Normalized-UNet Segmentation for COVID-19 Utilizing an Encoder-Decoder Connection Layer Block

Mohammed Al-Mukhtar [1,*], Ammar Awni Abbas [2], Aws H. Hamad [3], and Mina H. Al-hashimi [4]

[1] Computer Center, University of Baghdad, Baghdad, Iraq
[2] College of Mass Media Baghdad, University of Baghdad, Baghdad, Iraq
[3] The Iraqi Ministry of Higher Education, Research and Development Department, Baghdad, Iraq
[4] Department of Computer Engineering, Al-Mansour University College, Baghdad, Iraq
Email: mohammed.abdul@cc.uobaghdad.edu.iq (M.A.-M.); ammarawne@yahoo.com (A.A.A.); aws.hamed@rdd.edu.iq (A.H.H.); minaalhashimi082@gmail.com (M.H.A.)
*Corresponding author

***Abstract*—The COVID-19 pandemic has had a huge influence on human lives all around the world. The virus spread quickly and impacted millions of individuals, resulting in a large number of hospitalizations and fatalities. The pandemic has also impacted economics, education, and social connections, among other aspects of life. Coronavirus-generated Computed Tomography (CT) scans have Regions of Interest (ROIs). The use of a modified U-Net model structure to categorize the region of interest at the pixel level is a promising strategy that may increase the accuracy of detecting COVID-19-associated anomalies in CT images. The suggested method seeks to detect and isolate ROIs in CT scans that show the existence of ground-glass opacity, which is frequent in COVID-19 patients. This can assist healthcare practitioners in identifying and monitoring illness development, as well as making treatment decisions. Scale U-Net is a strong U-Net design modification that can increase the performance of semantic segmentation tasks. Our model, Normalized-UNet, uses batch normalization after each convolutional layer to decrease the internal covariate shift, which dramatically improves the network's learning efficiency.**

***Keywords*—normalized-UNet, U-Net, COVID-19, scale U-Net**

I. Introduction

An epidemic of a novel coronavirus known as Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2) was first detected in Wuhan, China, in December 2019 [1]. By early 2020, the outbreak had spread to other parts of the world, with instances recorded in Europe, North America, South America, Africa, and other places. Reverse Transcription-Polymerase Chain Reaction (RT-PCR) assays were the principal diagnostic method for COVID-19 in the early phases of the pandemic. These tests, however, have limitations, such as the possibility of false negatives due to insufficient sampling or testing too early in the disease's progression [2]. Computed Tomography (CT) imaging has emerged as a valuable adjunct to RT-PCR assays, particularly when the test findings are equivocal or the patient exhibits symptoms compatible with COVID-19 but has not yet tested positive. However, deep Convolutional Neural Network (CNNs) have been utilized in medical imaging for malaria classification and detection [3]. 2D-anisotropic Total-Variation Unet (TV-UNet) has evolved into a powerful tool for radiologists to use in diagnosing disorders such as brain tumors [4], lung cancer [5], diabetic retinopathy [6], and COVID-19 [7]. In this method, segmentation is used to separate regions of interest from other parts of the image so that quantitative measurements may be performed. In existing COVID-19 CT segmentation experiments, U-Net is frequently employed as a baseline model. Extra diagnostic information is obtained, e.g., by calculating the area and volume of segmented structures. The inherent challenges of segmentation algorithms are exacerbated by intensity heterogeneity, the presence of artifacts, and the gray-level closeness of various soft tissues. In order to combat the scale difference of lesions and the low contrast between lesions and normal tissues in CT images, a unique Multiscale Dilated Convolutional Network (MSDCNet) has been developed. It attempts to extract lesions of various scales and their borders correctly. To prevent information loss during the downsampling process, a multilayer feature aggregate module has been developed. After the last layer, an MSFF module is added to fuse the output of all the layers in the block [2]. Although the correct segmentation of infected lesions is crucial for treatment decisions, it is frequently necessary to first identify COVID-19 patients.

For COVID-19 detection from chest Computed Tomography (CT) images [8], Hybrid Harris Hawks Optimization Deep Learning (CovH2SD) is a Hybrid Learning (HL) and optimization technique. CovH2SD extracts characteristics from CT scans and learns from

Manuscript received August 10, 2023; revised September 19, 2023; accepted October 23, 2023; published March 8, 2024.

them using deep learning and pre-trained models. It optimizes the hyperparameters using the Harris Hawk Optimization (HHO) method. They applied Transfer Learning (TF) comparison for nine pre-trained CNNs. Chakraborty *et al.* [9] proposed and trained deep learning strategies that select the best vector components to construct its model. It can filter out errors to produce the best results. COVID-19 was detected utilizing chest X-ray (CXR) datasets and a Deep Learning Method (DLM). The datasets are supplied to the network first, then preprocessed before being sent to the segmentation algorithm. It passes the segmented lung shape to the Deep Neural Network (DNN) model to achieve the classification result after segmentation. The pretrained weightages of the ResNet18 deep learning model are employed in the DLM model. The Mutex Attention Network (MA-Net) [10], a deep learning network for COVID-19 supplementary diagnosis based on CT images, was proposed to improve the difference between positive and negative samples. Mutex Attention Block (MAB) uses the distance between mutex inputs as a weight to detect characteristics for more accurate diagnostic results. The Fusion Attention Block (FAB) layer is intended to fuse features in the channel direction by picking features from two inputs in order to generate more representative features.

In the context of COVID-19, an AI-based automated detection and diagnosis system can provide various benefits. There are several significant advantages: 1) rapid and accurate detection; 2) improved triage and resource allocation; 3) enhanced diagnostic accuracy; 4) continuous learning and improvement. AI models can learn from large datasets of annotated images and extract meaningful features that aid in the diagnosis of COVID-19. Several studies have focused on training AI models to identify COVID-19 and distinguish it from other lung illnesses in chest CT images. Our study scaled a U-Net model through a Normalize layer after each CONV, which enhanced the training model to learn to segment Regions of Interest (ROIs) faster. Another technique for scaling U-Net is to use mixed precision training, which involves using lower-precision floating point representations for certain of the model's calculations. Due to the time-consuming manual masking of CT scans, a larger weighted module inside a U-Net may be employed to enhance CT image lesion segmentation with limited labeled data, which is common in COVID-19.

