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

Journal of Biomedical Informatics 119 (2021) 103816

Contents lists available at ScienceDirect

Journal of Biomedical Informatics

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

Original Research

# Dynamic deformable attention network (DDANet) for COVID-19 lesions semantic segmentation

Kumar T. Rajamani [1,*], Hanna Siebert, Mattias P. Heinrich

*Institute of Medical Informatics, University of Lübeck, Germany*

ARTICLE INFO

*Keywords:*
Deformable attention
Attention mechanism
COVID-19
Ground-glass opacity
Consolidation
RT-PCR
Infection
Semantic segmentation
U-Net
Segmentation
Computed Tomography (CT)
Differentiable attention sampling
CCNet
Criss-cross attention

ABSTRACT

Deep learning based medical image segmentation is an important step within diagnosis, which relies strongly on capturing sufficient spatial context without requiring too complex models that are hard to train with limited labelled data. Training data is in particular scarce for segmenting infection regions of CT images of COVID-19 patients. Attention models help gather contextual information within deep networks and benefit semantic segmentation tasks. The recent criss-cross-attention module aims to approximate global self-attention while remaining memory and time efficient by separating horizontal and vertical self-similarity computations. However, capturing attention from all non-local locations can adversely impact the accuracy of semantic segmentation networks. We propose a new Dynamic Deformable Attention Network (DDANet) that enables a more accurate contextual information computation in a similarly efficient way. Our novel technique is based on a deformable criss-cross attention block that learns both attention coefficients and attention offsets in a continuous way. A deep U-Net (Schlemper et al., 2019) segmentation network that employs this attention mechanism is able to capture attention from pertinent non-local locations and also improves the performance on semantic segmentation tasks compared to criss-cross attention within a U-Net on a challenging COVID-19 lesion segmentation task. Our validation experiments show that the performance gain of the recursively applied dynamic deformable attention blocks comes from their ability to capture dynamic and precise attention context. Our DDANet achieves Dice scores of 73.4% and 61.3% for Ground-glass opacity and consolidation lesions for COVID-19 segmentation and improves the accuracy by 4.9% points compared to a baseline U-Net and 24.4% points compared to current state of art methods (Fan et al., 2020).

## 1. Introduction

The coronavirus COVID-19 pandemic is having a global impact affecting 213 countries so far. The cases world wide as reported on Worldometers [3] is about 154,998,238 as of early May 2021. Many of the countries have steadily flattened the curve by stringent social distancing measures. In the last several months of managing this pandemic globally, several screening options have become main stream from Nucleic Acid Amplification Tests (NAAT) assay tests, serological tests, and radiological imaging (X-rays, CT). Recent studies have also demonstrated that lack of taste and smell is a new indicator for this virus [4].

The gold-standard for COVID-19 diagnosis is currently using reverse-transcription polymerase chain reaction (RT-PCR) testing [5]. It has been observed that RT-PCR also has several vital limitations. The most pertinent of this limitation is that the test is not universally available. To further compound the drawbacks, the turnaround times for this test is currently lengthy and the sensitivities vary. Some studies have even pointed out that the sensitivity of this test is largely insufficient [5]. To mitigate some of the challenges in rapid screening given the large incidence rate of this virus and limited testing facility, radiological imaging complements and supports immensely stratify therapy options for more severe cases of COVID-19.

Radiological imaging equipment, such as X-ray, are more easily accessible to clinicians and also provide huge assistance for diagnosis of COVID-19. CT imaging and Chest radiographs (CXR) are two of the

* Corresponding author.
*E-mail addresses:* kumar.rajamani@uni-luebeck.de, kumartr@gmail.com (K.T. Rajamani), siebert@imi.uni-luebeck.de (H. Siebert), heinrich@imi.uni-luebeck.de (M.P. Heinrich).
*URL:* https://www.linkedin.com/in/kumartr/ (K.T. Rajamani).
[1] Kumar Rajamani is Postdoctoral Researcher at University of Luebeck

https://doi.org/10.1016/j.jbi.2021.103816
Received 1 September 2020; Received in revised form 5 May 2021; Accepted 16 May 2021
Available online 20 May 2021

currently used radiological imaging modalities for COVID-19 screening. Lung CT can detect certain characteristic manifestations associated with COVID-19. Several studies [6] [5] have demonstrated that CT is more sensitive to detect COVID-19, with 97–98%, compared to 71% for RT-PCR [5]. CXR might have lesser scope in the first stages of the disease as the changes are not evident on CXR. Studies have shown [7,8] that CXR may even present normal in early or mild disease, as demonstrated in Fig. 1 [9]. CT is hence preferred for early stage screening and is also generally better than X-rays as it enables three dimensional views of the lung.

The typical signs of COVID-19 infection observed in CT slices are Ground-glass opacities (GGO), which occur in the early stages and pulmonary consolidation, which occur in later stages. Detection of these regions in CT slices gives vital information to the clinicians and helps in combating COVID-19. Manual detection is laborious, highly time consuming, tedious and error prone. It has to be pointed out that COVID-19 associated abnormalities, such as Ground-glass opacities and consolidations, are not characteristic for only COVID-19 but can occur in other forms of pneumonia.

Deep Learning plays a vital role in processing these medical images and correctly diagnosing patients with COVID-19. In regular clinical workflow, while assessing the risks for progression or worsening, the images need to be segmented and quantified.

