ADU-Net: An Attention Dense U-Net based deep supervised DNN for automated lesion segmentation of COVID-19 from chest CT images

Sanjib Saha [a,b,*], Subhadeep Dutta [b], Biswarup Goswami [c], Debashis Nandi [a]

[a] Department of Computer Science and Engineering, National Institute of Technology, Durgapur, 713209, West Bengal, India
[b] Department of Computer Science and Engineering, Dr. B. C. Roy Engineering College, Durgapur, 713206, West Bengal, India
[c] Department of Respiratory Medicine, Health and Family Welfare, Government of West Bengal, Kolkata, 700091, West Bengal, India

ARTICLE INFO

ABSTRACT

An automatic method for qualitative and quantitative evaluation of chest Computed Tomography (CT) images is essential for diagnosing COVID-19 patients. We aim to develop an automated COVID-19 prediction framework using deep learning.

We put forth a novel Deep Neural Network (DNN) composed of an attention-based dense U-Net with deep supervision for COVID-19 lung lesion segmentation from chest CT images. We incorporate dense U-Net where convolution kernel size 5×5 is used instead of 3×3. The dense and transition blocks are introduced to implement a densely connected network on each encoder level. Also, the attention mechanism is applied between the encoder, skip connection, and decoder. These are used to keep both the high and low-level features efficiently. The deep supervision mechanism creates secondary segmentation maps from the features. Deep supervision combines secondary supervision maps from various resolution levels and produces a better final segmentation map. The trained artificial DNN model takes the test data at its input and generates a prediction output for COVID-19 lesion segmentation. The proposed model has been applied to the MedSeg COVID-19 chest CT segmentation dataset. Data pre-processing methods help the training process and improve performance.

We compare the performance of the proposed DNN model with state-of-the-art models by computing the well-known metrics: dice coefficient, Jaccard coefficient, accuracy, specificity, sensitivity, and precision. As a result, the proposed model outperforms the state-of-the-art models.

This new model may be considered an efficient automated screening system for COVID-19 diagnosis and can potentially improve patient health care and management system.

1. Introduction

The worldwide pandemic caused by COVID-19 has badly affected human life and healthcare services. The SARS-CoV-2 virus variants[1] are genetically changing over time, circulating globally [1], and causing COVID-19 disease. The World Health Organization[2] reports that COVID-19 confirmed cases are more than 526 million and COVID-19 death cases are more than 6.2 million all over the world as of 30th May 2022. The COVID-19 disease may cause cytokine storms involving multiple organ failures, including lungs, leading to COVID-19 pneumonia that ultimately leads to pulmonary fibrosis in all age groups [2]. Therefore, early diagnosis of pulmonary is very crucial. Chest CT [3–5] is an important tool in diagnosing clinically suspicious COVID-19 pneumonia early. The hospitals and health centres of developing and underdeveloped countries are equipped with High-Resolution CT (HRCT) and Magnetic Resonance Imaging (MRI) at a higher cost. The MRI applies magnetic waves instead of ionizing radiation like a CT or X-ray. However, a chest CT scan is more beneficial in detecting and quantifying COVID-19 than MRI. We are inspired to develop an automated screening system using deep learning that can classify COVID-19 cases and identify the level of lung infections from conventional chest CT images. Though X-ray images are more inexpensive than CT images, they produce a high error rate, so CT images are useful for more accurate diagnosis [6]. Several slices of CT images are produced for each patient at the time of scans. These create a high workload on clinicians to detect COVID-19 manually [7]. Deep neural network models have recently brought a groundbreaking revolution in developing automated computer-assisted diagnosis (CAD)
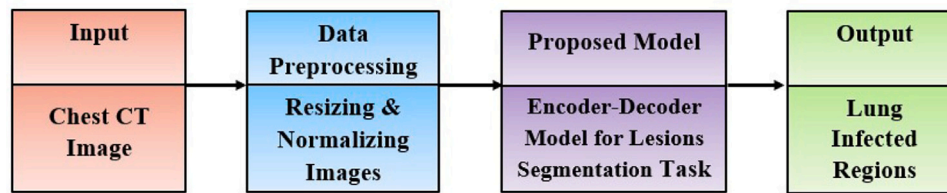
Fig. 1. The workflow diagram of our proposed model consists of four modules: input COVID-19 chest CT, data preprocessing, proposed encoder–decoder ADU-Net model, and output COVID-19 lesion segmentation.

systems to improve diagnostic accuracy in the health care system. Most technologists and researchers have concentrated their research on developing deep learning models to design automated CAD-based disease screening systems [8–12]. However, it is observed that not a single deep neural network model can give the same performance for all medical image modalities and diseases. A considerable amount of data with pre-defined classes or labels is required to train the DNN model. However, labelled data may only sometimes be available. All these issues compel the model developers to think about different models. During the last two years, the world has been severely stricken by the COVID-19 virus, and the virus has caused a massive loss of lives. The virus primarily attacks the lung of the infected person and causes the severe respiratory problem. Moreover, it is highly infectious. Hence, quick detection of COVID-19 can help to isolate the patients. Conventional RT-PCR (Reverse Transcription Polymerase Chain Reaction) [4,5] test takes more than 24 h to get reports which may delay the patient isolation. On the other hand, the imaging-based diagnosis of COVID-19 may be quicker and help early isolation of the patients. Besides this, if the lungs are already attacked, it is required to observe the areas of infection in the lungs. Because of this, the research on imaging-based COVID-19 diagnosis has gained tremendous momentum over the last two years. Many models have been proposed so far [13–18]. Various U-Net-based models like 3D U-Net [19], VB-Net [20], U-Net with VGG16 encoder [21], and nnU-Net [22] models have been developed for segmenting lesions in COVID-19 applications. However, they are unable to provide desirable accuracy in diagnosis. Hence, there is a stringent requirement to develop a better model that can improve the performance of COVID-19 detection.

In this paper, we have proposed a novel DNN model that improves the accuracy in the segmentation of COVID-19 lesions from chest CT. The novelty of this work lies in a two-fold contribution. (1) Development of a dense U-Net architecture. (2) Introducing attention module and deep supervision in the proposed dense U-Net. The dense architecture removes the vanishing gradient problem, the attention gate keeps the meaningful semantic features, and the deep supervision integrates additional useful multi-resolution features which ultimately improve the accuracy of state-of-the-art models.

The state-of-the-art literature on COVID-19 is summarized in Section 2. Section 3 presents the proposed deeply supervised, densely connected attention U-Net for COVID-19 lung lesion segmentation. A summary of the data set used in the research work is given in Section 4. Section 5 portrays the experiments, their outcomes, comparison with the state-of-the-art, analysis and discussion on the outcomes. Finally, the paper is concluded in Section 6.

2. Related works

A significant amount of research has been done to identify COVID-19 patients quickly using X-ray, CT and HRCT images to isolate the infected patients in the last two years. At the same time, imaging-based diagnosis is expected to measure the severity of the infection in the lungs. In addition, the segmentation of the infected area in the lungs can show how much area of the lungs is infected. Therefore, many researchers have put their efforts into implementing DNN models that can effectively apply to COVID-19 management. In this section, we provide

a summary of recent works. Different models and their performances are listed in Table 1. It has been observed that performance is not desired for transfer learning using standard deep neural models. However, transfer learning can extract useful features from fewer datasets in less time. However, transfer learning can transfer negative features. It works if reference and target problems are identical. Therefore, many researchers are interested in developing their proposed models.

A. Amyar et al. [17], A. K. Mishra et al. [23], A. Mobiny et al. [24], Md. Rahimzadeh et al. [25], P. Kalane et al. [26], G. Jia et al. [27], and V. Arora et al. [28] have developed new DNN models and improved their performance to some extent. However, there are some scopes to improve in the architectures, training procedures and hyperparameters selections of those existing models. The performance could be better using various generative adversarial networks implemented by P. Zhang et al. [29] and J. Zhang et al. [30] for lesion segmentation. The neural model developed by J. Ma et al. [22] could be getting better performance due to inappropriate training procedures. Y. Wang et al. [31] have used transfer learning in the model and failed to attain top performance in the literature.

The research domain posted here is based on deep neural network models and focuses on automatic COVID-19 lesion segmentation from chest CT. The recent works on COVID-19 lesion segmentation reflect that we have enough scope of work to design newer and more efficient models to achieve better accuracy than the state-of-the-art models.

3. Methodologies

3.1. Proposed deep neural network (ADU-Net)

We propose a new and efficient DNN, a deeply supervised, densely connected attention U-Net for the COVID-19 lung lesion segmentation. These are used to keep both the high and low-level features efficiently. Our model is trained on an adaptive optimizer. The hyper-parameters are optimized in hybrid training procedures. The workflow diagram is shown in Fig. 1. We take chest CT images as input. In the data preprocessing, we resize, normalize and augment input images. We share parameters between encoder and decoder tasks. The final output is the COVID-19 lesion segmentation. The proposed ADU-Net model is shown in Fig. 2.

3.1.1. Data preprocessing

All the COVID-19 chest CT images [32] are resized to 224×224. Before training, we normalize image pixel values to [0, 1]. The data augmentation approach creates new data that slightly differ from existing data. We have done the data augmentation using parameterized transformations [44] like rotation, scaling, and vertical–horizontal flipping on 20% of the original chest CT images. We have generated chest CT images and corresponding lesion masks by rotating and scaling with the same degrees and same scale factors. It helps to train the model with more and different data examples and prevents the model from over-fitting.

