and tested them on Dataset_1 (Radiologists, 2019), Dataset_3 (Ma et al., 2021), and Dataset_4 (Zhang et al., 2020), respectively. Similarly, we used the trained models on Dataset_3 (Ma et al., 2021) (trained on the training data 70%) and test them on Dataset_1 (Radiologists, 2019), Dataset_2 (Radiologists, 2019), and Dataset_4 (Zhang et al., 2020).

4.2. Evaluation metrics

To evaluate the performance of different approaches, we used five evaluation metrics, namely: F1 score (F1-S), Dice score (D-S), intersection over union (IoU), sensitivity (Sens), specificity (Spec), and precision (Prec).

F1 score, Intersection over Union (IoU), Sensitivity (Sens), Specificity (Spec) and Precision (Prec) are defined as follows:

$$\text{F1 score} = 100 \cdot \frac{2 \cdot TP}{2 \cdot TP + FP + FN} \tag{7}$$

$$IoU = 100 \cdot \frac{TP}{(TP + FP + FN)} \tag{8}$$

$$Sens = Recall = 100 \cdot \frac{TP}{TP + FN} \tag{9}$$

$$Spec = 100 \cdot \frac{TN}{FP + TN} \tag{10}$$

$$Prec = 100 \cdot \frac{TP}{TP + FP} \tag{11}$$

where TP is the True Positives, TN is the True Negatives, FP is the False Positives and FN is the False Negatives, all associated with the binary segmentation of the test images.

F1 score, intersection over union (IoU), sensitivity (Sens), specificity (Spec), and precision (Prec) are micro-metrics, which are calculated at the level of pixel. For these types of metrics, TP, TN, FP, and FN were first calculated for all test images, and then these metrics were calculated using the Eqs. (7), (8), (9), (10), and (11). However, the Dice score is the macro version of the F1 score (the F1 score is calculated for each test image, then the average over all test images is the Dice score). For N test images, the Dice score is defined as follows:

$$\text{Dice score} = 100 \cdot \frac{1}{N} \sum_{i=1}^{N} \frac{2 \cdot TP_i}{2 \cdot TP_i + FP_i + FN_i} \tag{12}$$

where TP_i, TN_i, FP_i and FN_i are True Positives, True Negatives, False Positives and False Negative for the ith image, respectively. In the experimental results, it is normal to have low dice score average since in the three datasets there is a large number of uninfected slices, where totally correct prediction of the masks (black mask) will give 0 dice score for these slices which will be included in calculating the average.

5. Experiments and results

5.1. Experimental setup

For deep learning training and testing, we used the Pytorch (Paszke et al., 2019) library with NVIDIA GPU Device GeForce TITAN RTX 24 GB. The batch size used consists of 6 images. We trained the networks for 60 epochs. The initial learning rate is 0.01, which decays by 0.1 after 30 epochs, followed by another decay of 0.1 after 50 epochs. Furthermore, we used active data augmentation techniques which are: rotation using random angle between −35 degrees to 35 degrees, and random horizontal and vertical flipping. The values of γ for the Edge loss in the proposed hybrid loss function (illustrated in Section 3.4) is experimentally set to 0.7, 2 and 0.9 for Dataset_2, Dataset_3 and Dataset_4, respectively.

5.2. Experimental results on intra-dataset scenario

In this section, we evaluate the performance of the proposed PDEAtt-Unet and compare it with Unet (Ronneberger et al., 2015), Att-Unet (Oktay et al., 2018), Unet++ (Zhou et al., 2018), InfNet (Fan et al., 2020), SCOATNet (Zhao et al., 2021a) and nCoVSegNet (Liu et al., 2021). Tables 3–5 show the experimental results using Dataset_2 (Radiologists, 2019), Dataset_3 (Ma et al., 2021) and Dataset_4 (Zhang et al., 2020), respectively. From the results of Tables 3, 4, and 5, our proposed approach (PDEAtt-Unet) outperforms three baseline segmentation architectures and three state-of-the-art approaches. Specifically, the experimental results of Dataset_2 show that our approach achieves 77.60%, 41.68%, and 63.40% for F1 score, Dice score, and IoU, respectively. In this dataset, the best competing approach is SCOATNet (Zhao et al., 2021a), which achieved 69.39%, 31.56%, and 53.13% for F1 score, Dice score, and IoU, respectively. Therefore, our approach outperformed the SCOATNet approach by 8.2%, 10.1%, and 10.3% for F1 score, Dice score, and IoU, respectively.

Using the intra-dataset experiment results for Dataset_3 (Table 4), we find that InfNet (Fan et al., 2020) outperforms the other competing methods, achieving 72.91%, 37.02%, and 57.37% for F1 score, Dice score, and IoU, respectively. On the other hand, our proposed PDEAtt-Unet approach outperforms all methods, especially InfNet by 2.60%, 4.25%, and 3.3% for F1 score, Dice score, and IoU, respectively.

Table 5 shows that nCoVSegNet (Liu et al., 2021) outperforms the other state-of-the-art comparison approaches (InfNet and SCOATNet) and the baseline architectures (Unet, Unet++, and Att-Unet). On the other hand, our proposed PDEAtt-Unet approach outperforms nCoVSegNet by 1.3%, 2.3%, and 1.8% for F1 score, Dice score, and IoU, respectively. The intra-dataset experiment results show that our proposed PDEAtt-Unet approach outperforms all competing methods and the second best approach differs from one dataset to another. This shows that our approach is efficient and consistent under different evaluation conditions.