For the diagnosis of lung diseases, CT scans have been the preferred modality, and this has therefore been actively utilized in managing COVID-19 [11] [12] [13]. AI in medical imaging has largely aided in automating the diagnosis of COVID-19 from medical images [14] [15]. A detailed review of AI in Diagnosis of COVID-19 has been presented by Shi et al. [11]. They broadly group AI based automated assistance for image acquisition, accurate segmentation of organs and infections and for clinical decision making. Under the segmentation approaches they have comprehensively covered nearly all the research that has happened so far in the automated segmentation of lung regions and lesion regions in CT and X-ray images.

Fan et al. [2] have reported a list of at this time available public COVID-19 imaging datasets. As mentioned in their paper, there is only one dataset which provides segmentation labels [16]. From this public database [16], we have combined the first dataset of 100 sparsely selected axial CT slices from over 40 patients with a dense set of slices from 9 patient's CT scans and use this larger datasets for our studies. A few exemplary slices are demonstrated below to get a visual impression of how the Ground-glass opacity lesions and consolidation lesions manifests itself in Fig. 2.

In this paper, we propose Dynamic Deformable Attention Network (DDANet), a novel deep network for COVID-19 infection segmentation in 2D CT slices. Our inspiration for this network is the recent success of self attention mechanisms and sparse deformable convolutions [17]. As attention blocks do not have to be regularly structured, this opens the

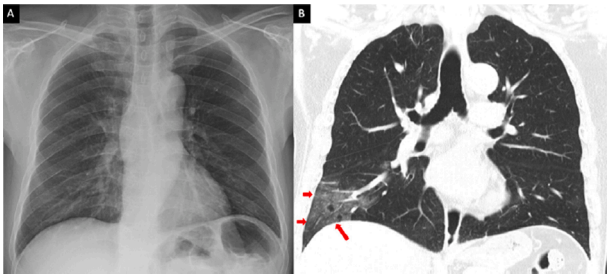


**Fig. 1.** Comparison of chest radiograph (A) and CT thorax coronal image (B). The Ground-glass opacities in the right lower lobe periphery on the CT (red arrows) are not visible on the chest radiograph, which was taken 1 h apart from the first study. Image courtesy - Ming-Yen et al. [10]. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

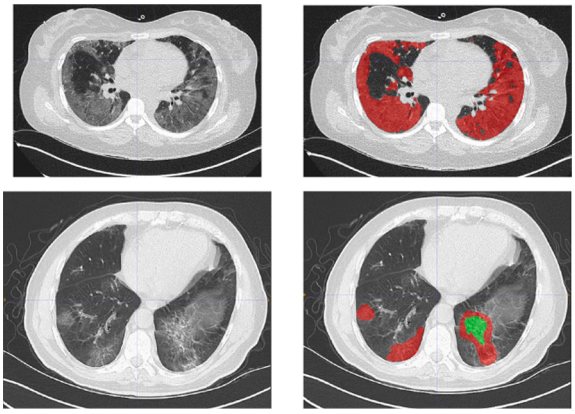

**Fig. 2.** Sample slice from one of the dataset and the corresponding Ground-glass opacity lesion (GGO) marking in first row and GGO and consolidation lesion marking in second row. Dataset from website [16].

novel research area that motivates our investigation of spatially-adaptive attention filters. In this work we generalize criss-cross attention [18] for semantic segmentation tasks. We enhance the criss-cross attention and propose a novel deformable attention module in which both the attention filter offsets and coefficients are learnt in a continuous, differentiable space. We carried out extensive experiments of our novel algorithm on a large publicly available COVID-19 dataset. Our proposed DDANet achieves very good lesion segmentation and outperforms most cutting-edge segmentation models reported so far on Ground-glass opacity and consolidation lesions. The proposed solution greatly enhances the performance of the baseline U-Net architecture [19]. The baseline U-Net we have employed in our work is from Oktay et al. [19], which has a well-proven semantic segmentation performance. Our novel adaptation of the criss-cross attention module is generic and can also be easily plugged into any state-of-art segmentation architecture. These results demonstrate that our proposed DDANet can be effectively used in image segmentation in general and COVID-19 automated image analysis in particular and can greatly aid in clinical workflow handling of these images.

In summary, our main contributions in our work are:

- We propose a novel deformable attention module in which sparse attention filter offsets are learnt in a continous differentiable space and can capture contextual information in an efficient way
- We demonstrate that employing this new deformable attention mechanism within the U-Net architecture [19] [1] achieves superior performance of lung infection segmentation compared to conventional U-Nets or U-Nets with criss-cross attention [18]
- Our DDANet reaches state-of-the-art segmentation performance of 73.4% and 61.3% dice scores for Ground-glass opacity and consolidation lesions, on a large publicly available CT COVID-19 infection dataset in a fivefold cross validation on GGO and consolidation labels. It improves the accuracy by 4.9% points compared to a baseline U-Net and 24.4% points compared to current state of art methods [2]

The rest of the paper is organized as follows. In Section II we present the works related to our current research. We start with a brief background to the recent approaches to Semantic Segmentation and then delve into Attention mechanism. Then we follow this with details about Criss-Cross attention. Section III describes the methods including our proposed Network architecture and details of our proposed differentiable attention sampling. We then present the experimental setup and share the results in Section IV. Finally we conclude the paper with a section on Discussion (Section V) and a few potential extension and ideas for future work in the Conclusion section (Section VI)

## 2. Related work

We discuss the areas of research that are related to our work - semantic segmentation and attention mechanisms as well as deep learning based methods for processing of COVID-19 images.