II. Literature Review

Because Chinese clinical centers lacked sufficient test kits at the outset of the pandemic, which resulted in a large number of false negative findings, physicians are recommended to base their diagnosis only on clinical and chest CT data [11]. In locations like Turkey, where there were few test kits available at the onset of the outbreak, CT is routinely used for COVID-19 identification. Deep learning (DL) techniques might be used to analyze CXRs [12] and detect COVID-19 pneumonia symptoms automatically. A huge dataset of CXRs may be used to train DL algorithms, allowing them to spot patterns and provide reliable diagnoses. In the fight against COVID-19, DL-based techniques are a potential weapon, and ongoing research in this field is expected to lead to significant breakthroughs in automated CXR analysis. DL-based techniques can considerably reduce the time required to analyze CXRs while also improving COVID-19 diagnostic accuracy. DL algorithms may also be used to prioritize CXRs for processing, allowing radiologists to prioritize the most essential patients.

The suggested system's core design [13] is based on the ResNet-50 concept. The effects of various hyperparameter values on the performance of the recommended systems are investigated. The idea is based on a modified ResNet-50. It expands the most recent levels with one Fully Connected Layer (FCL) of 512 and two more FC layers of 2048 and 1024. In both versions, the original FC layer and softmax layer are replaced. However, the authors implemented Stacked Generalization (SG), which is a method that includes training many models and then combining their predictions to outperform any single model. While SG can be a strong strategy for enhancing a model's prediction performance, it is not without drawbacks, such as complexity, overfitting, and data requirements. In ResNet-50 [14], a network (COVNet) was suggested to extract visual elements from volumetric chest CT using transfer learning. The U-Net model was used to accomplish lung segmentation as a pre-processing operation. The framework can extract both two-dimensional local and three-dimensional global representational elements.

Deep learning has resulted in greater performance in the field of radiology. Khan *et al.* [15] proposed a novel deep channel-boosting technique by distinguishing radiographic patterns, CB-STM-RENet, to identify COVID-19 in chest X-rays. The suggested approach was tested by implementing it and comparing it to numerous current CNN architectures and techniques that have been described in the literature. The suggested deep CB-STM-RENet has great classification performance, but it used a data processing technique called Split-Transform-Merge (STM). STM may be a difficult approach to implement, especially when working with big, dispersed datasets. Coordination of the processing and merging of the subgroups may present additional issues. Also, the framework proposed in [16] includes the generator and classifier networks. Its purpose is to produce more data samples for each class in a small class-unbalanced data set. In contrast, the classifier model is just used for classification problems. The Generator Model (GM) must first be trained before it can generate synthetic samples. The first deep models were a model of a convolutional variational Generative Auto-Encoder (CGA). Offline, it turned a tiny unbalanced dataset of classes into a bigger balanced dataset. Second, a large balanced dataset was used to train the ResNet50 classifier.

To address the objective of COVID-19 identification in chest X-rays, an ISNet based on a DenseNet121 classifier was used [17]. It compares the model to a U-Net, then to a DenseNet121, and finally to a solo DenseNet121. Then, the notion of optimizing DNNs for relevance segmentation in Layer-Wise Relevance Propagation (LRP) heatmaps was proposed. These heatmaps may be used for a number of purposes, including relevance segmentation, which divides the input into regions that are related to a certain output. In fact, one method is to change the loss function during training. The loss function, in particular, may be tuned to encourage the network to generate heatmaps that appropriately emphasize the key parts of the input. They were able to overcome the issue of binary and multiclass classification of COVID-19 infection by constructing two distinct COVNets that can analyze CXRs and CT scans to identify COVID-19 [18]. The authors used two datasets to complete this task: the binary classification dataset had 3877 images of COVID-19 and healthy images, while the multiclass dataset contained 6077 images of COVID-19, pneumonia, and healthy images. However, the best results depend on a CNN model, which is divided into two parts: 1) feature learning; 2) CNNs are composed of three basic layers: a convolutional layer (ConV blocks), a pooling layer, and a fully connected layer (FC). The convolution and pooling layers are used to extract features, while the fully connected layer is used to classify them. To identify COVID-19, they created two CNNs, one for binary classification and the other for multiclass classification.

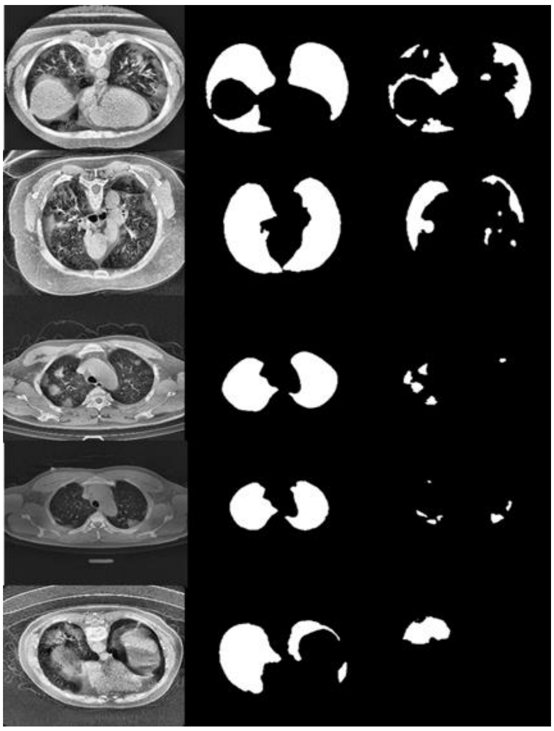
The Coronavirus Information Fusion and Diagnostic Network (CIFD-Net) [19] is a COVID-19 CT scan identification model that can handle the multidomain shift problem successfully by utilizing a unique robust weakly supervised learning technique. It uses the (SAM) a slice aggregation module to integrate multiple sections in order to estimate the patient’s joint likelihood of being COVID-19 positive or not. Furthermore, CIFD-Net was expanded with the Slice Noisy Correction Module (SNCM) to predict a single CT slice with no image-level annotations. Additionally, extensive tests were conducted using publicly accessible CT datasets to examine the proposed model’s prediction capability [20]. The model’s architecture is made up of three major processes. The first step is to prepare the image data for usage by the model through input image preprocessing. This might involve scaling, normalization, and data augmentation. The second step is to extract features from ROI pictures and train them. The Inception architecture is used to extract meaningful features from the input pictures and train the model on these features. This phase is crucial for the model’s capacity to categorize fresh photos appropriately. The last procedure is classification, which employs two fully linked layers and binary classifier prediction. The recovered features are used to categorize the input picture into one of two categories. The delimited ROIs were gathered in order to develop the classification model.