Table 1

Summary of related work for COVID-19 lesion segmentation on chest CT images.

Authors	Methods	Materials source	Performances
A. Amyar et al. [17]	Multi-task learning: COVID-19 or Non-COVID image reconstruction and classification in the encoder and followed by COVID-19 lesion segmentation in the decoder	SIRM and MedSeg COVID-19 chest CT segmentation dataset [32]	Dice coeff: 0.785
N. Paluru et al. [18]	Anam-Net: Anamorphic depth based lightweight, fully convolutional network	[32]	Dice coeff: 0.972 (Normal) 0.755 (COVID-19)
D. Muller et al. [19]	3D U-Net [33] and Residual U-Net [34]	Zenodo COVID-19 CT lung and infection segmentation dataset [35]	Dice coeff: 0.761
F. Shan et al. [20]	VB-Net	Chinese hospitals	Dice coeff: 0.916 ± 0.10
O. Gozes et al. [21]	U-Net with VGG-16 encoder	Chinese hospitals	AUC score: 0.948
J. Ma et al. [22]	nnU-Net (no-new-U-Net) [36]: automatically set the size of the patch, batch, kernel, stride, and pooling from a given dataset	[35]	Dice coeff: 0.673 ± 0.223
X. Chen et al. [34]	Aggregated Residual Network (ResNeXt) and Locality Sensitive Hashing (LSH) soft attention	[32]	Dice coeff: 0.83 (no data augmentation), 0.94 (with data augmentation)
Q. Yan et al. [37]	COVID-SegNet: U-Net with Feature Variation and Spatial Pyramid Pooling	Chinese hospitals	Dice coeff: 0.726
D. P. Fan et al. [38]	Inf-Net and Semi-Inf-Net: Res2Net [39] with parallel partial decoder and edge attention followed by reverse attention	[32]	Dice coeff (multi-class): 0.682 (Inf-Net) 0.739 (Semi-Inf)
Y. Qiu et al. [40]	MiniSeg: Attentive Hierarchical Spatial Pyramid (AHSP) in the encoder, Feature Fusion (FF) in the decoder, and depthwise separable convolution (DSConv) are used	[32]	Dice coeff: 0.759
V. K. Singh et al. [41]	LungINFseg: Receptive-Field-Aware (RFA) and attention mechanism	[35]	Dice coeff: 0.8034
P. Zhang et al. [29]	CoSinGAN: conditional generative adversarial networks model	[35]	Dice coeff: 0.615 ± 0.202
Y. Wang et al. [31]	The model is evaluated using 3D U-Net and four hybrid transfer learning methods. Dynamic feature selection is made using attention based fusion method	[35]	Dice coeff: 0.704
N. Saeedizadeh et al. [42]	TV-Unet: Regularization term 2D-anisotropic Total-Variation is used in the loss function	[32]	Dice coeff: 0.86
J. Zhang et al. [30]	Dense Generative Adversarial Network and U-Net-based multi-layer attention mechanism	[32]	Dice coeff: 0.68
MA. Basset [43]	Few-Shot Segmentation (FSS) using pre-trained model ResNet34	[32]	Dice Coeff: 0.798 (semi-supervised) 0.679 (supervised)

3.1.2. Encoder model

Our proposed modified U-Net [45,46] based DNN (as shown in Fig. 2) is built with down-sampling at the encoder stage (left side) and up-sampling at the decoder stage (right side). The bottleneck stage resides on the bottom side. The encoding stage consists of four convolution blocks. In every step of the encoding stage, 5×5 kernel convolution layers are frequently applied instead of 3×3 kernel, and 2×2 max-pooling downsampling layers are used with appropriate strides. Small kernels extract valuable features from the smaller receptive field of very few pixels at once. It slowly decreases the image dimensions to build the network deeper with more memory and share the weights better than the large kernel. The model performance is observed with convolution kernel size 3×3, 5×5, and 7×7, however, the best performance is achieved by convolution kernel size 5×5 for the COVID-19 chest CT [32]. It extracts in-depth features using convolution layers, as shown by the blue arrow from the specified image. It reduces the resolution by 2 using max-pooling layers, shown by the brown arrow. The convolution blocks will bring out the in-depth activated feature maps of the given image in the neural network. The feature map (no. of filter) is increased from 64 to 1024 feature maps from the first convolution layer to the last convolution layer of the encoder stage. The max-pooling layers will halve down the dimension of the image from 224×224 to 14×14 from the first max-pooling layer to the last max-pooling layer of this stage. The batch normalization techniques have an impressive effect on optimization performance for neural networks. It can be used to normalize the inputs to a layer.

We have applied batch normalization after each convolution (Conv 5×5) and before the activation function (ReLU). As a result, we have achieved a training speedup from higher learning rates and faster network convergence.

3.1.3. DenseNet

It is observed that the model performance decreases with the increase in convolutional layers. It does not eradicate the gradient loss or vanishing gradient problem, which means weights are insignificantly incremented to the initial convolution layers and degrade in backpropagation in the training phase. To alleviate the problem, the DenseNet [47,48] architecture is introduced, consisting of two stages, dense block operations followed by transition down block operations. The DenseNet architecture has two dense blocks and one transition-down block. In the dense block, features from all preceding layers are received, combined and provide better gradient propagation. As shown in Fig. 3, the dense block operations are batch normalization, ReLU activation, Conv 5×5, and channel-wise concatenation instead of element-wise addition. As the features come from all previous layers to the subsequent layers, information is easily kept between layers. In the encoder limb, the input of a dense block is concatenated to ensure that each block receives the entirely represented input, and so it recognizes the finer features. As a consequence of the modification direct connection between subsequent layers is established. In the transition down the block, batch normalization, ReLU activation, Conv 1×1, and max pool 2×2 operations are executed, as shown in Fig. 4.
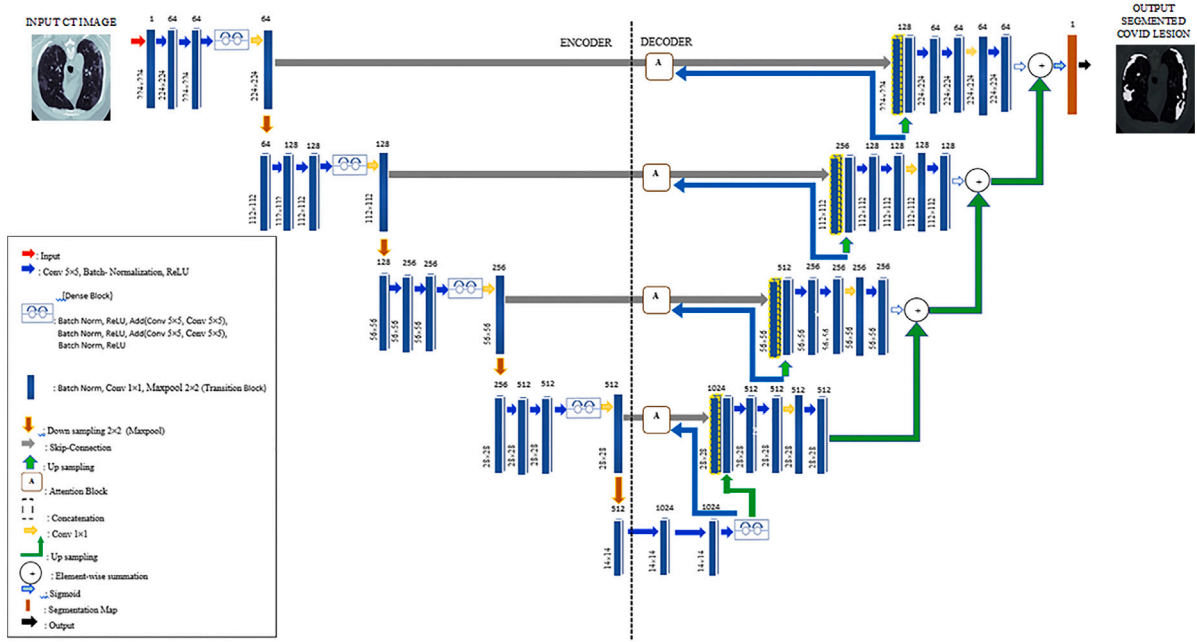
Fig. 2. Proposed ADU-Net architecture. Our modified dense U-Net architecture is built with down-sampling at the encoder stage (left side) and up-sampling at the decoder stage (right side) with the attention gate and deep supervision for COVID-19 lesion segmentation. The DenseNet is applied at each encoder level to overcome the vanishing gradient problem. The attention mechanism is incorporated between the encoder, skip-connection, and decoder. It is used to keep the high and low-level feature maps. Also, the deep supervision mechanism is included to combine secondary segmentation maps from various resolution feature maps from each decoder level in the proposed DNN.
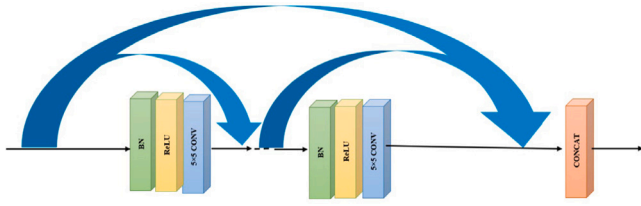

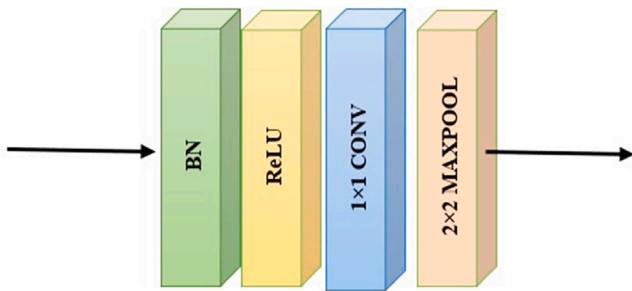**Fig. 3.** Dense block operations of DenseNet.