### 2.1. Semantic segmentation and attention mechanisms

Deep learning based algorithms are able to automatically segment images when trained on manually segmented lesion labels. Semantic segmentation has steadily progressed in the last few years evolving from Fully Convolutional Network (FCN) [20], to the use of dilated convolutions [21] and extensive adaptation of encoder decoder architectures - U-Net [22], Attention U-Net [19] [1], nnU-Net [23], DeepLabv3+ [24], Semantic Prediction Guidance (SPGNet) [25], Discriminative Feature Network (DFN) [26], RefineNet [27] and Multi- Scale Context Intertwining (MSCI) [28].

To detect objects within images of various scale, the convolution operator has been enhanced using Deformable Convolution [29] [1] and Scale adaptive convolutions [30]. Graphical models have also been employed effectively for the task of semantic segmentation [31] [21].

Attention models initially gathered a lot of traction after the successful introduction of transformer models in Natural Language Processing (NLP) domain [32]. It has been demonstrated that NLP models perform better when the encoder and decoder are connected through attention blocks.

Attention mechanism have subsequently been utilized in computer vision tasks to capture long-range dependencies. The earlier approaches have tried to augment convolutional models with content-based interactions [24] [33] [1]. The seminal work in attention mechanisms was non-local means [34], which was then followed by self-attention [33]. These have helped achieve better performance on computer vision tasks like image classification and semantic segmentation. Attention-gates have also shown promising results when incorporated into U-Nets for 3D medical segmentation [1]. There have also been successful experiments of building pure self-attention vision models [35].

Non-Local Networks [34] enable full-image context information by utilizing self-attention which helps reference features from any position to perceive the features of all other positions. The drawback of Non-Local network is the large time and space complexity ($\mathcal{O}(H\times W)\times(H\times W)$) to measure every pixel-pair relation, and also requiring large GPU memory to train such models.

CCNet [18] elegantly solves the complexity issue by using consecutive sparse attention. With two criss-cross attention modules, CCNet captures contextual information from all pixels with far less time and space complexity. Criss-cross attentiom (CCNet) [18] was shown to enable improvements in computer vision semantic segmentation tasks on Cityscapes, ADE20K datasets. Tang et al. [36] have successfully employed criss cross attention in medical organ segmentation (lung segmentation). In their XLSor paper [36] they used a pretrained ResNet101 replacing the last two down-sampling layers with dilated convolution operation.

### 2.2. Deep learning based processing of COVID-19 images

Several researchers have already established the efficacy of deep learning based algorithms for processing of COVID-19 images.

For classifying COVID-19 from healthy, a variant of inception network was proposed by Wang [15]. An U-Net++ architecture [37] has been effectively put to use for COVID-19 diagnosis, which worked better than expert radiologists. Mahmud et al. [38] use a large database containing X-rays from normal and other non-COVID pneumonia patients for transfer learning and are able to distinguish between normal, COVID-19, viral, and bacterial pneumonias. Their proposed CovXNet includes depthwise convolution with varying dilation rates for efficiently extracting diversified features from chest X-rays.

The COVID-19-20 MICCAI challenge [39] currently evaluates many different methods for the segmentation and quantification of lung lesions caused by COVID-19 from CT images. The challenge winners are not announced yet, but based on the leaderboard, it is possible to infer many successful methods. One of the early works for semantic segmentation on COVID-19 images was DenseUNet proposed by Chaganti et al. [40] to segment the lesions, lungs and lobes in 3D. They compute percentage of opacity and lung severity scores and report this on entire lung and lobe-wise. The algorithm was trained on 613 manually delineated CT scans (160 COVID-19, 172 viral pneumonia, and 296 ILD). They report Pearson Correlation coefficients between prediction and ground truth above 0.95 for all the four categories. In CovidENet [41] a combination of a 2D slice-based and 3D patch-based ensemble architectures is proposed, trained on 23423 slices. Their finding was that CovidENet performed equally well as trained radiologists, with a Dice coefficient of 0.7. In the realm of segmentation based methods, [42,43] use VB-net [44] to segment of lung and infection regions in CT images. In [45], the focus is on speed-up of segmentation while maintaining comparable accuracy to state-of-the-art models using a three-stage framework called KISEG for multi-level accelerating semantic segmentation of image series of CT.

Gao et al. [46] introduced a dual-branch segmentation-classification framework that simultaneously performs COVID-19 diagnosis and the segmentation of lesions based on chest CT images. They proposed a lesion attention module to utilize the intermediate results of both segmentation and classification branches to improve the classification performance.

## 3. Methods

In this section we explain the details of the proposed network architecture. We also capture the major differences between the existing work and our proposed approach. Our basic idea is integrating an attention module within the U-Net architecture [1] as an extension of the U-Net's bottleneck in order to capture contextual information from only the necessary and meaningful non-local contextual information in smart and efficient way. Our models utilize the approach of the criss-cross attention module proposed by [18] and modify it to enhance the segmentation performance on COVID-19 datasets. The aim of our work is twofold. First, we evaluate whether criss-cross attention can be employed within a U-Net [1] to improve medical image lesion segmentation for labelled data which is relatively small, a common scenario currently for COVID-19. Second, we incorporate our novel adaptation of this attention model and extend it with a dynamic deformable attention mechanism where the attention filter offsets are learnt in a continuous differentiable space. We strongly believe that the deformable attention module that automatically adapt their layout is an important step to get better insight into the computation mechanism of attention modules. We have discovered in our work that capturing attention from all non-local locations does negatively impact the accuracy of semantic segmentation networks. Capturing only the necessary and essential non-local contextual information in a smart and data driven way yields far more promising segmentation results. We also demonstrate that having the attention offsets learnable enables the network to smartly decide on its own the locations where to obtain non-local attention from for improved results.