We tweaked the standard inception network and refined the Modified inception (M-inception) model using pre-trained weights. This study proposes a deep Convolutional Neural Network (CNN) [21] that can automatically recognize patterns linked to lesions from chest CT images. With an accuracy of 93.26%, the proposed model detected patterns associated with COVID-19 lesions. The suggested model may be utilized as an auxiliary system by physicians in health facilities to diagnose and evaluate diseases with high accuracy. Although it has various advantages over Fully Convolutional Networks (FCN), enhancing FCN [22] aids in the construction of U-Net. FCNs for semantic segmentation greatly improve accuracy by transferring pretrained classifier weights, combining separate layer representations, and learning end-to-end on entire images. Learning and inference are simplified and accelerated through end-to-end, pixel-to-pixel activity. To begin, U-Net is a totally symmetric algorithm. Upsampling is done by neighboring interpolation [23], and decoding is done with convolution and deepening. Second, skip connections link global and local features together to create deep feature information. Finally, proper convolution is used throughout U-Net to ensure that no contextual information is lost during segmentation. Because it conducts multi-scale fusion, merges low- and high-resolution data, and serves as the foundation for item category identification as well as precise segmentation and placement, U-Net is highly suited for medical image segmentation.

The RDCTrans model [23] is a hybrid variable structure based on U-Net that has been proposed for Liver Tumor Segmentation (LTS) in Computed Tomography (CT) scans. The network depth and the perceptual area had increased with Feature Extraction (FE) efficiency However, the number of parameters has not increased. and to create a backbone network dominated by ResNeXt50 and supplemented by dilated convolution. U-Net 3+ [24] network parameters will be reduced, while feature extraction capabilities will be enhanced. It prunes the U-Net 3+ full-scale skip connections to reduce duplication and boost computing efficiency. Furthermore, it employs Convolutional Block Attention (CBA) to gather more essential features, hence improving feature expression capabilities. To get additional edge information from the tiny contaminated patches, enhanced dilation convolution is utilized to extend the receptive field of the encoder output. ADID-UNet [25] uses a dense network instead of typical convolution and max-pooling procedures to address the problem of gradient disappearance in deep learning networks. The dense network extracts and enriches dense characteristics. Furthermore, the dense network's training parameters are smaller, which minimizes the size and computational cost. Multimodal CNN architectures were created using neural network parameters and hyperparameters [26]. A total of nine evaluations were conducted to appraise the optimizers, learning rates, and epoch count. The experimental results have been utilized to establish pertinent parameters for the creation of a multimodal architecture aimed at detecting COVID-19. However, the development of a Multimodal COVID Network (MMCOVID-NET) capable of predicting the presence of COVID-19 or normal conditions based on the analysis of both CXRs and CT scans.

The extensive neural network CNN has numerous layers and is used to categorize COVID-19 X-ray images and assess the model [27]. In comparison to the other approaches described, this method has demonstrated its effectiveness in terms of treatment speed, proper diagnosis in a short period of time, and high accuracy.

The present report provides a detailed account of the data collection process from two distinct healthcare facilities, namely Union Hospital (HUST-UH) and Liyuan Hospital (HUST-LH) [28]. This database is free for academic use only via the website https://ngdc.cncb.ac.cn/ictcf/Resource.php. Between January 25 and February 20, 2020, a total of 1170 patients were enrolled in the study, comprising 775 patients from Union Hospital (HUST-UH) and 395 patients from Liyuan Hospital (HUST-LH). The sample size consisted of 1170 individuals diagnosed with cystic fibrosis, of whom 222 were classified as controls, 23 as mild cases, 417 as regular cases, 146 as severe cases, 63 as critically ill cases, and 299 as suspicious cases. The dataset used 1613 images for CT Scan, Lung mask and COVID-19 mask. The training phase feed with 1290 while the test phase feed with 323 images. We used another publicly accessible "COVID-19 CT Segmentation dataset" [29], which comprises 100 axial CT pictures from 40 individual COVID-19 patients. In traditional machine learning or deep learning methods, an imbalanced distribution of data sets to be classed has a detrimental impact on classification results. Fig. 1 shows some samples of the dataset with lung segmentation and COVID-19 masks.



Original CT Scan | Lung Mask Annotation | COVID-19 Mask Annotation

Fig. 1. Several dataset samples utilizing CT scan originals, Lung mask and COVID-19 mask lesions.

III. MATERIALS AND METHODS

The CNN model architecture is the suggested network design that we discuss in detail in this section. The modified CNN model distinguishes the proposed technique. Convolution layers serve as the foundation for several semantic segmentation approaches and models. The network can learn increasingly complicated representations of the input image by stacking many convolutional layers on top of each other. Convolutional layers are frequently utilized in an encoder–decoder architecture in the context of semantic segmentation. The encoder is often made up of many layers of convolutional and pooling algorithms that extract information from the input picture at various sizes. The decoder then upsamples the feature maps and generates a dense prediction for each pixel in the picture using transposed convolutional layers. This encoder–decoder architecture underpins several prominent semantic segmentation models, including U-Net [30] and SegNet [31]. Other systems, such as Mask R-CNN, conduct (OB) Object Detection and segmentation methods by combining convolutional layers with other types of layers (such as fully connected layers and region proposal networks).

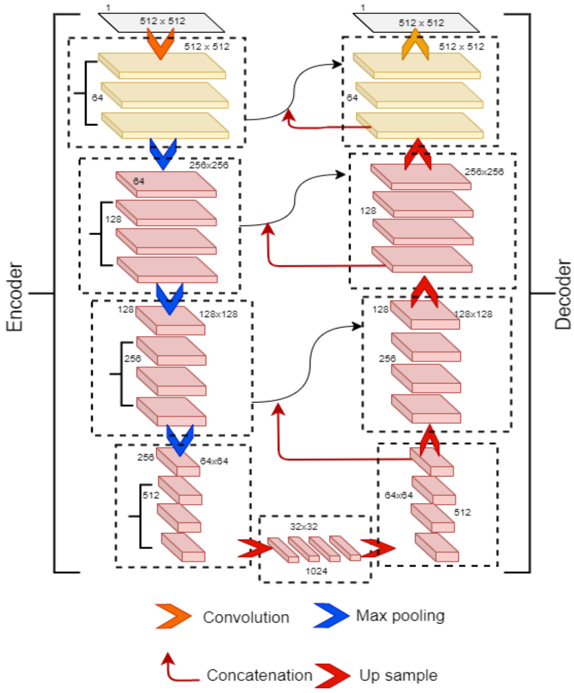


Fig. 2. The proposed segmentation model with encoder and decoder layers.