Fig. 4. Transition down block operations of DenseNet.

The convolution layer 1×1 operation reduces the channel count to half. The max pool layer 2×2 responsible for downsampling the image features. It is expected to have the proposed network would have better intermediate features. The DenseNet block has a specific advantage of a strong gradient flow and high efficiency with a lower number of parameters since each layer receives the feature maps from all the preceding layers. Besides increasing efficiency, the collective knowledge and dense connectivity allow DenseNet to obtain more diversified

and rich feature patterns in the feature images, which makes DenseNet blocks a perfect fit for image segmentation tasks. The resultant of the dense block is defined in Eq. (1):

$$x_l = H_l(\text{Concat}[x_0, x_1, \ldots, x_{l-1}]) \tag{1}$$

where $[\cdot]$ denotes channel-wise concatenation, while x_l denotes concatenated feature map generated from each previous layer $0, 1, \ldots, l-1$ at each transition label and $H_l(\cdot)$ represents the dense mapping function which is the combination of convolution, batch-normalization and activation. The convolutional procedure for lth layer is formulated in Eq. (2):

$$x_l = \sum_{i=0}^{n-1} x_{l-1} * w_l + b_l \tag{2}$$

Where x_l is the output feature map, x_{l-1} is the input data from $(l-1)^{th}$ layer, w_l and b_l represents weight and bias respectively, n is the number of input features and i ranges between 0 and $n-1$. The batch normalization operation for lth layer is defined in Eq. (3):

$$x_l = \gamma \cdot \frac{x_{l-1} - \mu_{x_l}}{\sqrt{\sigma_{x_l}^2}} + \beta \tag{3}$$

where β, γ are parameters and μ_{x_l}, σ_{x_l} are mean and standard deviation.

To establish the non-linearity of data across the layers, batch normalization and ReLU is applied on x_l in a cascaded manner which is formulated in Eq. (4):

$$x_l = \text{BatchNorm}(\text{ReLU}(w_l * x_{l-1})) \tag{4}$$

3.1.4. Decoder model

The bottom side of the neural segmentation network shows the bottleneck stage (as shown in the bottom side Fig. 2). This stage works between the encoding stage and the decoding stage. It was built with two convolution layers without any max-pooling layer. The bottleneck stage is needed because it gives a much more in-depth network. Bottleneck architecture will lead to a high-resolution image as input, requiring more complex feature maps. Therefore, the image size is upsized from 512×512 to 1024×1024, which will propagate to the
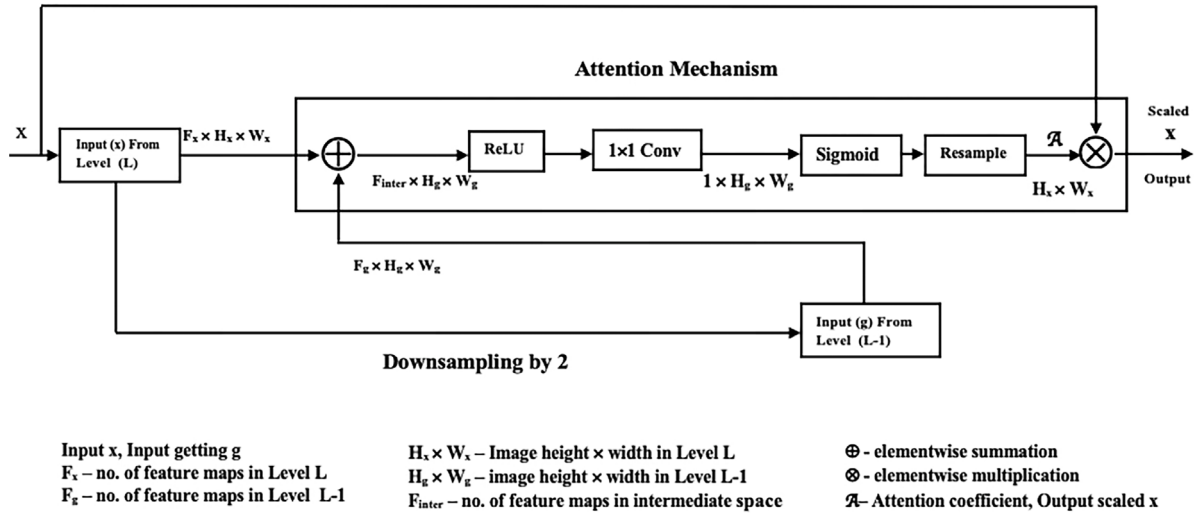
Fig. 5. Attention Mechanism. The input of the attention block is the element-wise addition of higher-level (L) feature-map x from the encoder and lower-level (L-1) feature-map g from the decoder. x and g are linearly mapped to the same dimensional space. Then there is the non-linear activation function ReLU, convolution and non-linear activation function Sigmoid. The output of the attention block is the multiplication of the Attention coefficient $\mathcal{A}$ and input x.
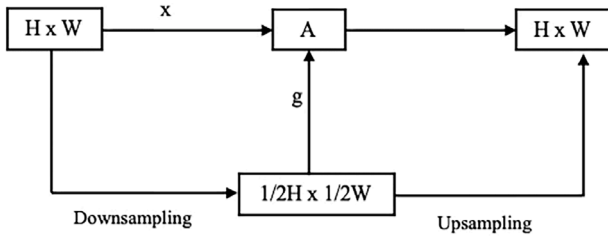

Fig. 6. Input and output dimensions of Attention block. The input of the attention block is higher-level input feature-map x from the encoder and lower-level gated feature-map g from the decoder.

decoding stage. In contrast, the decoding stage (as shown in the right side of Fig. 2) consists of four deconvolution layers with transposed convolution operations. It up-samples the image feature maps by two. Also, it halves the number of feature channels. The exact number of filters is used at every level in the convolution and deconvolution layers to maintain symmetry. At the decoder, the feature maps vary from 1024 to 64, shown by the green arrow in the neural network. In every step of the decoding stage, there is frequent application of 2×2 upsampling layers, each followed by 5×5 convolution layers instead of 3×3, as mentioned earlier. The deconvolution layer up samples the feature map that increases the image's dimension from 14 to 224.

In this semantic segmentation, the image size is gradually reduced while the image depth increases. It starts from $224 \times 224 \times 1$ and ends at $14 \times 14 \times 512$ in down-sampling. So, spatial information is lost. To restore lost information, the feature maps at a particular level of the encoder get copied to the corresponding level of the decoder. The principal idea of the concatenation process is to merge the feature maps from corresponding convolution layers to get an appropriate and precise segmentation network output at deconvolution layers. Before copying, it needs to be cropped to resolve the two feature maps are the same. Skip connections are used to skip some layers in the neural network. It explicitly copies feature maps or feeds one layer's output to another, skipping a few layers in between. Some information is captured in the initial layers and is required for reconstruction during the later layers of upsampling using the skip architecture. Skip-connection prevents information loss when passing through multiple layers. It also helps traverse information faster in deep neural networks. Skip-connection is liable to use redundant information from the encoder

low-level feature map. It leads to poor efficacy when appending with the decoder high-level feature map. That can produce dense predictions at multi-levels. The proposed neural network has some superiority compared to other studies in the literature.

3.1.5. Attention gate

To overcome deficiencies and weaknesses, introduce an attention mechanism into the skip-connection of the encoder feature and corresponding decoder feature. It keeps the essential feature maps or regions. Before concatenation, the attention mechanism is allowed in the segmentation network. The attention mechanism in [49–51] combines linear transformation and non-linear activation function. The residual attention [52] extracts local and non-local features. The attention mechanism keeps the high-level major and low-level minor feature maps. The block diagrams of the attention mechanism are shown in Figs. 5 and 6.

$$A = \sigma_2 \left[w \left\{ w_i \left(\sigma_1 \left(w_x x + w_g g + b_{x,g} \right) \right) + b_i \right\} + b \right] \tag{5}$$

$$\sigma_1 = ReLU(x) = max(0, x) \tag{6}$$

$$\sigma_2 = Sigmoid(x) = \frac{1}{1 + exp(-x)} \tag{7}$$

Attention Gate (AG) evaluates the attention coefficient $\mathcal{A}$ using Eq. (5), 6, and 7. x and g are the feature mapping of the AG inputs from the encoder and decoder, respectively. σ_1 and σ_2 are the ReLU and sigmoid functions, respectively. The σ_1 ranges between {0,1}. Let w represent the weights of the network. $w_x x$ and $w_g g$ are the linear transformation. The given bias is the $b_{x,g}$, b_i, and b. Finally, the output of AGs is 'Scaled x', which is the multiplication of attention coefficient $\mathcal{A}$ and input x is given in Eq. (8).

$$Scaled\ x = \mathcal{A} \cdot x \tag{8}$$

The feature maps are transposed by 1×1 convolution to generate desired segmentation output. However, the final output loss is not effectively gradient back-propagated.