### 3.1. Criss-cross attention module

The criss-cross attention module (CCA) proposed by Huang et al. [18] aggregates contextual information in horizontal and vertical directions for each pixel. The input image $\mathbf{X}$ is passed through convolutional neural network (CNN) to generate the feature maps $\mathbf{H}$ of reduced dimension. The CCA module comprises of three convolutional layers applied on $\mathbf{H}\in\mathbb{R}^{C\times H\times W}$ with $1\times 1$ as kernel size.

First, the local representation feature maps $\mathbf{H}$ are fed into two convolutional layers in order to obtain two feature maps - query $\mathbf{Q}$ and key $\mathbf{K}$ with the same reduced number of feature channels $C^{'}$. By extracting feature vectors at each position $u$ from $\mathbf{Q}$, a vector $\mathbf{Q_u} \in \mathbb{R}^{C'}$ is generated. From $\mathbf{K}$ feature vectors in the same row and column as $u$ are collected in $\Omega_{\mathbf{u}} \in \mathbb{R}^{(H+W-1)\times C^{'}}$ with elements $\Omega_{\mathbf{i,u}} \in \mathbb{R}^{C'}$.

Attention maps $\mathbf{A} \in \mathbb{R}^{(H+W-1)\times H\times W}$ are obtained by applying the affinity operation $d_{i,u} = \mathbf{Q_u}\Omega_u^T$ with $d_{i,u} \in \mathbf{D}$ being the degree of correlation between feature $\mathbf{Q_u}$ and $\Omega_{\mathbf{i,u}}, i = [1, \ldots, |\Omega_{\mathbf{u}}|], \mathbf{D} \in \mathbb{R}^{(H+W-1)\times H\times W}$ followed by a softmax layer on $\mathbf{D}$ over the channel dimension.

The third convolutional layer applied on $\mathbf{H}$ generates value $\mathbf{V} \in \mathbb{R}^{C\times H\times W}$ for feature adaption. Therefore, a feature vector $\mathbf{V_u} \in \mathbb{R}^{C}$ and a set $\Phi_{\mathbf{u}} \in \mathbb{R}^{(H+W-1)\times C}$ are extracted at each position $u$ in the spatial dimension of $\mathbf{V}$.

The contextual information is aggregated by

$$\mathbf{H_u^{'}} = \sum_{\mathbf{i}\in|\Phi_{\mathbf{u}}|} \mathbf{A_{i,u}}\Phi_{\mathbf{i,u}} + \mathbf{H_u} \tag{1}$$

with $\mathbf{H_u^{'}}$ being a feature vector in the module's output feature maps $\mathbf{H}^{'} \in \mathbb{R}^{C\times H\times W}$ at position $u$ and $\mathbf{A_i}, \mathbf{u}$ being a scalar value at channel $i$ and position $u$ in $\mathbf{A}$. Finally, the contextual information is weighted with a learnable scalar $\gamma$ and added to the feature map $\mathbf{H}$.

### 3.2. Network architecture

The architecture of our model combines the concepts of U-Net [22] and CCNet [18]. A block diagram of the proposed Deformable Attention Net (DDANet) is shown in Fig. 3.

We use a U-Net structure from Oktay et al. [19] [1], adapting it slightly by reducing one downsampling (and corresponding upsampling path), to best process our image dimension (256*256). It consists of three blocks in the downsampling path and three blocks in the upsampling block. Each block consists of 2 × (Batch Normalization - 2D Convolution (kernel size 3 × 3, stride 1, padding 1) - ReLU). The last block consists of a 2D convolution with kernel size 1 × 1. For downsampling, max pooling is applied in the downsampling path to halve the spatial dimension of the feature maps after each block. In the upsampling path ConvTranspose2d is used to double the size of the spatial dimension of the concatenated feature maps. The number of feature channels is increased 1 −64 −128 −256 −512 in the downsampling path and decreased again accordingly in the upsampling path. The U-Net's last layer outputs a number of feature channels matching the number of label classes for semantic segmentation.

The local representation feature maps $\mathbf{H}$ being output from the U-Net's last block within the downsampling path serve as input of reduced dimension to the criss-cross module. The attention module is inserted in the bottleneck, as the feature maps are of reduced dimension, and hence

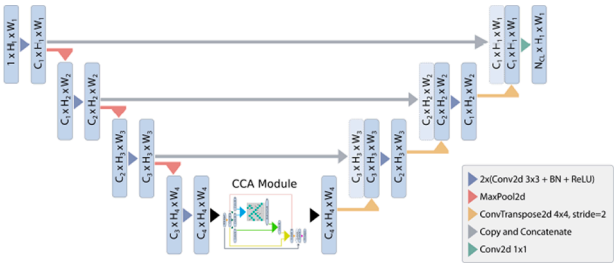


**Fig. 3.** A block diagram of the proposed Deformable Attention Net (DDANet). Input image is progressively filtered and downsampled by factor 2 at each scale in the encoding part. The deformable criss-cross attention is inserted as an extension of the U-Net's bottleneck in order to capture contextual information from only the necessary and meaningful non-local contextual information in smart and efficient way.

the attention maps have smaller, more manageable time and space complexity. In the orginial CCNet [18], the following attention module gathers contextual information in the criss-cross path of each pixel leading to feature maps $\mathbf{H}^{'}$.