At first, the encoder part uses the convolved operation with a filter of size 3×3 on the input feature map. To generate a new feature map, this filter glides over the input feature map and executes element-wise multiplication and summation operations. A pooling operation with a size of 2×2 and a stride size of 2 is used. By grouping neighboring features, pooling methods minimize the spatial dimensions of feature maps. In this scenario, the pooling procedure with a size of 2×2 and stride size of 2 decreases the feature map's spatial dimensions by a factor of 2. This method is repeated for all non-overlapping regions in the feature map, yielding a new feature map with smaller spatial dimensions. The most important goal for the encoder part is that the combination of the 3×3 convolution operation with the pooling operation aids in the extraction and aggregation of essential features from the input data while also lowering the spatial dimensions of the feature maps to improve the efficiency of future calculations. Fig. 2 shows the decoder is often followed by an upsampling process that uses transposed convolutions to transfer the low-resolution feature maps to a higher resolution. Transposed convolutions are often referred to as deconvolutions or fractionally strided convolutions.

The model is made up of eight blocks that capture spatial data from CT scans. The operation was repeated four times in an encoder. The number of convolution layers in each encoder and decoder design block was increased in this study. Fig. 3 depicts the passage of the input picture via the encoder pipeline. Each block defines the procedure as well as the number of convolution layers used.

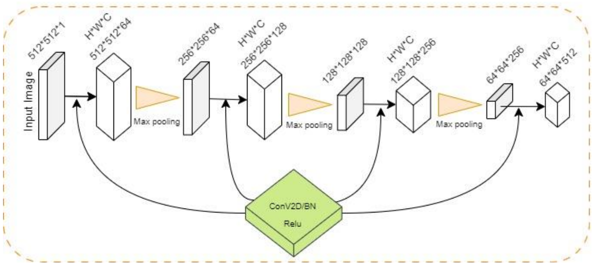


Fig. 3. Encoder block within the normalization layer.

Fig. 2 depicts the N-UNet design in further detail. The input picture is convolved using batch normalization and a nonlinear function to reduce the number of channels. A max-pooling process is then done to generate a pooled feature map. The encoder route repeats this sequence of actions four times to reduce the amount of the input. The softmax activation function is generally used in the last layer of a semantic segmentation model to convert the model's output into a probability distribution across the multiple class labels.

A popular strategy in the context of U-Net is to utilize Batch Normalization (BN) to equalize the input to each layer. This can help to stabilize the training process and prevent overfitting. However, when training with large batch sizes, BN can introduce issues with the batch statistics used for normalization, which can have a negative impact on performance. One option for scaling U-Net to bigger batch sizes is to utilize Group Normalization (GN). Our N-UNet used a batch normalization layer as described in Fig. 3. It is a normalization algorithm that has been proven to function well in high-batch-size situations and may be used to replace BN in CNN systems. Another method for scaling U-Net is to employ mixed-precision training, which entails utilizing lower-precision floating point representations for portions of the model's calculations. This can greatly reduce training memory consumption and computational cost, allowing for bigger batch sizes without running out of memory or surpassing computing resource restrictions.

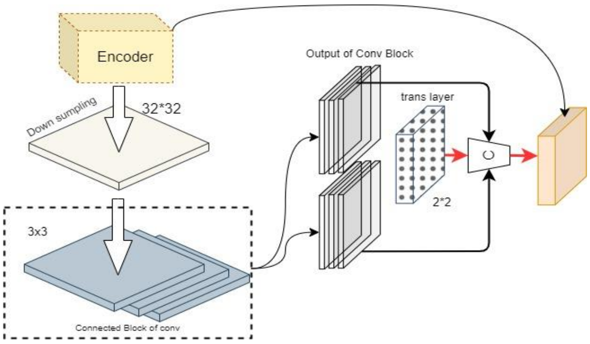


Fig. 4. Layer block with encoder–decoder connection.

The transposed convolution operation applies a filter (also known as a kernel or weight matrix) to each pixel in the input feature map during the upsampling process and returns a bigger feature map with the appropriate spatial resolution. The quantity of upsampling is determined by the size of the transposed convolutional filter. Fig. 4 illustrates layer block with encoder–decoder connection with the filter in this example is 2×2. It resulting in a doubling of the spatial dimensions of the feature map. The upsample operation was then followed by two convolution layers of size 3×3. The Feature Map (FM) from the decoder's last block performs a 1×1 convolution operation, yielding a Segmentation Map (SM) of the same size as the input image. The process mechanism is summarized as follows:

1) The revised model is utilized. Before each nonlinear function, batch normalization is performed. Batch normalization has no place in a traditional U-Net network. It operates by normalizing a specific layer's activations over a batch of training data so that the mean activation is near zero and the standard deviation is close to one. This can assist in limiting the influence of disappearing or expanding gradients, as well as regularizing the model by minimizing the impact of tiny changes in the input.
2) Each encoder structure has two convolution layers added to it. The network can learn to extract increasingly abstract and sophisticated characteristics as the input is downsampled by adding convolution layers to each encoder block. As the input is downsampled, the network may learn to extract increasingly abstract and sophisticated information.
3) Each block of the decoder circuit has the same number of convolution layers. This contributes to the network's ability to reconstruct the input with the same level of detail and complexity as the encoder. The network may learn to upsample the features while preserving the structure and content of the original input by employing the same number of convolution layers in the decoder.
4) Weight settings are the parameters that the network learns during training and that govern the

network's behavior. This architecture may be able to understand more complicated and subtle correlations between input and output data by having more weight settings, which might lead to improved performance on specific tasks.

The Batch U-Net employed eight blocks, four of which formed the downsampling path and four of which comprised the upsampling path. In general, each encoder and decoder pipeline block include two fundamental components:

1) Batch normalization: This process normalizes the activation function's input, making the network more resistant to variations in input distribution and speeding up training. It aids in preventing the internal covariate shift problem, which can arise in deep networks.
2) The ReLU activation function adds nonlinearity to the network, helping it to learn complicated representations of the input data.