5.3. Experimental results on cross-dataset scenario

In order to compare different segmentation architectures, it is important to investigate their generalization capabilities in cross-databases scenarios. Indeed, cross-database experiments play a crucial role in evaluating the effectiveness of each architecture in real-world scenarios. Tables 6 and 7 show the cross-dataset results of the trained models on Dataset_2 (Radiologists, 2019) and Dataset_3 (Ma et al., 2021), respectively. On the other hand, F1 score (F1-S), Dice score (D-S), and Intersection over Union (IoU) are used as evaluation metrics.

For the cross-dataset experiments, Dataset_2 (Radiologists, 2019) is used as training dataset and Dataset_1 (Radiologists, 2019), Dataset_3 (Ma et al., 2021) and Dataset_4 (Zhang et al., 2020) are used as test datasets. From the results of Table 6, the proposed PDEAtt-Unet architecture performs better than the state-of-the-art architectures and the baseline architectures in the three test datasets.

For the cross-dataset experiments with Dataset_3 (Ma et al., 2021) as the training dataset, Dataset_1 (Radiologists, 2019), Dataset_2 (Radiologists, 2019), and Dataset_4 (Zhang et al., 2020) are used as test datasets. Based on the results of Table 7, we can notice that the proposed PDEAtt-Unet outperforms all competing architectures in the three cross-dataset scenario.

From Tables 6 and 7, the comparison between the state-of-the-art architectures and the baselines shows that for each experiment, the best architecture among these architectures changes. More specifically, the best competing architecture for ***Experiments II, III and VI*** is the SCOATNet architecture. On the other hand, for ***Experiments I, IV and V***, the best architectures are Att-Unet, InfNet and Unet++, respectively. In contrast, our proposed approach achieves consistent performance across experiments. Similar to the intra-dataset results, the cross-dataset results also demonstrate the efficiency of the proposed architectures and the use of Edge loss to segment Covid-19 infections.

5.4. Ablation study

As shown in Fig. 3, our proposed approach consists of a pyramid encoder input, a dual decoders, and a hybrid loss function. To test the effectiveness of each component, we perform an ablation study using both intra-database and inter-database evaluation scenarios by discarding some components of the proposed PDEAtt-Unet.

Table 8 shows the ablation study for intra-dataset experiments. From this table, it can be seen that using Attention Pyramid paths improves the results of the three intra-dataset experiments compared to using the standard Att-Unet. For Dataset_2, the PAtt-Unet architecture improves Att-Unet results by 4%, 3.1%, and 4.4% for F1 score, Dice score, and IoU, respectively. Similarly, experiments with Dataset_3 and Dataset_4 show the efficiency of using the encoder Attention Pyramids. For Dataset_3 and Dataset_4, the F1 score is improved by 3.1% and 0.5%, respectively, the Dice score is improved by 3.6% and 0.7%, respectively, and the IoU is improved by 4.8% and 0.7%, respectively. The second component of our proposed approach is the use of a dual decoders. The results compared to using Att-Unet show the superiority of the proposed DAtt-Unet architecture on three datasets (Dataset_2, Dataset_3, and Dataset_4) and three metrics (F1 score, Dice score, and IoU).

Moreover, the combination of Attention Pyramid input and dual decoders (PDAtt-Unet) improves the results even more. As can be

Table 3
Experimental results of intra-dataset scenario using Dataset_2 (Radiologists, 2019). The following architectures are evaluated for Covid-19 infection segmentation: InfNet (Fan et al., 2020), SCOATNet (Zhao et al., 2021a), nCoVSegNet (Liu et al., 2021), Unet, Unet++, and Att-Unet and the proposed PDEAtt-Unet.

Model	F1-S	D-S	IoU	Sens	Spec	Prec
InfNet (Fan et al., 2020)	40.86	15.02	25.68	35.11	99.06	48.88
SCOATNet (Zhao et al., 2021a)	69.39	31.56	53.13	56.67	99.83	89.46
nCoVSegNet (Liu et al., 2021)	55.41	26.57	38.32	41.10	99.81	85.01
Unet	66.38	28.95	49.68	52.15	99.87	91.30
UNet++	59.29	26.49	42.13	43.23	99.93	94.32
Att-Unet	59.32	27.55	42.17	45.62	99.79	84.79
PDEAtt-Unet	**77.60**	**41.68**	**63.40**	**80.69**	**99.30**	**74.74**

Table 4
Experimental results of intra-dataset scenario using Dataset_3 (Ma et al., 2021). The following architectures are evaluated for Covid-19 infection segmentation: InfNet (Fan et al., 2020), SCOATNet (Zhao et al., 2021a), nCoVSegNet (Liu et al., 2021), Unet, Unet++, and Att-Unet and the proposed PDEAtt-Unet.

Model	F1-S	D-S	IoU	Sens	Spec	Prec
InfNet (Fan et al., 2020)	72.91	37.02	57.37	66.48	99.78	80.72
SCOATNet (Zhao et al., 2021a)	71.02	36.74	55.06	61.35	99.84	84.29
nCoVSegNet (Liu et al., 2021)	62.48	35.95	45.44	50.16	99.86	82.85
Unet	65.82	35.73	49.05	53.32	99.88	85.97
UNet++	69.89	37.43	53.72	60.69	99.82	82.39
Att-Unet	68.67	33.42	52.29	58.34	99.84	83.45
PDEAtt-Unet	**75.52**	**41.28**	**60.67**	**81.94**	**99.52**	**70.03**

Table 5
Experimental results of intra-dataset scenario using Dataset_4 (Zhang et al., 2020). The following architectures are evaluated for Covid-19 infection segmentation: InfNet (Fan et al., 2020), SCOATNet (Zhao et al., 2021a), nCoVSegNet (Liu et al., 2021), Unet, Unet++, and Att-Unet and the proposed PDEAtt-Unet.