The major difference of our approach to the existing approaches of using criss-cross attention [18,36] is the efficient and effective methodology to capture non-local interactions. Earlier criss-cross attention approaches [18,36] capture attention from all non-local locations, which negatively impact the accuracy of semantic segmentation networks. Our approach captures only the necessary and essential non-local contextual information in a smart and data driven way using differentiable attention sampling introduced later in this section, which yields superior segmentation results. Our experiments validate our hypothesis that having the attention offsets learnable enables the network to sparsely and smartly decide on its own the locations where to obtain non-local attention from thereby yielding improved results.

In our proposed novel DDANet, the pattern is dynamic and learnable, and hence a dynamic deformable criss-cross path is used to obtain the attention feature maps $\mathbf{DH}^{'}$. These feature maps are again passed through the dynamic deformable attention module again which results in feature maps $\mathbf{DH}^{''}$ capturing attention information from the most relavant locations from the whole-image at each of its positions. The contextual features $\mathbf{DH}^{''}$ obtained after passing $R = 2$ loops through the attention module are concatenated with the feature maps $\mathbf{X}$ and merged by a convolutional layer. The resulting feature maps are then passed through the U-Net's upsampling path.

We implement the following modifications of the criss-cross attention module: Deformable CCA module with $R = 2$ loops, $\mathbf{X} + \gamma\mathbf{DH}^{''}$.

**Differentiable attention sampling**: Consider a classical criss-cross attention operation which gathers non-local information on a feature map of Height $H$ and width $W$. The initial shape of the criss-cross pattern is a cross as the orginial CCNet [18] which aggregates contextual information for each pixel in its criss-cross path. We have realized the baseline criss-cross attention by first initializing statically defined locations in a 2D flow field (sampling grid) of size $H \times W$. The attention filter offsets for the vertical direction are defined as the locations where the $x$ coordinates match a tensor of length $H$ of equally spaced points between $-1$ and 1. Similarly, the attention filter offsets for the horizontal direction are defined as the locations where the $y$ coordinates match a tensor of length $W$ of equally spaced points between $-1$ and 1. These vertical and horizontal offsets help to compute the attention along a cross pattern at $\mathbf{H}+\mathbf{W}$ non-local locations.

To make the attention map differentiable, we compute the displacements for the horizontal and vertical offsets. For computing the displacement for each of the horizontal and vertical locations we use $H + W$ random locations sampled from a standard normal distribution. We distribute these displacement locations smoothly by convolving them three times with a Gaussian kernel with a kernel size 5. We then use a spatial transformer network to sample the attention values from the offset locations coupled with the displacements. To obtain the attention output for inputs on a discrete grid, we use differentiable bilinear interpolation. This makes our attention sampling differentiable and the attention locations are dynamic and deformable.

We realized our dynamic deformable attention mechanism by the differentiable attention sampling described above which deforms the criss-cross pattern. In our deformable attention implementation, we have included $\mathbf{H}+\mathbf{W}$ learnable attention offset parameters in our deep neural network definition. These are the learnt displacements for each of the criss-cross locations. The learnt displacement vector ($\mathbf{x}$ and $\mathbf{y}$ displacement) for each of the criss-cross locations is used to displace the horizontal and veritcal offsets, while sampling the attention maps. For the second recurrence, a second set of different $\mathbf{H}+\mathbf{W}$ learnable attention parameters is used for determining the displacements.

We use differentiable bilinear interpolation to differentiably sample the attention values for the query, key and value feature maps from the

deformed and dynamically learnt positions of criss-cross offset locations. Hence the attention filter offsets for each of the original criss-cross pattern are learnt in continuous differentiable space. The proposed deformable criss-cross attention is depicted in the CCA-Module in Fig. 4. As depicted in the figure, the criss-cross pattern is learnt and dynamically deformed to best capture the most relevant non-local information. (See 5).

The infection class in COVID-19 data is generally under represented as compared to the background class, especially in early stages of the disease. This leads to a large class imbalance problem. As found in several studies, Ground-glass opacities generally precede consolidations lesions. This progression of the lesion development in COVID-19 leads to the another scenario of class-imbalance. In some patients only one of the lesions is largely present and the second lesion is highly under-represented (less than 10% of the total infection labels). This also leads to a second category of class-imbalance. To address all of these class-imbalance issues, especially present in COVID-19 lesion segmentation scenarios, we propose to use the inverse class-weighted cross-entropy loss. The weights are computed to be inversely proportional to the square root of class frequency. Given a sample with class label $y$, this inverse class-weighted cross-entropy loss can be expressed as

$$\mathrm{CE}(z, y) = w_y \left( -\log\left( \frac{\exp(z_y)}{\sum_{j=1}^{C} \exp(z_j)} \right) \right) \tag{2}$$

with $C$ being the total number of classes and $z$ the output from the model for all classes. The weighting factor

$$w_y = \frac{\sqrt{\frac{1}{z_y}}}{\frac{1}{C}\sum_{j=1}^{C} \sqrt{\frac{1}{z_j}}} \tag{3}$$

is determined with help of the inverse square root of the number of samples in each label class to address the problem of training from imbalanced data. The training and validation sets also have different distributions, hence we have computed the inverse weighting separately for the train and validation sets. We have also used learning rate finder [47] to find the optimal learning rate, and a 1cycle learning rate policy scheduler, where the maximum learning rate was also determined using the learning rate finder.