The Rectified Linear Unit (ReLU) simply sets negative numbers to zero while leaving positive ones alone. The use of max-pooling for downsampling is a common technique in convolutional neural networks for image processing. Max-pooling reduces the spatial dimensions of the feature maps by taking the maximum value within a fixed window size. Max-pooling is applied after each block to downsample the feature maps by a factor of 2 in each spatial dimension. The output of each block, which consists of convolution layers produced prior to the pooling process in the encoder, is transferred to the appropriate decoder block through a skip link. Using skip connections, the encoder route's feature maps are concatenated with the decoder route's upsampled feature maps at the respective scales. This allows the decoder to retrieve spatial information from the input picture that was lost during the encoder's downsampling, boosting the network's segmentation accuracy. It is also an important part of the encoder–decoder design. Skip connections allow the decoder to retrieve spatial features from the input picture that were lost during the encoder's downsampling, boosting overall segmentation or reconstruction accuracy. Before being transmitted to the next block in the decoder pipeline, the concatenated feature maps are generally processed by a sequence of convolution layers, followed by batch normalization and ReLU activation functions. The size and number of convolution layers can be adjusted based on the task's difficulty and the size of the input picture. The convolution layer can be represented as *S*, while the feature map is denoted by *f* and the layer number is denoted by *n*. The bias term *b* is a model parameter that indicates the intercept or offset of the linear equation. The feature map in the first layer f^1 derives from convolving the input matrix by a kernel M^1. The feature map number is denoted by *i* in Eq. (1). And before the filtered output is transmitted to the convolution layer, a nonlinear function denoted as *f(y)* is applied to it, where *X* denotes the input neuron.

$$S_i^1 = f\left(b_i^1 + M_i^1 \times X\right) i = 1, 2, ..., f^{n-1}, f^n \quad (1)$$

The first convolution layer S^1 convolves the input layer *x* with a weight matrix to create a Feature Map (FM), such as Eq. (2). The f^1 feature map is created by sliding across the input matrix's different points based on the stride value. Because the weight parameters are shared across all infection classes (ground-glass opacity (GGO), consolidation) in our dataset, they are retrieved in this manner. As a consequence, the layer obtains equivariance and is no longer affected by image changes. To extract features, many layers turn the input data into higher-level representations that capture more complicated and abstract properties. These layers are made up of connected neurons, and each neuron modifies its input linearly before applying a non-linear activation function.

$$S_i^1 = f\left(b_i^n + \sum_{x=1}^{n-1} M_i^n \times X\right) i = 1, 2, ..., f^{n-1}, f^n \quad (2)$$

The mean mini-batch μ_b and mini-batch variance σ_B^2 were used in [32], along with batch normalization in Eq. (3). We can use much higher learning rates while being less cautious about initialization because they are criteria that can be learned. If a fluctuation in input distribution is required for the model to learn a specific class better, the layers learn the optimal scaling factor gamma (γ) and offset beta (β) values for each mini-batch.

$$\hat{x} = \frac{x_i - \mu_b}{\sqrt{\sigma_b^2 + \varepsilon}} \quad (3)$$

To stabilize the learning process and decrease the number of epochs, batch normalization is often inserted between the convolutional and ReLU layers. The batch normalization is performed after each convolutional layer to minimize the internal covariate shift, which considerably enhances the network's learning efficiency.

IV. Result and Discussion

The outcomes in detail present both quantitative results produced with the N-UNet model and a quantitative comparison of two popular algorithms (SegNet and U-Net) with the improved model. The Segmentation Evaluation Index (SEI) is a quantitative indicator for evaluating the efficacy of image segmentation algorithms. It is a synthesis of numerous indicators that provides an overall assessment of the model's performance. Furthermore, in comparison to the other models in our study, we obtained good segmentation when compared to the ground truth. We used accuracy to evaluate the network; it computes the ratio of the sum of true positive and true negative predictions to the total number of predicted values (both true and false). The F1 score is simply one of many possible assessment metrics for models. Fig. 5 shows the accuracy, area under the receiver operating characteristic curve metrics (ROC), and area under the precision–recall curve (AUC) for both models.

The proposed method used to segment the infection was a successful improvement as a result of the improvements we applied for intense learning in the baseline algorithm U-Net.

The model's evaluation metric findings, displayed in Table I, further demonstrate the soundness and robustness of the proposed technique. Based on these data, we can conclude that the proposed strategy outperforms both the baseline U-Net approach and other cutting-edge methodologies.

Table II displays the quantitative results of the suggested model and the baseline approach U-Net. The comparative results of other models described in Table II show the Dice coefficient, sensitivity, and F1-score. The normalized U-Net achieves 0.8284 for the Dice coefficient, 0.8096 for sensitivity, and 0.8357 for F1-score.

TABLE I. ACCURACY OF OUR N-UNET IN COMPARISON WITH OTHER STATE-OF-THE-ART MODELS

Algorithms	Na/ images	CT or X-ray	ACC
COVIDX-Net [12]	50	X-ray	90.00%
Deep convolutional neural network (CNN) [21]	16,040	CT	93.26%
-Random forest classifier -Logistic regression [24]	1848	CT	92% severe class 79% severe class
Multi-task learning [33]	1396 100	CT	91.03% E1 95.23% E2
DRE-Net [34]	-	Chest CT	86%
ResNet with location attention [35]	-	CT-COVID-19	86.7%
Our U-Net	**1613**	**U-Net Segmentation**	**95.78%**
Our SegNet	**1613**	**Segmentation**	**91.65%**
Our N-UNet	**1613**	**Normalized U-Net Segmentation**	**97.04%**

TABLE II. COMPARATIVE RESULTS OF THE SUGGESTED MODELS WITH N-UNET MODEL

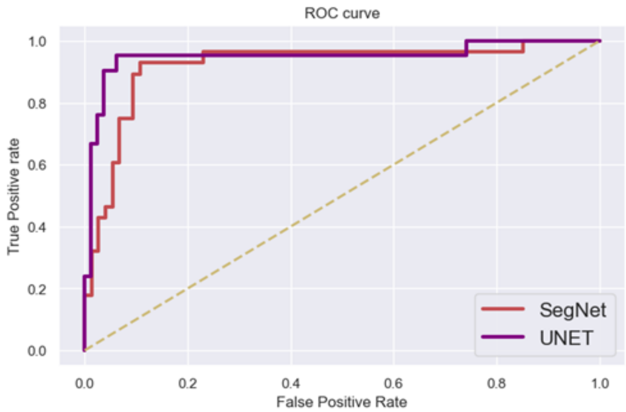
Model	Dice	Sen	F1-Score
MSDC-Net [2]	0.824	0.811	-
DDANET [11]	0.77	0.88	-
Deep learning algorithm [20]	0.75	0.88	0.77
ADIDUNET	0.8031	0.7973	0.8200
UNET	0.7998	0.8052	0.8154
SegNet [25]	0.7408	0.7608	0.7558
Encoder-Decoder Architecture [36]	0.786	0.711	-
SegNet C1 multi segmentor	0.479	0.638	0.562
UNET C1 multi segmentor [37]	0.483	0.804	0.636
N-UNet	**0.8284**	**0.8096**	**0.8357**