Model	F1-S	D-S	IoU	Sens	Spec	Prec
InfNet (Fan et al., 2020)	72.19	40.80	56.49	63.62	99.86	83.44
SCOATNet (Zhao et al., 2021a)	77.57	46.99	63.25	70.66	99.86	85.12
nCoVSegNet (Liu et al., 2021)	78.06	47.20	64.02	74.69	99.82	81.74
Unet	75.49	45.34	60.63	69.34	99.84	82.84
Unet++	77.45	46.99	63.20	71.55	99.85	84.42
AttUnet	77.27	46.32	62.96	70.48	99.87	85.51
PDEAtt-Unet	**79.38**	**49.50**	**65.81**	**82.52**	**99.72**	**76.47**

Table 6
Cross datasets results by using the trained model of Dataset_2 (Radiologists, 2019).

Testing dataset	Model	F1-S	D-S	IoU
Experiment I: Dataset_1 (Radiologists, 2019)	InfNet (Fan et al., 2020)	39.28	32.83	24.44
	SCOATNet (Zhao et al., 2021a)	49.24	40.82	32.66
	nCoVSegNet (Liu et al., 2021)	37.83	31.74	23.33
	Unet	49.34	34.57	32.75
	Unet++	26.59	24.79	15.33
	Att-Unet	50.31	41.01	30.34
	PDEAtt-Unet	**57.50**	**46.13**	**40.35**
Experiment II: Dataset_3 (Ma et al., 2021)	InfNet (Fan et al., 2020)	44.96	19.24	29.01
	SCOATNet (Zhao et al., 2021a)	60.55	27.36	43.43
	nCoVSegNet (Liu et al., 2021)	57.56	28.34	40.41
	Unet	51.24	25.29	34.14
	Unet++	51.94	23.03	35.08
	Att-Unet	56.61	25.90	39.48
	PDEAtt-Unet	**65.20**	**29.45**	**47.08**
Experiment III: Dataset_4 (Zhang et al., 2020)	InfNet (Fan et al., 2020)	39.43	18.16	24.55
	SCOATNet (Zhao et al., 2021a)	42.41	21.40	26.91
	nCoVSegNet (Liu et al., 2021)	28.21	12.61	16.42
	Unet	19.39	14.67	17.17
	UNet++	25.36	18.77	20.58
	Att-Unet	41.89	17.24	20.29
	PDEAtt-Unet	**46.69**	**24.39**	**30.45**

Table 7
Cross datasets results by using the trained model of Dataset_3 (Ma et al., 2021).

Testing dataset	Model	F1-S	D-S	IoU
Experiment IV: Dataset_1 (Radiologists, 2019)	InfNet (Fan et al., 2020)	67.83	60.19	50.03
	SCOATNet (Zhao et al., 2021a)	63.36	59.52	46.37
	nCoVSegNet (Liu et al., 2021)	66.37	58.19	49.66
	Unet	62.15	55.37	45.08
	Unet++	67.45	58.43	50.88
	Att-Unet	44.95	38.37	28.99
	PDEAtt-Unet	**69.99**	**63.18**	**53.83**
Experiment V: Dataset_2 (Radiologists, 2019)	InfNet (Fan et al., 2020)	51.88	21.66	35.03
	SCOATNet (Zhao et al., 2021a)	20.61	18.23	20.50
	nCoVSegNet (Liu et al., 2021)	54.29	25.05	37.26
	Unet	42.99	19.74	27.38
	Unet++	70.30	26.95	54.20
	Att-Unet	55.48	25.10	38.39
	PDEAtt-Unet	**72.51**	**27.96**	**56.43**
Experiment VI: Dataset_4 (Zhang et al., 2020)	InfNet (Fan et al., 2020)	57.69	30.29	40.54
	SCOATNet (Zhao et al., 2021a)	63.77	34.22	46.81
	nCoVSegNet (Liu et al., 2021)	60.03	32.90	42.89
	Unet	41.54	16.54	26.22
	Unet++	39.55	16.84	24.65
	Att-Unet	53.13	23.78	36.17
	PDEAtt-Unet	**66.74**	**38.90**	**50.08**

seen in Table 8, PDAtt-Unet improves the results of Dataset_2 compared to Att-Unet by 14.5%, 8%, and 16.5% for F1 score, Dice score, and IoU, respectively. The improvements for Dataset_3 are 5.15% and 5.76%, 6.2%, respectively. Similarly, the results for Dataset_4 are improved by 1.37%, 1.52%, and 1.85% for F1 score, Dice score, and IoU, respectively. On the other hand, training the proposed PDAtt-Unet architecture with the proposed hybrid loss (PDEAtt-Unet) achieves better performance than with the global loss, as shown in Table 8.

Tables 9 and 10 show the ablation study for cross-dataset experiments where Dataset_2 and Dataset_3 are used as training data, respectively. From the results of Tables 9 and 10, we notice that both PAtt-Unet and DAtt-Unet architectures outperform the baseline result

Table 8
Ablation study of intra-dataset evaluation scenario. The experimental results of Dataset_2, Dataset_3 and Dataset_4 are summarized with investigating the effectiveness of the following components, Attention Pyramid Input, Dual-Decoder and Hybrid loss function.