## 4. Experimental setup and results

We have used the publicly available COVID-19 CT segmentation dataset [16]. We have taken the 100 axial CT images from different COVID-19 patients. This first collection of data is from the Italian Society of Medical and Interventional Radiology. We have also utilized the second dataset of axial volumetric CTs of nine patients from Radiopaedia [16]. This second dataset with whole volumes includes slices which have COVID-19 lesions (373 positive slices) and slices without COVID-19 lesions (455 negative slices). We perform experiments with a 5-fold cross validation on this combined dataset consisting of 471 two-dimensional axial lung CT images with segmentations for Ground-glass opacities (GGO) and consolidation lesions. Each fold comprises data acquired from multiple patients plus one third of images from the 100 slice CT stack taken from more than 40 different patients. The CT images are cropped and rescaled to a size of $256 \times 256$. During training, we perform random affine deformations for data augmentation.

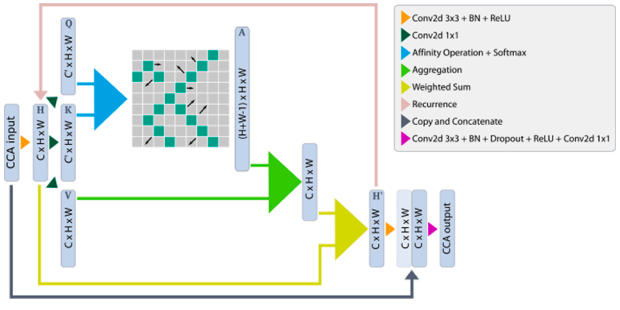


**Fig. 4.** A block diagram of the proposed deformable criss-cross attention module. In our deformable criss-cross, we have the $\mathbf{H}+\mathbf{W}-\mathbf{1}$ learnable attention offset parameters for each of the criss-cross locations. Differentiable bilinear interpolation is used to sample the attention values for the query, key and value feature maps from the learnt positions of deformed criss-cross offset locations.

Training is performed for 500 epochs using the Adam optimizer and an initial learning rate of 0.002. We further use a cyclic learning rate with an upper boundary of 0.005 and a class-weighted cross-entropy loss to address the problem of training from imbalanced data.

For the infection region experiments and multi-class labeling we compared our model with two cutting-edge models: U-Net [19] and Criss-Cross Attention [18]. The number of trainable parameters for the U-Net [19] is 611 K. For the U-Net incorporated with the criss cross attention the parameter count is 847 K. Our proposed variant of modified CCNet has slightly more parameters at 849 K. We have used four widely adopted metrics, *i.e.*, Dice similarity coefficient, Sensitivity (Sen.), Specificity (Spec.) and Mean Absolute Error *(MAE)*. If we denote the final prediction as $\mathbf{F_p}$ and the object-level segmentation ground-truth as $\mathbf{G}$, then the Mean Absolute Error which measures the pixel-wise error between final prediction and ground truth is defined as

$$\mathrm{MAE} = \frac{1}{w \times h} \sum_{x}^{w} \sum_{y}^{h} |F_p(x, y) - G(x, y))| \tag{4}$$

We have adopted a similar approach to Fan et al. [2] and present first the results of our proposed DDANet on detecting lung infections. Our network is trained on multi-class lung infection (GGO and consolidation) and during evaluation we combine these multiple classes into one infection label. We present our 5-fold cross-validation studies results in Table 1 which is averaged over multiple runs that we have conducted. We have also included the results from Fan et al. [2] in each of our experiments. It has to be noted that Inf-Net was only trained with the first dataset which is smaller (100 axial slices) and Semi-Inf-Net was trained with pseudo labels from unlabelled CT images. As captured in the Table 1, our proposed DDANet achieves the best Dice scores in each of the folds. The best Dice score obtained is **0.849** and least mean absolute error (MAE) is **0.0163**. We have also captured the average infection segmentation performance of our network in the same Table 1. Our proposed DDANet has the best infection segmentation performance in average with the average Dice score of **0.781**). In terms of Dice, our proposed DDANet out-performs the cutting-edge U-Net model [19] by **1.56%** on average infection segmentation.

We have also included the infection segmentation performance of our DDANet on each of the patients in the supplementary materials. For each of the patients, our proposed DDANet had the best Dice score and the minimum MAE. The average across all the patients is captured in Table 2. In terms of Dice, our DDANet method achieves the best competitive performance of **0.7789** averaged across all the patients. It outperforms the baseline best U-Net model Dice by **3.658**% on infection segmentation.

The fold-wise performance of our DDANet on multi-class labeling is included in the supplement section. We have captured the average multi-label segmentation performance of our network in Table 3. We have also compared our results with the results from Inf-Net by Fan et al. [2]. Our baseline U-Net [19] and proposed DDANet has far less trainable parameters at (**611K**) and (**849K**) as compared to **33M** in Inf-Net [2]. Our proposed DDANet has the best multi-label segmentation performance also in average with the best Dice score of **0.734**) for GGO lesions and best Dice score of **0.613**) for consolidation lesions. Our proposed DDANet has average best dice score of **0.673** for detecting COVID-19