TABLE III. COMPARISON OF TEST RESULTS BETWEEN EXISTING MODELS AND PROPOSED MODEL ON COVID-19 CHEST CT SEGMENTATION DATASET [37]

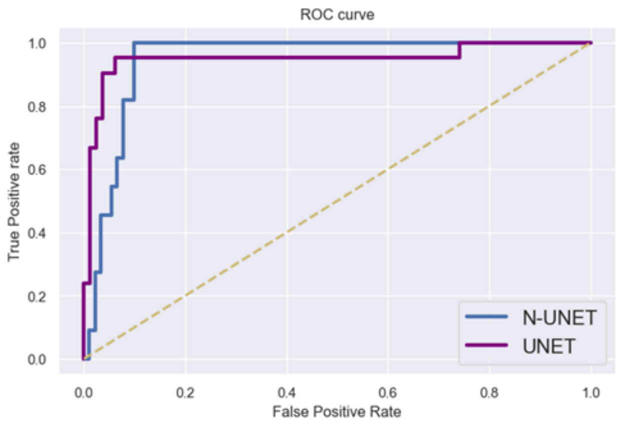
Algorithm	Dice_coef	Sens	Spec
TV-UNet [7]	80.1%	80.8%	96.0%
DenseNet [17]	74.34%	78.35%	91.43%
U-Net [29]	74.03%	76.51%	97.34%
SegNet [30]	71.12%	78.07%	95.23%
N-UNet	**81.453%**	**82.254%**	**97.687%**

Table III shows the comparison results for different models trained on public datasets [37]. The N-UNet model's performance is compared with many other approaches in terms of sensitivity, specificity, and the Dice coefficient for the GG-mask on the COVID-SemiSeg dataset. The best performance is given in bold text. We trained three different networks as described in Table III: SegNet, U-Net, and N-UNet.

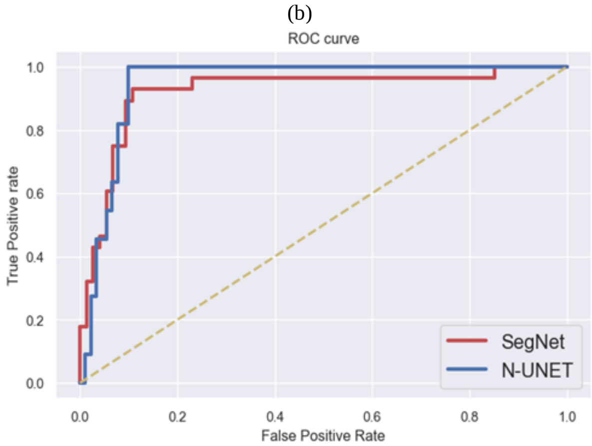
Four semantic segmentation models are used to compare with our model, Normalized U-Net, as represented in Table III. Our N-UNet outperforms the compared networks in teams of the Dice coefficient, sensitivity, and specificity, with values of 81.0%, 82.25%, and 97.68%, respectively.



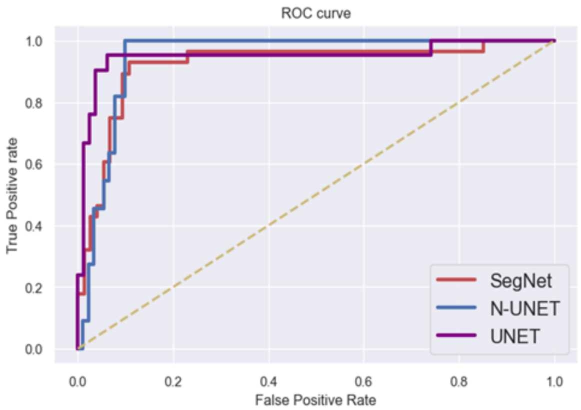
(a)



(b)

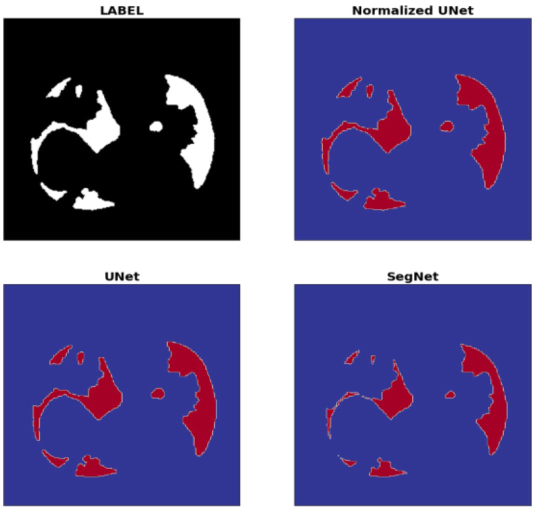


(c)

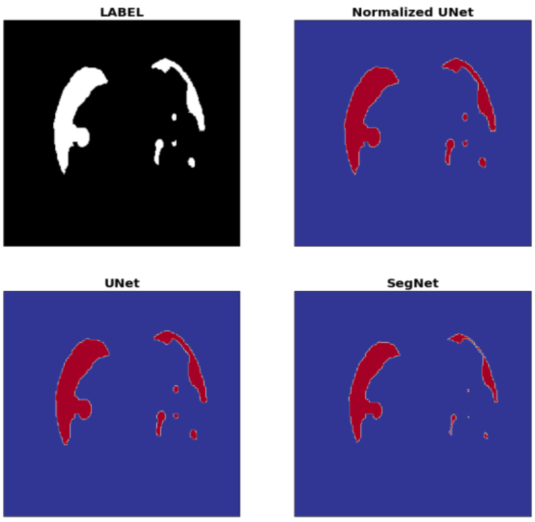


(d)