3.1.6. Deep supervision

To minimize this issue, the deep supervision [53,54] is implemented to combine secondary supervision maps by considering various resolution feature maps from the previous levels to the final level in the segmentation network. Then, it integrates predictions from various
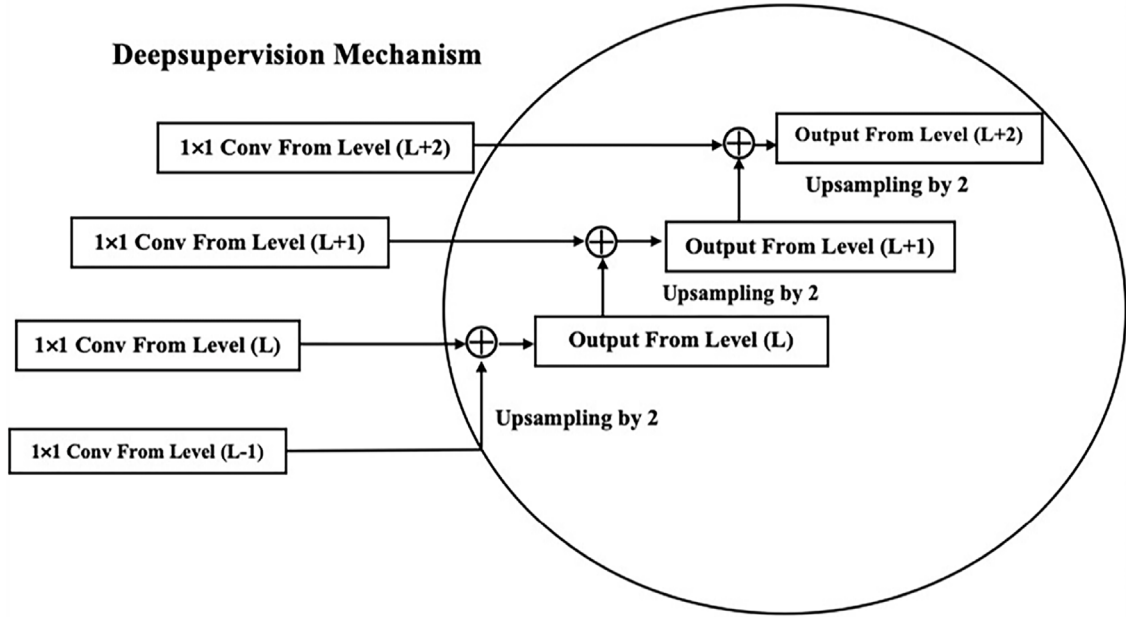
Fig. 7. Deep Supervision Mechanism. The secondary segmentation maps are created from feature maps transposed by 1×1 convolutions at each decoder level. Also, the lower level (L-1) low-resolution segmentation map is up-sampled to get the same size as the immediately higher level (L) segmentation map. Then, these two segmentation maps are added element-wise, then again up-sampled and added to the next higher level (L+1) segmentation map and so on till the final level (L+2).

resolution feature maps from earlier levels and reproduces the final in-depth supervision combined segmentation output, as shown in Fig. 7. The losses associated with the individual segmentation prediction are weighted and appended with the last loss. Therefore, more efficient gradient back-propagation is achieved. Also, further refinement of the final segmentation prediction at the final level is optional, as multi-scale context information can be achieved due to the presence of skip connections. As a result, the deep supervision mechanism produces better segmentation results and faster convergence than only with skip connections. The evaluation of the deep supervision mechanism is based on Eqs. (9) and (10). Let w represents weights of the main network and w_L be the weights of levels where $L \in \{\ell-1, \ell, \ell+1, \ell+2\}$. For, the predicted output y corresponds to input feature x, the dice loss function is given in Eq. (9).

The dice loss function:

$$DiceLoss\big(y, x\big) = 1 - \frac{2 \sum_L xy}{\sum_L x + \sum_L y} \tag{9}$$

When $\mathcal{R}$ is the $\mathcal{L}_2$ regularization term with hyper-parameter $\mathcal{H}$ then total dice loss function:

$$\begin{aligned} DiceLoss_L\big(y, x; w, w_{l-1}, w_l, w_{l+1}, w_{l+2}\big) = \\ w\Bigg(\sum_{L \in \{\ell-1, \ell, \ell+1, \ell+2\}} w_L DiceLoss\big(y, x; w, w_L\big)\Bigg) \\ + \mathcal{H}\Bigg(\mathcal{R}(w) + \sum_{L \in \{\ell-1, \ell, \ell+1, \ell+2\}} \mathcal{R}\big(w_L\big)\Bigg) \end{aligned} \tag{10}$$

3.1.7. Training process

The aim is to learn a model function $f(x, y; \theta)$ in a supervised manner through a training process which parameterizes with θ and maps input features x to target features y, and minimizes the objective or loss function $L(x, y; \theta)$ which returns-optimized θ, as defined in Eq. (11). It predicts the values of the target features y based on depictions of the input features x of test data.

$$\theta_{opt} = \arg \min_{\theta}(L(x, y; \theta)) \tag{11}$$

The proposed COVID-19 lesion segmentation neural network is trained and validated on chest CT images of size 224×224. The input training
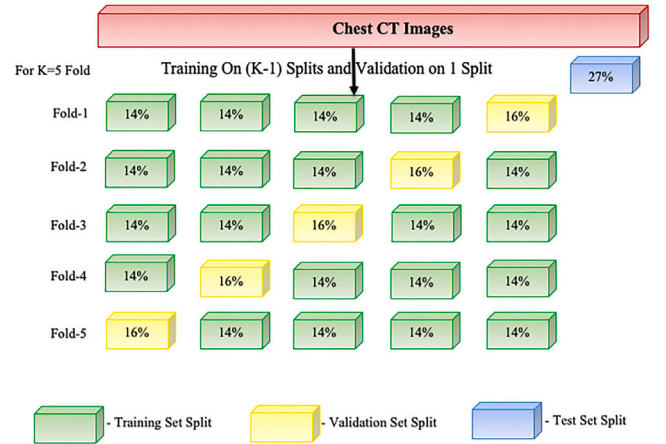

Fig. 8. Dataset split in 5-fold cross-validation.

image samples are divided into four mini-batches. The proposed model is trained on 5-fold cross-validation for better assessment. There are 100 epochs in each fold. One epoch is passed forward and backwards through the entire training dataset only once in the neural networks. In each epoch, there are n-number of iterations, calculated as the division between the number of training image samples and batch size. The proposed model is trained to minimize the cost functions of dice loss. An Adam optimizer reduces the cost function, and an initial learning rate of 0.0001 is used to update the model's weights. The learning rate reduction (decay) manner minimizes by a value of 0.000001. If the validation loss is stopped minimizing after ten consecutive epochs (patience), then an early stop is occurred and restores the best weights. The early stopping condition and dropout of 0.5 are used to avoid model overfitting. We have trained and validated our proposed deep neural network model with an evenly balanced weight so that the positive and negative training samples would contribute equally to the cost function. The evenly proportional weights help to arrive at faster convergence. The activation function ReLU and batch-normalized are used in neural networks. We have experimented with parameters

Table 2
Experimental setup for proposed models.

Folds	Epochs	Early Stopping Patience	Early Stopping Monitor	Image Size	Batch	Optimizer	Learning Rate	Decay	Loss Function	Activation Function
5	100	10	val loss	224×224	4	adam	1e−4	1e−6	dice loss	ReLU,Sigmoid

in kernel sizes, pooling operations, and strides, followed by hyper-parameter settings—the best set of parameter and hyper-parameter values chosen by parameter optimization techniques that provide minimum validation loss. We have designed a typical experimental setup given in Table 2.

3.1.8. Proposed algorithm

```
input: D : Dataset containing COVID-19 CT images and corresponding
            infection mask images
       d₁ : Training set from D
       d₂ : Validation set from D
       d₃ : Testing set from D
       μ : Learning rate
       f : Number of folds
       ϵ : Number of epochs
       s : Callbacks(Early Stop)
       ξ : Number of iterations
       β : Batch size i.e. number of images trained in one iteration
output: ω : DNN weights in training
        m : Trained proposed models
        roi : Corresponding predicted region of interest for input image x in
              testing
preprocessing: Resize each CT image to dimension 224×224
               Generate COVID-19 CT images based on data augmentation
               Normalize each CT image between [0,1]
begin: 1. Evaluate initial DNN class weights
       2. Selecting d₁, d₂, and d₃ from split data
       3: Proposed encoder-decoder model for COVID-19
          lesion segmentation
          Set proposed model layers with encoder, densenet, attention gate, decoder,
          and deep supervision
       4: Initialize hyper-parameters: μ, f, ϵ, s, ξ and β
       5: Train the proposed model
       for f = 1 to 5 do   // for training and validation
           for ϵ = 1 to s do   // if validation loss is not improving for 10 epochs
               6. Randomly select β from d₁   // select d1 and d2 for training and validation
               for ξ = 1 to d₁ / β do        // select d1 and d2 for training and validation
                   7. Forward propagation and compute dice loss with μ, dice coeff,
                      jaccard coeff, accuracy, specificity, sensitivity, precision, f1 score
                   8. Back propagation, update ω with Adam optimizer and
                      train m
               end
           end
       end
end

begin:
       for each x in d₃ do   // for testing
           predict roi = m(x)
           compute dice loss, dice coeff, and jaccard coeff
       end
end
```

3.1.9. 5-Fold cross validation

5-fold cross-validation [55] is used for better evaluation. For 5-fold cross-validation, the dataset is split into five folds. We pick four folds for training and the rest one fold for validation at every time, as shown in Fig. 8. 5-fold cross-validation takes more training data. Therefore, the learning efficiency of the neural network is faster, and the result is relatively better.