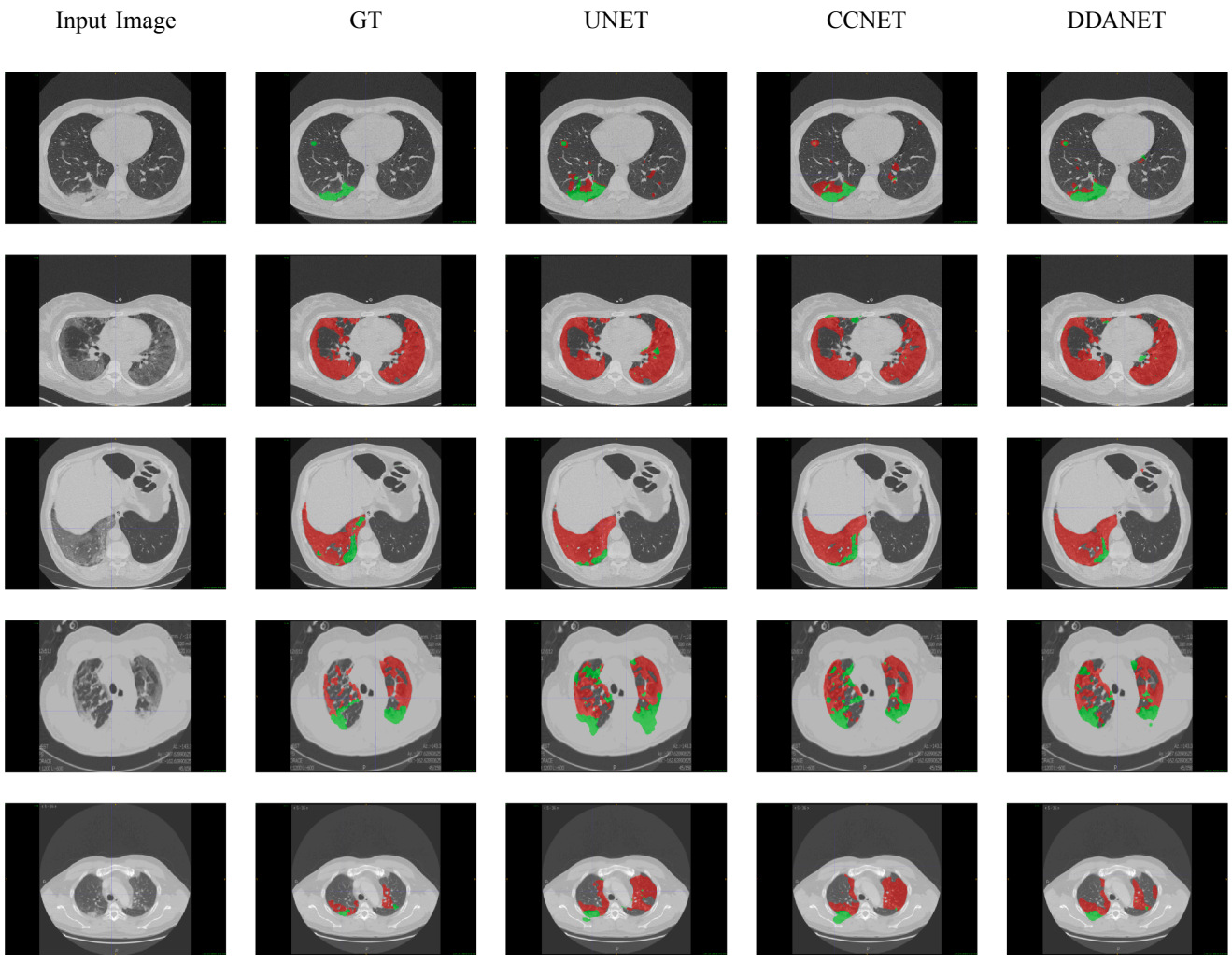


**Fig. 5.** Visual comparison of multi-class lung segmentation results, where the red and green labels indicate the GGO and Consolidation, respectively. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

lesions. In terms of Dice, our proposed DDANet out-performs the cutting-edge U-Net model [19] by **4.90**% on average multi-label segmentation. We have increased the trainable parameters in our proposed DDANet only by a negligible amount of 2450 (or 0.3%) in comparison to the original model with criss-cross attention.

We have also captured the multi-label segmentation performance of our DDANet on each of the Patients in the supplementary materials. In terms of Dice, our DDANet method achieves the best competitive performance of **0.702** for GGO lesion and **0.681** for consolidation lesion averaged across all the patients. In average the proposed DDANet outperforms the baseline best U-Net model Dice by **2.86**% on GGO, **4.73**% on consolidation and in average **3.52**% on multi-label segmentation. The distribution of the GGO and consolidation lesions are not even among the different patient scans. Some patients had predominantly only GGO (Patient-8) while other patients had predominantly consolidation (Patient-3). This skew in distribution impacts the segmentation dice scores significantly, when the lesions are minimally represented in the patients.

## 5. Discussion

COVID-19 lesion segmentation is a very challenging problem. One of the major challenge is the regional manifestation of lesions especially in the early stages of the disease, and it can be very hard to get good segmentation in those high class-imbalance scenarios. A similar challenge arises when one of the lesion classes is majorly represented and the other class is highly under-represented which makes it very difficult. This also is a challenging scenario of skewed class-imbalance and gets very hard to get good segmentation in this context as well. The third challenge is the very limited availability of large public datasets, which has been the case until recently. Slowly a number of COVID-19 datasets are made publicly available and this scenario could change quite dramatically in the future. This would then enable further research into more compelling algorithms to address this challenging problem.