Architecture	Ablation			Dataset_2			Dataset_3			Dataset_4		
	Pyramid	Dual	Edge	F1-S	D-S	IoU	F1-S	D-S	IoU	F1-S	D-S	IoU
Att-Unet	✗	✗	✗	59.32	27.55	42.17	68.67	33.42	52.29	77.27	46.32	62.96
PAtt-Unet	✓	✗	✗	63.51	30.69	46.53	71.75	38.22	55.95	77.76	47.06	63.62
DAtt-Unet	✗	✓	✗	64.04	29.35	47.10	70.73	36.49	54.71	78.15	47.33	64.14
PDAtt-Unet	✓	✓	✗	73.89	35.55	58.59	73.83	39.18	58.52	78.64	47.72	64.81
PDEAtt-Unet	✓	✓	✓	**77.60**	**41.68**	**63.40**	**75.52**	**41.28**	**60.67**	**79.38**	**49.50**	**65.81**

Table 9
Ablation study of Cross-datasets evaluation scenario. The experimental results of using Dataset_2 as training data and Dataset_1, Dataset_3 and Dataset_4 as testing datasets are summarized with investigating the effectiveness of the following components, Attention Pyramid Input, Dual-Decoder and Hybrid loss function.

Architecture	Ablation			Dataset_1			Dataset_3			Dataset_4		
	Pyramid	Dual	Edge	F1-S	D-S	IoU	F1-S	D-S	IoU	F1-S	D-S	IoU
Att-Unet	✗	✗	✗	50.31	41.01	30.34	56.61	25.90	39.48	41.89	17.24	20.29
PAtt-Unet	✓	✗	✗	55.54	44.43	38.45	58.20	26.36	43.47	44.77	21.34	28.25
DAtt-Unet	✗	✓	✗	52.30	42.51	31.84	59.65	25.52	42.51	43.75	21.15	27.47
PDAtt-Unet	✓	✓	✗	56.12	44.22	39.01	64.02	28.65	48.36	45.75	21.44	29.66
PDEAtt-Unet	✓	✓	✓	**57.50**	**46.13**	**40.35**	**65.20**	**29.45**	**47.08**	**46.69**	**24.39**	**30.45**

Table 10
Ablation study of Cross-datasets evaluation scenario. The experimental results of using Dataset_3 as training data and Dataset_1, Dataset_2 and Dataset_4 as testing datasets are summarized with investigating the effectiveness of the following components, Attention Pyramid Input, Dual-Decoder and Hybrid loss function.

Architecture	Ablation			Dataset_1			Dataset_2			Dataset_4		
	Pyramid	Dual	Edge	F1-S	D-S	IoU	F1-S	D-S	IoU	F1-S	D-S	IoU
Att-Unet	✗	✗	✗	44.95	38.37	28.99	55.48	25.10	38.39	53.13	23.78	36.17
PAtt-Unet	✓	✗	✗	68.07	61.08	51.60	70.42	27.84	54.35	63.64	34.90	46.67
DAtt-Unet	✗	✓	✗	68.70	61.77	52.33	63.21	27.32	46.21	64.56	36.96	47.67
PDAtt-Unet	✓	✓	✗	69.25	62.84	52.96	70.60	27.21	55.33	65.17	37.91	48.02
PDEAtt-Unet	✓	✓	✓	**69.99**	**63.18**	**53.83**	**72.51**	**27.96**	**56.43**	**66.74**	**38.90**	**50.08**

of Att-Unet by a significant margin for all cross-dataset experiments, as shown by the three evaluation metrics. Moreover, the combination of the encoder Attention Pyramids and the dual encoders further improves all results. Similar to the results in Table 8, using the proposed hybrid loss function to train PDAtt-Unet achieves better performance than using the global loss.

On the other hand, comparing the intra-dataset results within Tables 4, 3, 5, and 8 shows that the performance of PAtt-Unet, DAtt-Unet, and PDAtt-Unet is better than the baseline architectures in the three datasets. Moreover, PAtt-Unet, DAtt-Unet, and PDAtt-Unet perform better or close to better than the best state-of-the-art approach of InfNet (Fan et al., 2020), SCOATNet (Zhao et al., 2021a), and nCoVSegNet (Liu et al., 2021). Similarly, the comparison of cross-dataset results shows that PAtt-Unet, DAtt-Unet, and PDAtt-Unet perform better than the state-of-the-art approaches in most experiments (Tables 6, 7, 9, and 10).

6. Discussion

To show the effectiveness of the proposed PDAtt-Unet trained with the proposed hybrid loss function (PDEAtt-Unet), we visualized the generated segmented masks with nCoVSegNet (Liu et al., 2021), SCOATNet (Zhao et al., 2021a), Unet++ (Zhou et al., 2018), Att-Unet (Oktay et al., 2018), and PDEAtt-Unet, as shown in Fig. 4. The first row shows an example of an infected slice with multiple infection regions. From the segmented mask of the proposed PDEAtt-Unet, we can see that PDEAtt-Unet is able to accurately segment the large and small infection regions. In contrast, the other segmented masks show that the other segmentation methods are not able to accurately segment the Covid-19 infections, especially the small and border regions. The second row shows an example of a Covid-19 infection that is a mixture of GGO and consolidation with posterior distribution. Based on the segmented masks, we see that the competing methods could not segment the GGO infection in the upper lobe of the left lung and the borders of the other infections. In contrast, the proposed PDEAtt-Unet is able to segment the GGO infection, although the contrast with the uninfected lung is low. The third and fourth rows show examples of two slices with severe infection in which most lung regions are infected with a mixture of GGO and consolidation. The competing architectures suffer from non-lung tissue being segmented as infection. In contrast, the proposed PDEAtt-Unet is able to accurately segment Covid-19 infection. This demonstrates the efficiency of using two decoders for two tasks (infection and lung segmentation).