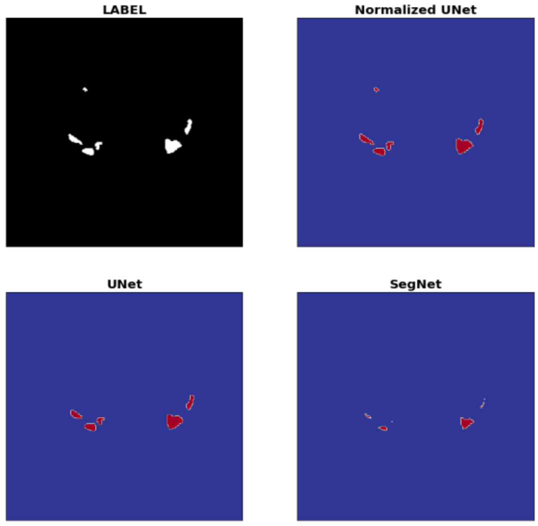
Fig. 5. SegNet, U-Net, and N-UNet training accuracy and ROC curve. Four plots of training and accuracy for the optimal configuration of each model. (a) refers to a separate U-Net, where the SegNet model has accuracy of 91.65% and U-Net has accuracy of 95.78%. (b) reflects a more accurate N-UNet model with accuracy of 97.04%. In our investigation, we achieved good segmentation results when compared to other state-of-the-art models in our study. (c) shows the deep learning (N-UNet and SegNet) algorithm's receiver operating characteristic plots for COVID-19 identification. (d) shows the performance for three algorithm's receiver operating characteristic plots for COVID-19 identification.



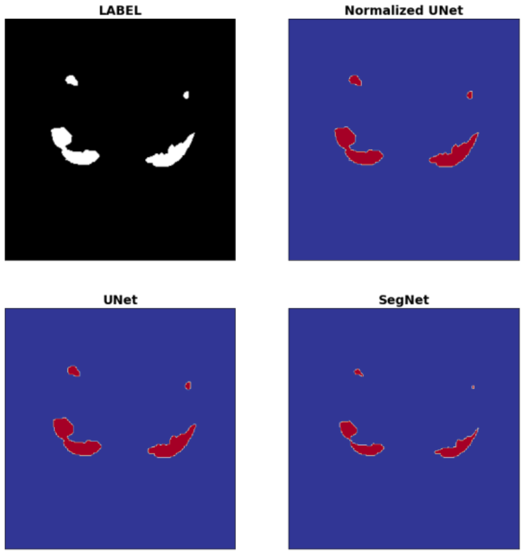
(a)



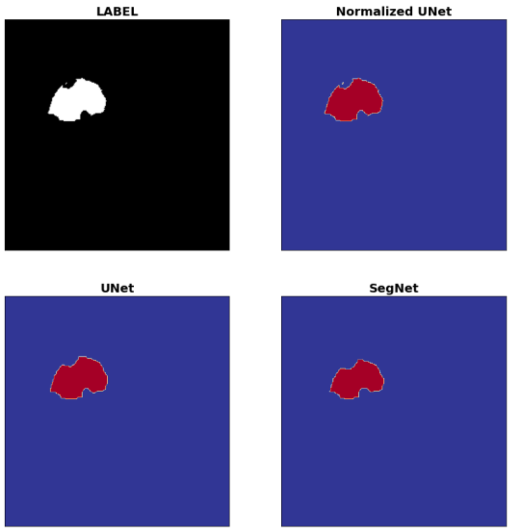
(b)



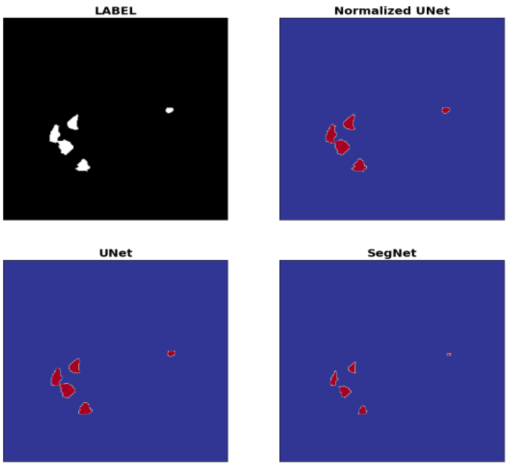
(c)



(d)



(e)



(f)

Fig. 6. Comparative results (a)–(f) of the proposed network utilizing N-UNet, U-Net, and SegNet segmentations.

Because of the changes we incorporated for intense learning in the baseline technique U-Net. the recommended strategy utilized to segment the infection was a successful improvement. The model involved data augmentation and tuning of the main settings for the algorithm hyperparameters, such as the learning rate, batch size, regularization strength, and number of hidden layers. Fig. 6 presents some examples of the data obtained in order to evaluate the effectiveness of the suggested method for segmenting lung infection lesions.

V. Conclusion

This paper proposes a deep normalized U-Net model-based CNN for detecting COVID-19 in a CT scan dataset that scales up with a normalized layer. The proposed N-UNet method was implemented and compared to several current CNN architectures and techniques reported in the literature to assess it. Scaling U-Net entails a combination of optimization approaches and architectural changes appropriate to the individual work at hand and the computational resources available. Normalization methods such as GN or IN, mixed precision training, and data augmentation to expand the size of the training dataset are some of the most frequent strategies. The learning curves were used to assess the robustness of the proposed system. Furthermore, the proposed normalized technique was compared to other recent systems, and the results show that the proposed model outperforms the others. Normalized upscaling of the model with fine-tuning had a great impact on developing the segment lesions properly. As a consequence, our proposed N-UNet approach can aid in the rapid identification, testing, and quantification of COVID-19-infected areas while also improving overall detection of the lesions.

Conflict of Interest

The authors declare no conflict of interest.

Author Contributions

Mohammed Al-Mukhtar Ammar Awni Abbas conducted study concepts, study design, data acquisition, data analysis, and interpretation; Mohammed Al-Mukhtar, Ammar Awni Abbas did experimental studies, statistical analysis, manuscript editing with figures and writing original draft; Aws H. Hamad, Mina H. Al-hashimi revised the methodology, write the introduction and literature research studies, writing-review and editing; all authors had approved the final version.

References

[1] A. Bartoli *et al.*, "Value and prognostic impact of a deep learning segmentation model of COVID-19 lung lesions on low-dose chest CT," *Research in Diagnostic and Interventional Imaging*, vol. 1, 100003, 2022.

[2] J. Zhang *et al.*, "Semantic segmentation of COVID-19 lesions with a multiscale dilated convolutional network," *Scientific Reports*, vol. 12, 1847, 2022.

[3] M. A. Jabbar *et al.*, "Diagnosis of malaria infected blood cell digital images using deep convolutional neural networks," *Iraqi Journal of Science*, pp. 380–396, 2022.

[4] M. Aloraini *et al.*, "Combining the transformer and convolution for effective brain tumor classification using MRI images," *Applied Sciences*, vol. 13, 3680, 2023.