4. Materials

The main concern regarding the involvement of machine learning and the deep learning approach concentrates on relevant and appropriate data. We have used publicly available COVID-19 chest CT segmentation dataset [32] found at the Italian Society of Medical and Interventional Radiology (SIRM) and MedSeg database. The dataset consists of 100 axial chest CT images from more than 40 patients with COVID-19 and its corresponding infection masks. The segmented masks or the ground truths were already done by expert radiologists using MedSeg tool [32], and they labelled lesions in the chest CT images. Therefore, it has been considered that all the lesions are labelled as COVID-19 lesions. All image files are in greyscale.
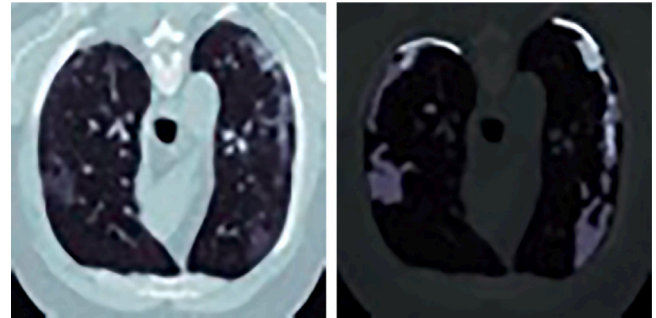

Fig. 9. COVID-19 Chest CT and corresponding ground truth with pleural effusion(fluid in pleura), ground glass opacities(hazy opacity), and consolidation(fluid in alveoli).

Table 3
Statistics of data split [32] for COVID-19 lesion segmentation.

Dataset	Training set (57%)	Validation set (16%)	Test set (27%)	Total (100%)
COVID-19 CT	57	16	27	100
Augmented COVID-19 CT	114	32	54	200

The dataset is split into three subsets. The percentage of the dataset used in training, validation, and testing are 57%, 16%, and 27%, respectively. The COVID-19 Chest CT and its relevant ground truth are shown in Fig. 9. In addition, the statistical distribution of images is given in Table 3 for COVID-19 chest CT [32].

5. Evaluation metrics

The widely used evaluation metrics [56] are Accuracy, Specificity, Sensitivity or Recall, Precision, F1-Score, Dice Similarity Coefficient (DSC), Dice Loss, and Jaccard Coefficient or IoU (Intersection over Union) are defined in Equations 13, 14, 15, 16, 17, 18, 19, and 20 respectively. This metrics value ranges between 0 and 1. Zero denotes the lowest similarity, and one denotes the highest similarity. Accuracy, Specificity, Sensitivity/Recall, Precision, and F1-Score are calculated from the Confusion Matrix, defined in Eq. (12).

$$ConfusionMatrix = Actual \begin{pmatrix} Predicted & \\ TP & FN \\ FP & TN \end{pmatrix} \tag{12}$$

Where TP: True Positive, TN: True Negative, FP: False Positive, FN: False Negative

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{13}$$

$$Specificity = \frac{TN}{TN + FP} \tag{14}$$

$$Sensitivity = Recall = \frac{TP}{TP + FN} \tag{15}$$

$$Precision = \frac{TP}{TP + FP} \tag{16}$$

$$F1 Score = \frac{2(Recall * Precision)}{Recall + Precision} \tag{17}$$

Dice Coefficient and Jaccard Coefficient calculated the similarity among the ground truth infectious regions (G) and predicted infectious

Table 4
Validation performance of COVID-19 lesion segmentation on chest CT dataset [32].

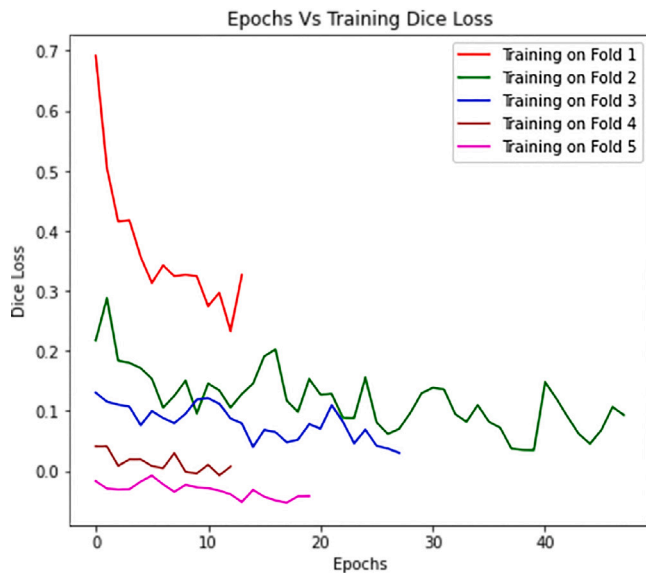| Fold	Performance							
	Dice Loss	Dice Coeff	Jaccard	Accuracy	Specificity	Sensitivity or Recall	Precision	F1 Score
Fold-1	0.99	0.06	0.03	0.90	0.99	0.02	0.37	0.03
Fold-2	0.42	0.52	0.35	0.92	0.97	0.44	0.56	0.49
Fold-3	0.07	0.87	0.76	0.92	0.98	0.74	0.87	0.80
Fold-4	0.03	0.88	0.78	0.93	0.98	0.77	0.91	0.83
Fold-5	0.11	0.83	0.70	0.95	0.98	0.79	0.89	0.83

Fig. 10. Graph plot showing the training dice loss results from the 5Fold-CV on COVID-19 lesion segmentation.
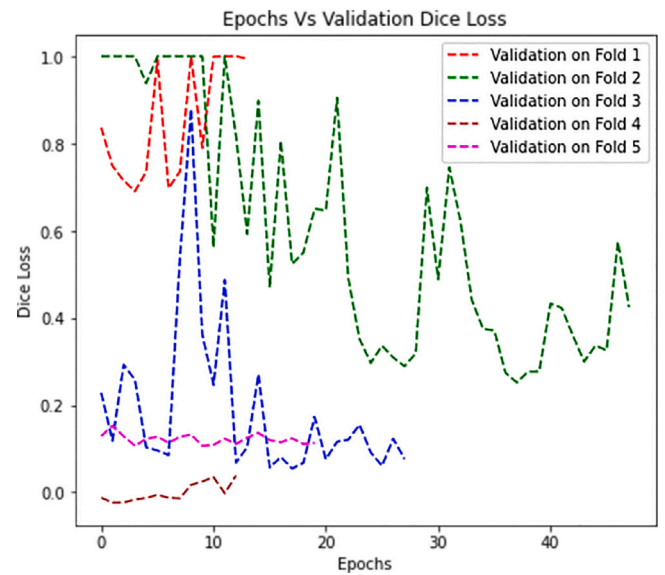

Fig. 11. Graph plot showing the validation dice loss results from the 5Fold-CV on COVID-19 lesion segmentation.

regions (P).

$$DiceCoef = \frac{2TP}{2TP + FP + FN} = \frac{2|P \cap G|}{|P| + |G|} \tag{18}$$

$$DiceLoss = 1 - DiceCoef \tag{19}$$

$$JaccardCoef = \frac{TP}{TP + FP + FN} = \frac{|P \cap G|}{|P \cup G|} \tag{20}$$

6. Results and discussion

The proposed models are implemented using Tensor-flow and Keras framework in Python. The experiments use 32 GB of RAM, an Intel Core i7 processor of 2.3 GHz, and 8 GB of GPU (NVIDIA GE-FORCE RTX).

6.1. Results of proposed deep neural network (ADU-Net)

We have trained and validated our proposed COVID-19 lesion segmentation neural network. We have repeated 100 epochs for five folds with batch size 4. We have observed the COVID-19 lesion segmentation encoder–decoder model performance in each fold. We have achieved the dice coefficient, Jaccard coefficient, accuracy, specificity, sensitivity or recall, precision, and f1-score are 0.83, 0.70, 0.95, 0.98, 0.79, 0.89, and 0.83, respectively, from the 5-folds cross-validation (5Fold-CV) which is demonstrated in Table 4 and shown violin plot in Fig. 12 and graph plots in Figs. 10 and 11. The average results are calculated on 160 data (= 5 folds ×32 validation data per fold). In addition, we have observed and compared the test performance of the proposed encoder–decoder ADU-Net model with state-of-the-art models illustrated in Table 7. The proposed COVID-19 lesion segmentation model has achieved a 0.82 test dice coefficient and shown better performance than state-of-the-art models.
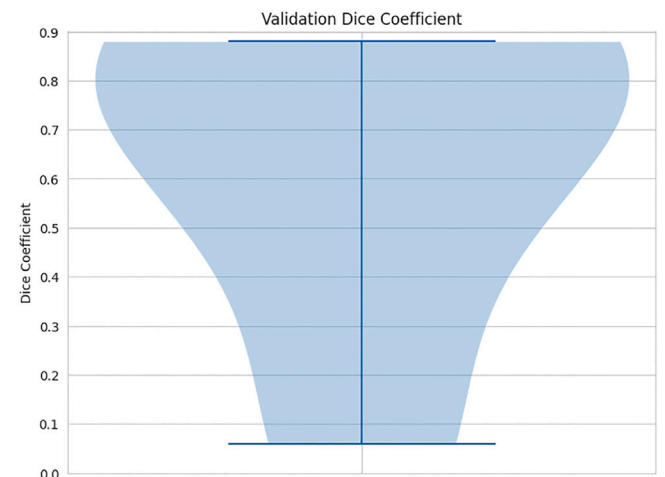

Fig. 12. Violin plot showing the validation dice coeff results from the 5Fold-CV on COVID-19 lesion segmentation.