7. Conclusion

In this work, we proposed an approach to segment Covid-19 infections from CT scans. Our proposed PDEAtt-Unet is an encoder–decoder CNN architecture. The encoder is a Unet-like encoder with pyramid paths and attention gates aimed to preserve the general structure awareness for all encoder layers. The decoder consists of two Att-Unet-like decoders that simultaneously segment the Covid-19 infection and lung regions. To circumvent the limitations of the BCE loss function, edge loss is proposed.

To prove the efficiency of our approach, we evaluated its performance in both scenarios; within a dataset and across data. Moreover, we compared the performance of our proposed approach with three baseline architectures (Unet, Att-Unet, Unet++) and three state-of-the-art architectures for Covid-19 segmentation (InfNet, SCOATNet, and nCoVSegNet). The experimental results showed the superiority of our proposed approach compared to the baseline segmentation architectures and to the three state-of-the-art architectures in both intra-database and cross-database evaluation scenarios. The proposed components of our approach have proved their efficiency individually. Moreover, their combination leads to better performance

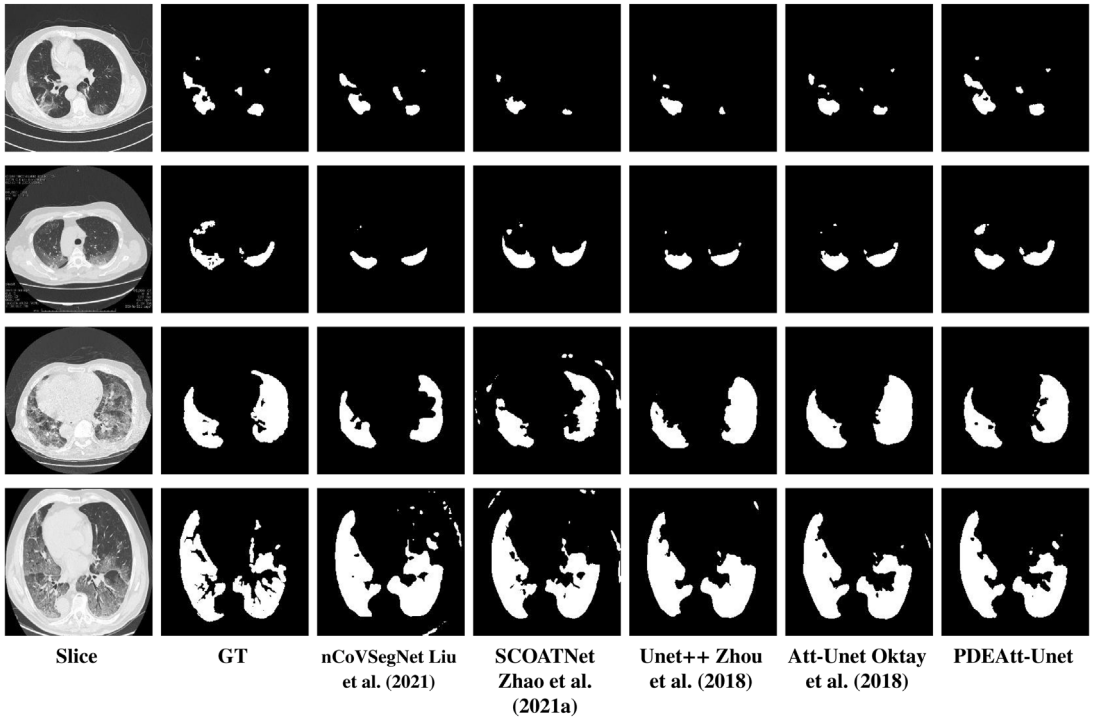

Fig. 4. Visual comparison of a segmentation model trained with different segmentation architectures. The first row shows visualization example from the validation data of Dataset_3 (Ma et al., 2021). The rest three examples are from the cross-datasets scenario, where the models were trained using Dataset_3 (Ma et al., 2021) and tested on Dataset_1 (Radiologists, 2019).