[5] S. Makaju *et al.*, "Lung cancer detection using CT scan images," *Procedia Computer Science*, vol. 125, pp. 107–114, 2018.

[6] M. Al-Mukhtar *et al.*, "Weakly supervised sensitive heatmap framework to classify and localize diabetic retinopathy lesions," *Scientific Reports*, vol. 11, 23631, 2021.

[7] N. Saeedizadeh *et al.*, "COVID TV-Unet: Segmenting COVID-19 chest CT images using connectivity imposed Unet," *Computer Methods and Programs in Biomedicine Update*, vol. 1, 100007, 2021.

[8] H. M. Balaha *et al.*, "CovH2SD: A COVID-19 detection approach based on Harris Hawks Optimization and stacked deep learning," *Expert Systems with Applications*, vol. 186, 115805, 2021.

[9] S. Chakraborty *et al.*, "An efficient deep learning model to detect COVID-19 using chest X-ray images," *International Journal of Environmental Research and Public Health*, vol. 19, 2013, 2022.

[10] B. B. Zheng *et al.*, "MA-Net: Mutex attention network for COVID-19 diagnosis on CT images," *Applied Intelligence*, vol. 52, pp. 18115–18130, 2022.

[11] K. T. Rajamani *et al.*, "Dynamic Deformable Attention Network (DDANet) for COVID-19 lesions semantic segmentation," *Journal of Biomedical Informatics*, vol. 119, 103816, 2021.

[12] E. E. D. Hemdan *et al.*, "Covidx-net: A framework of deep learning classifiers to diagnose COVID-19 in x-ray images," arXiv preprint, arXiv:2003.11055, 2020.

[13] W. Gouda *et al.*, "Detection of COVID-19 based on chest x-rays using deep learning," *Healthcare*, vol. 10, no. 2, 343, 2022.

[14] L. Li *et al.*, "Using artificial intelligence to detect COVID-19 and community-acquired pneumonia based on pulmonary CT: Evaluation of the diagnostic accuracy," *Radiology*, vol. 296, pp. E65–E71, 2020.

[15] S. H. Khan *et al.*, "COVID-19 detection in chest X-ray images using a new channel boosted CNN," *Diagnostics*, vol. 12, 267, 2022.

[16] W. M. Salama *et al.*, "A generalized framework for lung cancer classification based on deep generative models," *Multimedia Tools and Applications*, vol. 81, pp. 32705–32722, 2022.

[17] P. Bassi, "ISNet: Costless and implicit image segmentation for deep classifiers, with application in COVID-19 detection," arXiv preprint, PPR: PPR516966, 2022.

[18] S. Thakur and A. Kumar, "X-ray and CT-scan-based automated detection and classification of covid-19 using Convolutional Neural Networks (CNN)," *Biomedical Signal Processing and Control*, vol. 69, 102920, 2021.

[19] Q. Ye *et al.*, "Robust weakly supervised learning for COVID-19 recognition using multi-center CT images," *Applied Soft Computing*, vol. 116, 108291, 2022.

[20] S. Wang *et al*., "A deep learning algorithm using CT images to screen for Corona Virus Disease (COVID-19)," *European Radiology*, vol. 31, pp. 6096–6104, 2021.

[21] H. Polat *et al*., "Automatic detection and localization of COVID-19 pneumonia using axial computed tomography images and deep convolutional neural networks," *International Journal of Imaging Systems and Technology*, vol. 31, pp. 509–524, 2021.

[22] J. Long *et al*., "Fully convolutional networks for semantic segmentation," in *Proc. the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 3431–3440, 2015.

[23] L. Li and H. Ma, "Rdctrans Unet: A hybrid variable architecture for liver CT image segmentation," *Sensors*, vol. 22, 2452, 2022.

[24] A. A. Alrajhi *et al*., "Data-driven prediction for COVID-19 severity in hospitalized patients," *International journal of Environmental Research and Public Health*, vol. 19, 2958, 2022.

[25] A. N. J. Raj *et al*., "ADID-UNET—A segmentation model for COVID-19 infection from lung CT scans," *PeerJ Computer Science*, vol. 7, p. e349, 2021.

[26] T. Padmapriya, T. Kalaiselvi, and V. Priyadharshini, "Multimodal covid network: multimodal bespoke convolutional neural network architectures for COVID-19 detection from chest X-ray's and computerized tomography scans," *International Journal of International Journal of Imaging Systems and Technology*, vol. 32, pp. 704–716, 2022.

[27] O. Kadhim *et al*., "Detection of COVID-19 in X-Rays by convolutional neural networks," *Iraqi Journal of Science*, pp. 1963–1974, 2023.

[28] W. Ning *et al*., "iCTCF: An integrative resource of chest computed tomography images and clinical features of patients with COVID-19 pneumonia," arXiv preprint, PPR: PPR141530, 2020.

[29] COVID-19 Chest CT dataset. [Online]. Available: http://medicalsegmentation.com/covid19

[30] O. Ronneberger *et al*., "Unet: Convolutional networks for biomedical image segmentation," in *Proc. 18th International Conference on Medical Image Computing and Computer-Assisted Intervention, Part III 18, MICCAI 2015: Munich, Germany*, October 5–9, 2015, pp. 234–241.

[31] V. Badrinarayanan *et al*., "Segnet: A deep convolutional encoder-decoder architecture for image segmentation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 39, pp. 2481–2495, 2017.

[32] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," in *Proc. International Conference on Machine Learning*, 2015, pp. 448–456.

[33] A. Amyar *et al*., "Multi-task deep learning based CT imaging analysis for COVID-19 pneumonia: Classification and segmentation," *Computers in Biology and Medicine*, vol. 126, 104037, 2020.

[34] Y. Song *et al*., "Deep learning enables accurate diagnosis of novel coronavirus (COVID-19) with CT images," *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, vol. 18, pp. 2775–2780, 2021.

[35] X. Xu *et al*., "A deep learning system to screen novel coronavirus disease 2019 pneumonia," *Engineering*, vol. 6, pp. 1122–1129, 2020.

[36] O. Elharrouss, N. Subramanian, and S. Al-Maadeed, "An encoder-decoder-based method for segmentation of COVID-19 lung infection in CT images," *SN Computer Science*, vol. 3, pp. 1–12, 2022.

[37] A. Saood and I. Hatem, "COVID-19 lung CT image segmentation using deep learning methods: Unet versus SegNet," *BMC Medical Imaging*, vol. 21, pp. 1–10, 2021.