We have evaluated the role of attention gate, densenet, and deep supervised module on the backbone of UNet in our work by performing an ablation study as given in Table 5.

We have tested the performance of the model by using 3×3, 5×5, and 7×7 kernel size. We have performed an ablation study as shown in Table 6. We have got the best performance using 5×5 kernel size.

The dice coefficient of ResNeXt [34] is 0.83. The ResNeXt [34] is also trained with 500 CT augmented from 100 COVID-19 CT [32] images. It achieved a high dice coefficient of 0.94 with data augmentation. The number of training samples affects the performance of a

Table 5
Ablation study on test results of the model.

Models Used in CT	Dice Coeff	Jaccard Coeff
UNet	0.71	0.55
UNet+AttentionBlock	0.77	0.63
UNet+DenseBlock	0.79	0.66
UNet+DenseBlock+AttentionBlock	0.81	0.68
UNet+DenseBlock+AttentionBlock +DeepSupervisedBlock (Proposed ADUNet)	0.82	0.70
Proposed ADUNet (with data augmentation)	0.86	0.75

Table 6
Ablation study on test results of the model.

Models Used in CT	Dice Coeff	Jaccard Coeff
ADUNet (3×3 kernel size)	0.81	0.68
ADUNet (5×5 kernel size)	0.82	0.70
ADUNet (7×7 kernel size)	0.75	0.61

Table 7
Comparison of test results between existing models and proposed model on COVID-19 chest CT segmentation dataset [32].

Models Used in CT	Dice Coeff	Jaccard Coeff	Dice Loss
MLAttentionUNet+DGAN [30]	0.68	0.53	0.32
MiniSeg [40]	0.75	0.61	0.25
Anam-Net [18]	0.76	0.62	0.24
Multi Task Learning [17]	0.78	0.64	0.22
FSS-2019-nCov [43]	0.79	0.66	0.21
ResNeXt [34]	0.83	0.70	0.17
TV-Unet [42]	0.86	0.73	0.14
Proposed ADUNet Model	**0.82**	0.70	0.18
Proposed ADUNet Model (with data augmentation)	**0.86**	0.75	0.14

deep neural network. In contrast, the test dice coefficient of MultilayerAttentionUNet DenseGAN [30], MiniSeg [40], Anam-Net [18], Multi Task Learning [17], FSS-2019-nCov [43], and TV-Unet [42] models are 0.68, 0.75, 0.76, 0.78, 0.79, and 0.86 respectively as given in Table 7. We have augmented data using parameterized transformations and generated 200 CT augmented from 100 COVID-19 CT [32] images as shown in Table 3. We observed that our proposed ADUNet model achieved a test dice coefficient of 0.86 with data augmentation. The obtained test dice coefficient by the proposed method is close to the TV-Unet [42]. However, the Jaccard coefficient of the proposed method and the TV-Unet are 0.75 and 0.73, respectively. Therefore, the performance of the proposed COVID-19 lesion segmentation model is comparable and shows better performance than state-of-the-art models as shown in Table 7.

6.1.1. Reasons for superiority of proposed method

The reasons for the superiority of the proposed method are:

The convolution kernel size 5×5 extracts useful features from a small field view at once and reduces image sizes to create a deep network. The best performance is achieved by kernel 5×5 for the COVID-19 chest CT.

The batch normalization technique is applied after each convolution operation and before the ReLU activation to normalize the inputs to a layer, achieve a training speedup from higher learning rates, and attain optimized performance for the neural network.

The encoding stage of U-Net is a CNN which extracts patterns by increasing features resolution from 64 to 1024 and reducing image size from 224×224 to 14×14. As a result, the spatial information is lost due to downsampling at encoding. To restore lost information, the feature maps at a particular level of the encoder get concatenated to the corresponding level of the decoder. However, the image size and depth

start from $14 \times 14 \times 1024$ and end at $224 \times 224 \times 1$ in the decoding stage to generate the image output from the input image.

DenseNet removes the vanishing gradient problem, and weights are updated adequately to the initial convolution layers, improving backpropagation in the training phase.

Skip-connection is liable to use redundant information from the encoder, leading to poor efficacy when appending with the decoder. On the other hand, it can produce dense predictions at multi-levels. Therefore, the attention gate is applied on skip-connection and before concatenation to keep the high-level and low-level essential features using linear transformation and non-linear activation functions.

However, the final output loss needs to be effectively gradient back-propagated. Finally, deep supervision integrates predictions from various resolution feature maps from earlier levels and reproduces the final in-depth supervised segmentation output.

Therefore, the proposed attention dense U-Net based deep supervised DNN (ADU-Net) is a new, efficient, and hybrid model for automated lesion segmentation of COVID-19 from chest CT images. As a result, the proposed model achieves a test dice coefficient of 82% and 86% with data augmentation, which outperforms the state-of-the-art models.

### 6.1.2. Analysis of layer-wise visualizations on ADU-Net

To visualize the images and representations of the feature maps which are being passed on from each layer, a layer-wise output for each layer has been depicted for the used models in Fig. 13(a-d). The input data of one COVID-19 chest CT image is fed into the layers or functional blocks in the encoder and decoder limbs of the proposed DNN. There are five levels and 228 layers in the proposed DNN model. The decoded embeddings from both the sub-nets are sent to the final covering layers of the proposed network for the actual output.

The primary difference present in the Attention U- Net and Dense U-Net networks are between the feature targets accumulated by the networks. The Dense U-Net network focus on the more minor features and the Attention U-Net network focuses on the more significant features. This inference is constructed based on the fact that the concatenated structure of the Dense U-net does not allow it to lose focus of the minute details of the feature maps, while for Attention U-Net dives deep to a point where only the sharp edges are persistent, to bring out the best features, thus losing the minor details in the process. Therefore, the proposed model combines attention U-Net and dense U-Net. Fig. 13(a-d) shows that the feature maps have a consistent amount of fine details throughout all the layer outputs. In contrast, in the outputs of levels 1 and 2 for the encoder limb of the model, we can only see the sharp edges, and again the feature maps are reconstructed as we move up the decoder limb. A quick comparison between the three output layers shows that the features are best highlighted within the embedding layers of the final proposed network. The features are distinctly separated from one another and allow the identification of individual components in the image. The layer-wise visualization provides concrete qualitative validation that the results brought about by the conjunction of the two separate networks outperform the individual results.

6.1.3. Comparison between ground truth and predicted output

The output prediction of the proposed COVID-19 lesion segmentation model is shown in Fig. 14(a-c). The proposed model generates the predicted value, and the true value is the ground truth.
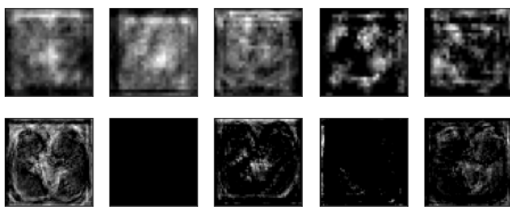
The error comparison between the ground truth and the predicted output of the proposed method is visualized in Fig. 15 for true positive and false positive lesion regions. We have observed that the proposed method has shown comparable results with high accuracy.
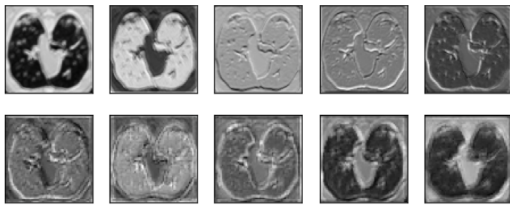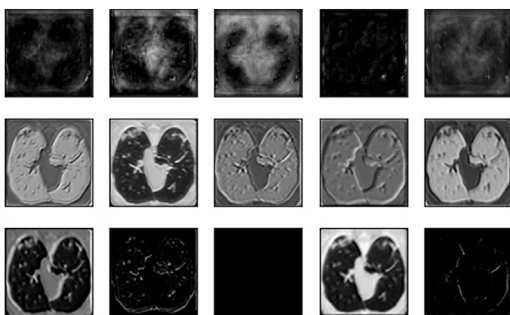
(a) above: encoder 1st level, below: encoder 2nd level
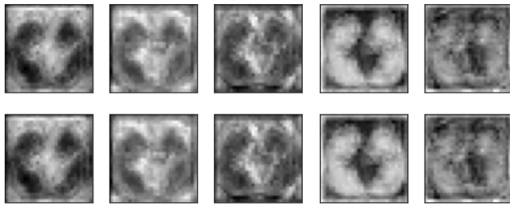
(b) above: encoder 3rd level, below: encoder 4th level

(c) above: decoder 1st level, below: decoder 2nd level

(d) above: decoder 3rd level, below2: decoder 4th level

Fig. 13(a-d). Intermediate outputs of proposed ADU-Net model for lesion segmentation.