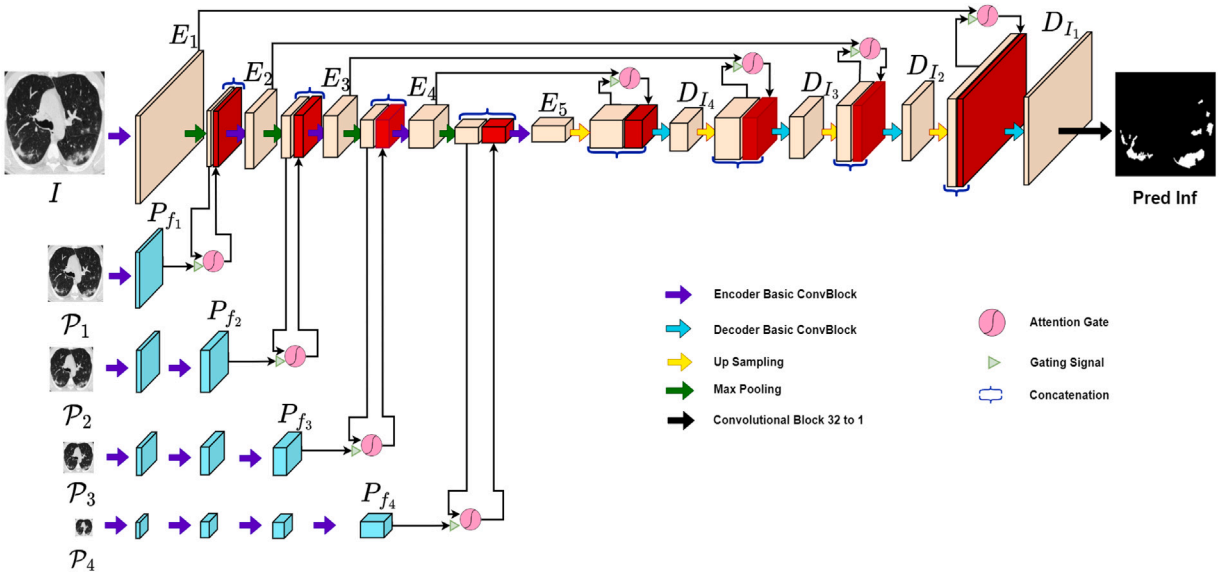

Fig. A.5. Our proposed PAtt-Unet architecture.

with stable and consistent results in all evaluated scenarios. Future work could investigate the performance of the proposed solution in other segmentation tasks in medical imaging with other infections and organs.

Declaration of competing interest

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

Data availability

Data will be made available on request.

Acknowledgment

The authors would like to thank Arturo Argentieri from CNR-ISASI Italy for his support on the multi-GPU computing facilities.

Appendix. PAtt-Unet and DAtt-Unet detailed architectures

See Figs. A.5 and A.6.

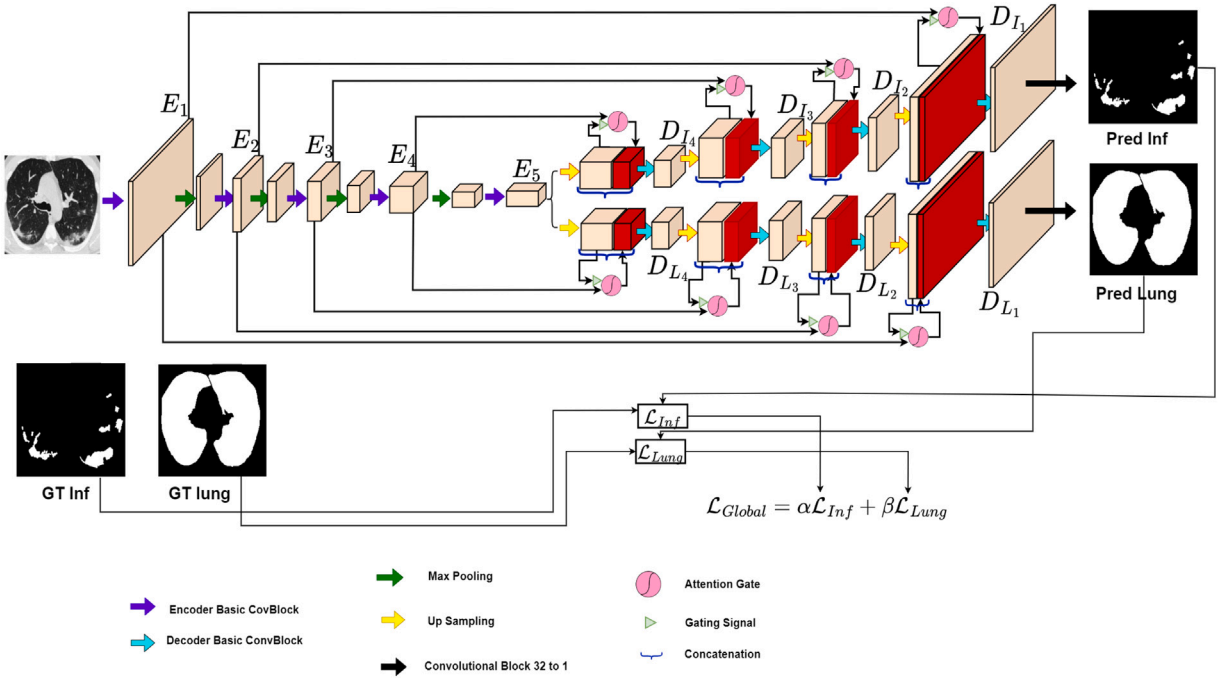

Fig. A.6. Our proposed DAtt-Unet architecture.

References

Abdel-Basset, M., Chang, V., Hawash, H., Chakrabortty, R.K., Ryan, M., 2021. FSS-2019-nCov: A deep learning architecture for semi-supervised few-shot segmentation of COVID-19 infection. Knowl.-Based Syst. 212, 106647. http://dx.doi.org/10.1016/j.knosys.2020.106647.

Bougourzi, F., Contino, R., Distante, C., Taleb-Ahmed, A., 2021a. CNR-IEMN: A deep learning based approach to recognise COVID-19 from CT-scan. In: ICASSP 2021-2021 IEEE International Conference on Acoustics, Speech and Signal Processing. ICASSP, IEEE, pp. 8568–8572.

Bougourzi, F., Contino, R., Distante, C., Taleb-Ahmed, A., 2021b. Recognition of COVID-19 from CT Scans Using Two-Stage Deep-Learning-Based Approach: CNR-IEMN. Sensors 21 (17), 5878. http://dx.doi.org/10.3390/s21175878, Number: 17 Publisher: Multidisciplinary Digital Publishing Institute.