Chest CT   Ground Truth   Predicted

Fig. 14(a-c). (a) Column1: Chest CT, (b) Column2: Ground Truth, (c) Column3: Predicted COVID-19 lesion segmentation on proposed ADU-Net encoder-decoder model.

Chest CT   Ground Truth   Predicted

Fig. 15. The true positive lesion regions are shown in the green border, whereas red represents false positive lesion regions at the predicted output of the proposed ADU-Net.

7. Conclusion

In this research work, we have proposed a novel attention-based dense U-Net architecture that is trained under deep supervision and improves the quality of segmentation of COVID-19 lesions from the lung CT images of COVID-19-infected patients. The U-Net is utilized as the backbone of the segmentation network. The dense net effectively removes the vanishing g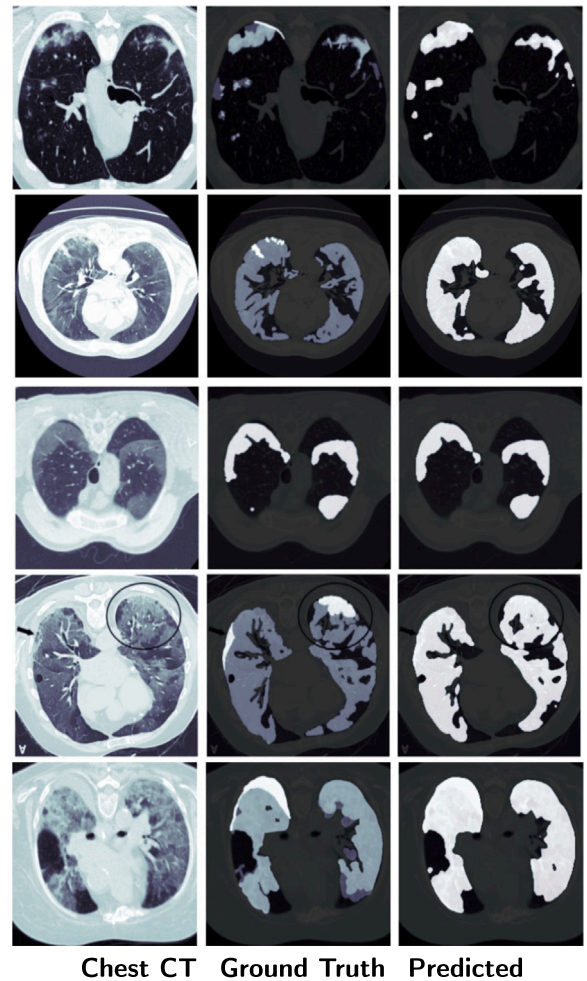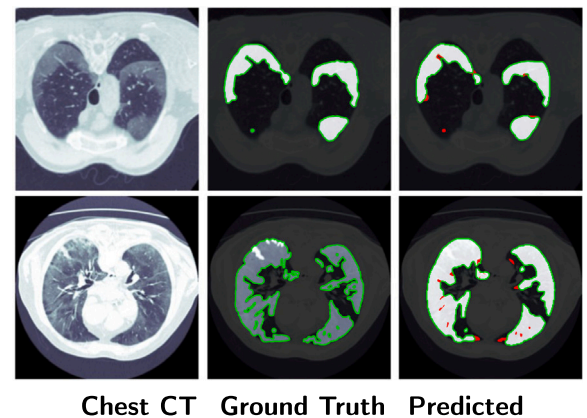radient problem, updates weights adequately, and improves back-propagation. The attention gate efficiently keeps attention to essential features using linear transformation and non-linear activation functions while suppressing irrelevant features. The deep supervision effectively integrates predictions from various resolution feature maps from earlier levels and reproduces the final in-depth supervised segmentation output. We evaluated the model by performing

an ablation study. The proposed model can better segment the COVID-19-affected areas in the lungs from the chest CT images. The publicly available COVID-19 chest CT segmentation dataset has trained the proposed model. The efficacy of the trained model has experimented with test images. The proposed DNN model outperforms the state-of-the-art models and has achieved the segmentation test dice coefficient of 0.86 and Jaccard coefficient of 0.75 with data augmentation. The proposed methodology may become a helpful tool for physicians for COVID-19 disease management.

CRediT authorship contribution statement

Sanjib Saha: Conceptualization, Visualization, Methodology, Implementation, Editing. **Subhadeep Dutta:** Conceptualization, Visualization. **Biswarup Goswami:** Relevant medical assessment and viability checking. **Debashis Nandi:** Conceptualization, Supervision, Editing.

Declaration of competing interest

No author associated with this paper has disclosed any potential or pertinent conflicts which may be perceived to have impending conflict with this work. For full disclosure statements refer to https://doi.org/10.1016/j.bspc.2023.104974.

Data availability

No data was used for the research described in the article

Acknowledgement

References

[1] J.T. Wu, K. Leung, G.M. Leung, Nowcasting and forecasting the potential domestic and international spread of the 2019-nCoV outbreak originating in Wuhan, China: a modelling study, Lancet 395 (10225) (2020) 689–697.

[2] A. Fontanet, B. Autran, B. Lina, M.P. Kieny, S.S.A. Karim, D. Sridhar, SARS-CoV-2 variants and ending the COVID-19 pandemic, Lancet 397 (10278) (2021) 952–954.

[3] J. Rodrigues, S. Hare, A. Edey, A. Devaraj, J. Jacob, A. Johnstone, R. McStay, A. Nair, G. Robinson, An update on COVID-19 for the radiologist-A British society of Thoracic Imaging statement, Clin. Radiol. 75 (5) (2020) 323–325.

[4] Y. Fang, H. Zhang, J. Xie, M. Lin, L. Ying, P. Pang, W. Ji, Sensitivity of chest CT for COVID-19: comparison to RT-PCR, Radiology 296 (2) (2020) E115–E117.

[5] T. Ai, Z. Yang, H. Hou, C. Zhan, C. Chen, W. Lv, Q. Tao, Z. Sun, L. Xia, Correlation of chest CT and RT-PCR testing for coronavirus disease 2019 (COVID-19) in China: a report of 1014 cases, Radiology 296 (2) (2020) E32–E40.

[6] R. Alizadehsani, Z. Alizadeh Sani, M. Behjati, Z. Roshanzamir, S. Hussain, N. Abedini, F. Hasanzadeh, A. Khosravi, A. Shoeibi, M. Roshanzamir, et al., Risk factors prediction, clinical outcomes, and mortality in COVID-19 patients, J. Med. Virol. 93 (4) (2021) 2307–2320.

[7] A. Shoeibi, M. Khodatars, R. Alizadehsani, N. Ghassemi, M. Jafari, P. Moridian, A. Khadem, D. Sadeghi, S. Hussain, A. Zare, et al., Automated detection and forecasting of COVID-19 using deep learning techniques: A review, 2020, arXiv preprint arXiv:2007.10785.

[8] F. Shi, J. Wang, J. Shi, Z. Wu, Q. Wang, Z. Tang, K. He, Y. Shi, D. Shen, Review of artificial intelligence techniques in imaging data acquisition, segmentation, and diagnosis for COVID-19, IEEE Rev. Biomed. Eng. 14 (2020) 4–15.

[9] S. Agarwala, M. Kale, D. Kumar, R. Swaroop, A. Kumar, A.K. Dhara, S.B. Thakur, A. Sadhu, D. Nandi, Deep learning for screening of interstitial lung disease patterns in high-resolution CT images, Clin. Radiol. 75 (6) (2020) 481–e1.

[10] C.-H. Liang, Y.-C. Liu, M.-T. Wu, F. Garcia-Castro, A. Alberich-Bayarri, F.-Z. Wu, Identifying pulmonary nodules or masses on chest radiography using deep learning: external validation and strategies to improve clinical practice, Clin. Radiol. 75 (1) (2020) 38–45.

[11] A. Kumar, N. Upadhyay, P. Ghosal, T. Chowdhury, D. Das, A. Mukherjee, D. Nandi, CSNet: A new DeepNet framework for ischemic stroke lesion segmentation, Comput. Methods Programs Biomed. 193 (2020) 105524.

[12] P. Ghosal, T. Chowdhury, A. Kumar, A.K. Bhadra, J. Chakraborty, D. Nandi, MhURI: A supervised segmentation approach to leverage salient brain tissues in magnetic resonance images, Comput. Methods Programs Biomed. 200 (2021) 105841.

[13] S. Minaee, R. Kafieh, M. Sonka, S. Yazdani, G.J. Soufi, Deep-COVID: Predicting COVID-19 from chest x-ray images using deep transfer learning, Med. Image Anal. 65 (2020) 101794.

[14] L. Wang, Z.Q. Lin, A. Wong, COVID-net: A tailored deep convolutional neural network design for detection of COVID-19 cases from chest x-ray images, Sci. Rep. 10 (1) (2020) 1–12.