Bougourzi, F., Distante, C., Dornaika, F., Taleb-Ahmed, A., 2023. CNR-IEMN-CD and CNR-IEMN-CSD approaches for Covid-19 detection and Covid-19 severity detection from 3D CT-scans. In: Computer Vision–ECCV 2022 Workshops: Tel Aviv, Israel, October 23–27, 2022, Proceedings, Part VII. Springer, pp. 593–604.

Bougourzi, F., Distante, C., Ouafi, A., Dornaika, F., Hadid, A., Taleb-Ahmed, A., 2021c. Per-COVID-19: A Benchmark Dataset for COVID-19 Percentage Estimation from CT-Scans. J. Imaging 7 (9), 189. http://dx.doi.org/10.3390/jimaging7090189, Number: 9 Publisher: Multidisciplinary Digital Publishing Institute.

Bougourzi, F., Dornaika, F., Mokrani, K., Taleb-Ahmed, A., Ruichek, Y., 2020. Fusing Transformed Deep and Shallow features (FTDS) for image-based facial expression recognition. Expert Syst. Appl. 156, 113459. http://dx.doi.org/10.1016/j.eswa.2020.113459.

Bougourzi, F., Dornaika, F., Taleb-Ahmed, A., 2022. Deep learning based face beauty prediction via dynamic robust losses and ensemble regression. Knowl.-Based Syst. 242, 108246.

Elharrouss, O., Subramanian, N., Al-Maadeed, S., 2021. An Encoder–Decoder-Based Method for Segmentation of COVID-19 Lung Infection in CT Images. SN Comput. Sci. 3 (1), 13. http://dx.doi.org/10.1007/s42979-021-00874-4.

Fan, D.-P., Zhou, T., Ji, G.-P., et al., 2020. Inf-Net: Automatic COVID-19 Lung Infection Segmentation From CT images. IEEE Trans. Med. Imaging 39 (8), 2626–2637. http://dx.doi.org/10.1109/TMI.2020.2996645, Conference Name: IEEE Transactions on Medical Imaging.

Fung, D.L.X., Liu, Q., Zammit, J., Leung, C.K.-S., Hu, P., 2021. Self-supervised deep learning model for COVID-19 lung CT image segmentation highlighting putative causal relationship among age, underlying disease and COVID-19. J. Transl. Med. 19 (1), 318. http://dx.doi.org/10.1186/s12967-021-02992-2.

Jin, Y.-H., Cai, L., Cheng, Z.-S., et al., 2020. A rapid advice guideline for the diagnosis and treatment of 2019 novel coronavirus (2019-nCoV) infected pneumonia (standard version). Mil. Med. Res. 7 (1), 4. http://dx.doi.org/10.1186/s40779-020-0233-6.

Kim, J.Y., Choe, P.G., Oh, Y., et al., 2020. The first case of 2019 novel coronavirus pneumonia imported into Korea from Wuhan, China: implication for infection prevention and control measures. J. Korean Med. Sci. 35 (5), Publisher: The Korean Academy of Medical Sciences.

Kucirka, L.M., Lauer, S.A., Laeyendecker, O., Boon, D., Lessler, J., 2020. Variation in False-Negative rate of reverse transcriptase polymerase chain Reaction–Based SARS-CoV-2 Tests by time since exposure. Ann. Internal Med. 173 (4), 262–267. http://dx.doi.org/10.7326/M20-1495, Publisher: American College of Physicians.

Kumar Singh, V., Abdel-Nasser, M., Pandey, N., Puig, D., 2021. LungINFseg: Segmenting COVID-19 Infected Regions in Lung CT Images Based on a Receptive-Field-Aware Deep Learning Framework. Diagnostics 11 (2), 158. http://dx.doi.org/10.3390/diagnostics11020158, Number: 2 Publisher: Multidisciplinary Digital Publishing Institute.

Laradji, I., Rodriguez, P., Manas, O., et al., 2021. A weakly supervised consistency-based learning method for COVID-19 segmentation in CT images. In: Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision. WACV, pp. 2453–2462.

Liu, J., Dong, B., Wang, S., Cui, H., Fan, D.-P., Ma, J., Chen, G., 2021. COVID-19 lung infection segmentation with a novel two-stage cross-domain transfer learning framework. Med. Image Anal. 74, 102205. http://dx.doi.org/10.1016/j.media.2021.102205.

Ma, J., Wang, Y., An, X., et al., 2021. Toward data efficient learning: A benchmark for COVID 19 CT lung and infection segmentation. Med. Phys. 48, 1197–1210. http://dx.doi.org/10.1002/mp.14676.

Morozov, S., Andreychenko, A., Pavlov, N., et al., 2020. MosMedData: Chest CT Scans with COVID-19 Related Findings Dataset. Radiol. Imaging http://dx.doi.org/10.1101/2020.05.20.20100362, preprint.

Mu, N., Wang, H., Zhang, Y., et al., 2021. Progressive global perception and local polishing network for lung infection segmentation of COVID-19 CT images. Pattern Recognit. 120, 108168. http://dx.doi.org/10.1016/j.patcog.2021.108168.

Muhammad, U., Hoque, M.Z., Oussalah, M., Keskinarkaus, A., Seppänen, T., Sarder, P., 2022. SAM: Self-augmentation mechanism for COVID-19 detection using chest X-ray images. Knowl.-Based Syst. 241, 108207.

Oktay, O., Schlemper, J., Folgoc, L.L., et al., 2018. Attention U-Net: Learning Where to Look for the Pancreas. arXiv:1804.03999 [Cs].