[15] T. Ozturk, M. Talo, E.A. Yildirim, U.B. Baloglu, O. Yildirim, U.R. Acharya, Automated detection of COVID-19 cases using deep neural networks with X-ray images, Comput. Biol. Med. 121 (2020) 103792.

[16] A. Narin, C. Kaya, Z. Pamuk, Automatic detection of coronavirus disease (COVID-19) using x-ray images and deep convolutional neural networks, Pattern Anal. Appl. (2021) 1–14.

[17] A. Amyar, R. Modzelewski, H. Li, S. Ruan, Multi-task deep learning based CT imaging analysis for COVID-19 pneumonia: Classification and segmentation, Comput. Biol. Med. 126 (2020) 104037.

[18] N. Paluru, A. Dayal, H.B. Jenssen, T. Sakinis, L.R. Cenkeramaddi, J. Prakash, P.K. Yalavarthy, Anam-Net: Anamorphic depth embedding-based lightweight CNN for segmentation of anomalies in COVID-19 chest CT images, IEEE Trans. Neural Netw. Learn. Syst. 32 (3) (2021) 932–946.

[19] D. Müller, I.S. Rey, F. Kramer, Automated chest CT image segmentation of COVID-19 lung infection based on 3D U-Net, 2020, arXiv preprint arXiv:2007.04774.

[20] 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, 2020, arXiv preprint arXiv:2003.04655.

[21] O. Gozes, M. Frid-Adar, N. Sagie, H. Zhang, W. Ji, H. Greenspan, Coronavirus detection and analysis on chest CT with deep learning, 2020, arXiv preprint arXiv:2004.02640.

[22] J. Ma, Y. Wang, X. An, C. Ge, Z. Yu, J. Chen, Q. Zhu, G. Dong, J. He, Z. He, et al., Towards efficient COVID-19 CT annotation: A benchmark for lung and infection segmentation, 2020, arXiv e-prints. arXiv–2004.

[23] A.K. Mishra, S.K. Das, P. Roy, S. Bandyopadhyay, Identifying COVID19 from chest CT images: a deep convolutional neural networks based approach, J. Healthc. Eng. 2020 (2020).

[24] A. Mobiny, P.A. Cicalese, S. Zare, P. Yuan, M. Abavisani, C.C. Wu, J. Ahuja, P.M. de Groot, H. Van Nguyen, Radiologist-level COVID-19 detection using ct scans with detail-oriented capsule networks, 2020, arXiv preprint arXiv:2004.07407.

[25] M. Rahimzadeh, A. Attar, S.M. Sakhaei, A fully automated deep learning-based network for detecting COVID-19 from a new and large lung CT scan dataset, Biomed. Signal Process. Control 68 (2021) 102588.

[26] P. Kalane, S. Patil, B. Patil, D.P. Sharma, Automatic detection of COVID-19 disease using U-Net architecture based fully convolutional network, Biomed. Signal Process. Control 67 (2021) 102518.

[27] G. Jia, H.-K. Lam, Y. Xu, Classification of COVID-19 chest X-Ray and CT images using a type of dynamic CNN modification method, Comput. Biol. Med. 134 (2021) 104425.

[28] V. Arora, E.Y.-K. Ng, R.S. Leekha, M. Darshan, A. Singh, Transfer learning-based approach for detecting COVID-19 ailment in lung CT scan, Comput. Biol. Med. (2021) 104575.

[29] P. Zhang, Y. Zhong, Y. Deng, X. Tang, X. Li, CoSinGAN: learning COVID-19 infection segmentation from a single radiological image, Diagnostics 10 (11) (2020) 901.

[30] J. Zhang, L. Yu, D. Chen, W. Pan, C. Shi, Y. Niu, X. Yao, X. Xu, Y. Cheng, Dense GAN and multi-layer attention based lesion segmentation method for COVID-19 CT images, Biomed. Signal Process. Control (2021) 102901.

[31] Y. Wang, Y. Zhang, Y. Liu, J. Tian, C. Zhong, Z. Shi, Y. Zhang, Z. He, Does non-COVID-19 lung lesion help? investigating transferability in COVID-19 CT image segmentation, Comput. Methods Programs Biomed. 202 (2021) 106004.

[32] COVID-19 Chest CT dataset. http://medicalsegmentation.com/covid19.

[33] Ö. Çiçek, A. Abdulkadir, S.S. Lienkamp, T. Brox, O. Ronneberger, 3D U-Net: learning dense volumetric segmentation from sparse annotation, in: International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer, 2016, pp. 424–432.

[34] X. Chen, L. Yao, Y. Zhang, Residual attention U-Net for automated multi-class segmentation of COVID-19 chest CT images, 2020, arXiv preprint arXiv:2004.05645.

[35] COVID-19 Chest CT dataset. https://zenodo.org/record/3757476#.YZZdnr1ByGQ.

[36] F. Isensee, J. Petersen, S.A. Kohl, P.F. Jäger, K.H. Maier-Hein, NNU-Net: Breaking the spell on successful medical image segmentation, 2019, arXiv preprint arXiv:1904.08128. 1:1–8, Apr.

[37] Q. Yan, B. Wang, D. Gong, C. Luo, W. Zhao, J. Shen, Q. Shi, S. Jin, L. Zhang, Z. You, COVID-19 chest CT image segmentation–a deep convolutional neural network solution, 2020, arXiv preprint arXiv:2004.10987.

[38] 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.

[39] S. Chen, X. Tan, B. Wang, X. Hu, Reverse attention for salient object detection, in: Proceedings of the European Conference on Computer Vision, ECCV, 2018, pp. 234–250.

[40] Y. Qiu, Y. Liu, S. Li, J. Xu, Miniseg: An extremely minimum network for efficient COVID-19 segmentation, 2020, arXiv preprint arXiv:2004.09750.

[41] V. Kumar Singh, M. Abdel-Nasser, N. Pandey, D. Puig, LungINFseg: Segmenting COVID-19 infected regions in lung CT images based on a receptive-field-aware deep learning framework, Diagnostics 11 (2) (2021) 158.

[42] N. Saeedizadeh, S. Minaee, R. Kafieh, S. Yazdani, M. Sonka, COVID TV-Unet: Segmenting COVID-19 chest CT images using connectivity imposed Unet, Comput. Methods Programs Biomed. Update 1 (2021) 100007.

[43] M. Abdel-Basset, V. Chang, H. Hawash, R.K. Chakrabortty, M. Ryan, FSS-2019-nCov: A deep learning architecture for semi-supervised few-shot segmentation of COVID-19 infection, Knowl.-Based Syst. 212 (2021) 106647.

[44] A. Mikołajczyk, M. Grochowski, Data augmentation for improving deep learning in image classification problem, in: 2018 International Interdisciplinary PhD Workshop, IIPhDW, IEEE, 2018, pp. 117–122.

[45] O. Ronneberger, P. Fischer, T. Brox, U-Net: Convolutional networks for biomedical image segmentation, in: International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer, 2015, pp. 234–241.

[46] B. Kayalibay, G. Jensen, P. van der Smagt, CNN-based segmentation of medical imaging data, 2017, arXiv preprint arXiv:1701.03056.

[47] F. Iandola, M. Moskewicz, S. Karayev, R. Girshick, T. Darrell, K. Keutzer, Densenet: Implementing efficient convnet descriptor pyramids, 2014, arXiv preprint arXiv:1404.1869.

[48] S. Li, M. Dong, G. Du, X. Mu, Attention dense-U-Net for automatic breast mass segmentation in digital mammogram, IEEE Access 7 (2019) 59037–59047.

[49] O. Oktay, J. Schlemper, L.L. Folgoc, M. Lee, M. Heinrich, K. Misawa, K. Mori, S. McDonagh, N.Y. Hammerla, B. Kainz, et al., Attention U-Net: Learning where to look for the pancreas, 2018, arXiv preprint arXiv:1804.03999.

[50] J. Schlemper, O. Oktay, M. Schaap, M. Heinrich, B. Kainz, B. Glocker, D. Rueckert, Attention gated networks: Learning to leverage salient regions in medical images, Med. Image Anal. 53 (2019) 197–207.

[51] A.N.J. Raj, H. Zhu, A. Khan, Z. Zhuang, Z. Yang, V.G. Mahesh, G. Karthik, ADID-UNET—a segmentation model for COVID-19 infection from lung CT scans, PeerJ Comput. Sci. 7 (2021) e349.

[52] Y. Zhang, K. Li, K. Li, B. Zhong, Y. Fu, Residual non-local attention networks for image restoration, 2019, arXiv preprint arXiv:1903.10082.

[53] F. Isensee, P. Kickingereder, W. Wick, M. Bendszus, K.H. Maier-Hein, Brain tumor segmentation and radiomics survival prediction: Contribution to the brats 2017 challenge, in: International MICCAI Brainlesion Workshop, Springer, 2017, pp. 287–297.

[54] A. Tureckova, T. Turecek, Z. Kominkova Oplatkova, A.J. Rodriguez-Sanchez, Improving CT image tumor segmentation through deep supervision and attentional gates, Front. Robot. AI 7 (2020) 106.

[55] P. Refaeilzadeh, L. Tang, H. Liu, Cross-validation, in: Encyclopedia of Database Systems, Vol. 5, Springer, 2009, pp. 532–538.

[56] M. Sokolova, G. Lapalme, A systematic analysis of performance measures for classification tasks, Inf. Process. Manage. 45 (4) (2009) 427–437.