Pan, Y., Guan, H., Zhou, S., et al., 2020b. Initial CT findings and temporal changes in patients with the novel coronavirus pneumonia (2019-nCoV): a study of 63 patients in Wuhan, China. Eur. Radiol. 30 (6), 3306–3309, Publisher: Springer.

Pan, F., Ye, T., Sun, P., et al., 2020a. Time Course of Lung Changes at Chest CT during Recovery from Coronavirus Disease 2019 (COVID-19). Radiology 295 (3), 715–721. http://dx.doi.org/10.1148/radiol.2020200370.

Paszke, A., Gross, S., Massa, F., et al., 2019. Pytorch: An imperative style, high-performance deep learning library. In: Advances in Neural Information Processing Systems. pp. 8026–8037.

Prevedello, L.M., Halabi, S.S., Shih, G., et al., 2019. Challenges Related to Artificial Intelligence Research in Medical Imaging and the Importance of Image Analysis

Competitions. Radiol.: Artif. Intell. 1 (1), e180031. http://dx.doi.org/10.1148/ryai.2019180031, Publisher: Radiological Society of North America.

Qin, X., Zhang, Z., Huang, C., Gao, C., Dehghan, M., Jagersand, M., 2019. Basnet: Boundary-aware salient object detection. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 7479–7489.

Radiologists, 2019. COVID-19 CT-scans segmentation datasets. available at: http://medicalsegmentation.com/covid19/. Last visited: 18-08-2021.

Ronneberger, O., Fischer, P., Brox, T., 2015. U-Net: Convolutional Networks for Biomedical Image Segmentation. In: Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015. Springer International Publishing, Cham, pp. 234–241.

Roth, H.R., Xu, Z., Tor-Díez, C., Jacob, R.S., Zember, J., Molto, J., Li, W., Xu, S., Turkbey, B., Turkbey, E., et al., 2022. Rapid artificial intelligence solutions in a pandemic—The COVID-19-20 lung CT lesion segmentation challenge. Med. Image Anal. 82, 102605.

Salehi, S., Abedi, A., Balakrishnan, S., Gholamrezanezhad, A., 2020. Coronavirus disease 2019 (COVID-19): a systematic review of imaging findings in 919 patients. Ajr Am. J. Roentgenol. 215 (1), 87–93.

Santa Cruz, B.G., Bossa, M.N., Sölter, J., Husch, A.D., 2021. Public covid-19 x-ray datasets and their impact on model bias–a systematic review of a significant problem. Med. Image Anal. 74, 102225.

Sun, Z., Zhang, N., Li, Y., Xu, X., 2020. A systematic review of chest imaging findings in COVID-19. Quant. Imaging Med. Surg. 10 (5), 1058–1079. http://dx.doi.org/10.21037/qims-20-564.

Vantaggiato, E., Paladini, E., Bougourzi, F., Distante, C., Hadid, A., Taleb-Ahmed, A., 2021. COVID-19 recognition using ensemble-CNNs in two new chest X-ray databases. Sensors 21 (5), http://dx.doi.org/10.3390/s21051742.

Wang, R., Ji, C., Zhang, Y., Li, Y., 2022. Focus, Fusion, and Rectify: Context-Aware Learning for COVID-19 Lung Infection Segmentation. IEEE Trans. Neural Netw. Learn. Syst. 33 (1), http://dx.doi.org/10.1109/TNNLS.2021.3126305, Conference Name: IEEE Transactions on Neural Networks and Learning Systems.

Wang, G., Liu, X., Li, C., Xu, Z.e.a., 2020. A Noise-Robust Framework for Automatic Segmentation of COVID-19 Pneumonia Lesions From CT Images. IEEE Trans. Med. Imaging 39 (8), 2653–2663. http://dx.doi.org/10.1109/TMI.2020.3000314, Conference Name: IEEE Transactions on Medical Imaging.

Wu, Y.-H., Gao, S.-H., Mei, J., et al., 2021. JCS: An Explainable COVID-19 Diagnosis System by Joint Classification and Segmentation. IEEE Trans. Image Process. 30, 3113–3126. http://dx.doi.org/10.1109/TIP.2021.3058783, Conference Name: IEEE Transactions on Image Processing.

Yang, D., Xu, Z., Li, et al., 2021. Federated semi-supervised learning for COVID region segmentation in chest CT using multi-national data from China, Italy, Japan. Med. Image Anal. 70, 101992. http://dx.doi.org/10.1016/j.media.2021.101992.

Zhang, K., Liu, X., Shen, J., et al., 2020. Clinically Applicable AI System for Accurate Diagnosis, Quantitative Measurements, and Prognosis of COVID-19 Pneumonia Using Computed Tomography. Cell 181 (6), 1423–1433.e11. http://dx.doi.org/10.1016/j.cell.2020.04.045.

Zhao, S., Li, Z., Chen, Y., et al., 2021a. SCOAT-net: A novel network for segmenting COVID-19 lung opacification from CT images. Pattern Recognit. 108109.

Zhao, X., Zhang, P., Song, F., et al., 2021b. D2A U-Net: Automatic segmentation of COVID-19 CT slices based on dual attention and hybrid dilated convolution. Comput. Biol. Med. 135, 104526. http://dx.doi.org/10.1016/j.compbiomed.2021.104526.

Zhou, Z., Rahman Siddiquee, M.M., Tajbakhsh, N., Liang, J., 2018. UNet++: A Nested U-Net Architecture for Medical Image Segmentation. In: Stoyanov, D., Taylor, Z., Carneiro, G., et al. (Eds.), Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support. Springer International Publishing, Cham, pp. 3–11.