Our proposed deformable attention is only one of the potential ways to realize learnable attention mechanisms that are smarter, more elegant, and have better performance than earlier proposed criss-cross attention or non-local methods. There are lots of research possibilities to make this even better. There is no requirement or limitation to gather attention from $\mathbf{H}+\mathbf{W}$ locations as we are currently computing. We have currently computed it that way to make it comparable to criss-cross attention. The attention could be gathered from lesser or more locations. One of the next research problems could be to explore what could be the optimal or minimal number of non-local attention locations that needs to be gathered to get the best results. It would also be interesting to establish theoretical upper and lower bounds for number of locations to get non-local attention and its impact on performance. Our work opens up all these and more possible research directions and can be the trigger for more fundamental work on learnable attention mechanisms.

**Table 1**
Performance (averaged) of Infection regions on COVID-19 datasets. We have split our data into five folds, and the results here are averaged over multiple runs for each fold. These are quantitative results of infection regions computed fold-wise and we report 3D Dice-Scores.

PERFORMANCE (AVERAGED) OF INFECTION REGIONS ON COVID-19 DATASETS. WE HAVE SPLIT OUR DATA INTO FIVE FOLDS, AND THE RESULTS HERE ARE AVERAGED OVER MULTIPLE RUNS FOR EACH FOLD. THESE ARE QUANTITATIVE RESULTS OF INFECTION REGIONS COMPUTED FOLD-WISE AND WE REPORT 3D DICE-SCORES

| Model | Dice (Fold) | Dice (Average) | MAE |
|---|---|---|---|
| Inf-Net [2] | | 0.682 | 0.082 |
| Semi-Inf-Net [2] | | 0.739 | 0.064 |
| UNET | 0.716 | 0.769 | 0.0227 |
| | 0.792 | | 0.0301 |
| | 0.738 | | 0.0273 |
| | 0.833 | | 0.0360 |
| | 0.766 | | 0.0182 |
| +CCA | 0.737 | 0.768 | 0.0207 |
| | 0.790 | | 0.0302 |
| | 0.728 | | 0.0292 |
| | 0.843 | | 0.0305 |
| | 0.741 | | 0.0207 |
| **DDANet** | **0.74** | **0.781** | **0.0206** |
| | **0.794** | | 0.0305 |
| | **0.745** | | **0.0262** |
| | **0.849** | | **0.0299** |
| | **0.776** | | **0.0163** |

**Table 2**
Performance (averaged) on Nine real CT patient data. These are quantitative results of infection regions computed patient-wise and we report 3D Dice-Scores. The best results are shown in Blue font and the Gain with respect to baseline UNet is shown in Green.

PERFORMANCE (AVERAGED) ON NINE REAL CT PATIENT DATA. THESE ARE QUANTITATIVE RESULTS OF INFECTION REGIONS COMPUTED PATIENT-WISE AND WE REPORT 3D DICE-SCORES. THE BEST RESULTS ARE SHOWN IN BLUE FONT AND THE GAIN WITH RESPECT TO BASELINE UNET IS SHOWN IN GREEN.

| Model | Dice | Sen. | Spec. | MAE | % Gain |
|---|---|---|---|---|---|
| Inf-Net [2] | 0.579 | 0.87 | 0.974 | 0.047 | |
| Semi-Inf-Net [2] | 0.597 | 0.865 | 0.977 | 0.033 | |
| UNET | 0.7515 | 0.8811 | 0.9904 | 0.0149 | |
| +CCA | 0.7633 | **0.8934** | 0.9908 | 0.0143 | **1.5819** |
| **DDANet** | **0.7789** | 0.8840 | **0.9915** | **0.0135** | **3.658** |

## 6. Conclusion

In this paper, we have proposed a novel adaptation to the criss-cross attention module with deformable criss-cross attention. This has been incorporated into the U-Net framework (DDANet) to improve the segmentation of lesion regions in COVID-19 CT scans. Our extensive experiments have demonstrated that both adapting the U-Net with a straightforward incorporation of the CCNet module and also extending this CCNet with multiple recurrent application does not yield substantial improvements in segmentation quality. Our novel solution and smart combination of adapted dynamic deformable spatial attention have shown to be a working combination yielding superior and promising results. This solution has immense potential in better aiding clinicians with state-of-art infection segmentation models.

Our proposed approach works well on medical data, specifically to segment small lesion structures like the COVID lesion segmentation example in this article. It should be highlighted that our technique and concept is generic and could very well generalize to other medical data segmentation problems and potentially to other semantic segmentation problems in computer vision. This would need further investigation, experiments and validation. This could be an interesting area for future work. Our proposed architecture can also be trained with less number of data, as we have used the UNet architecture of Oktay et al. [19], which has only 611 K parameters and after integrating our Deformable attention the modified network has only 850 K parameters.

For our future studies, we plan to explore its adaptation in ResNet like architectures for 2D and once more labelled 3D scans become available the module can easily be adapted to 3D V-Net architectures. We will make our source-code and trained models publicly available.

**Table 3**
Quantitative results of Ground-Glass opacities and consolidation. The results are averaged across multiple folds and multiple runs. The best results are shown in Blue font.

QUANTITATIVE RESULTS OF GROUND-GLASS OPACITIES AND CONSOLIDATION. THE RESULTS ARE AVERAGED ACROSS MULTIPLE FOLDS AND MULTIPLE RUNS. THE BEST RESULTS ARE SHOWN IN BLUE FONT.

| Model | GGO | Consol. | Avg | %Gain | #Params |
|---|---|---|---|---|---|
| Semi-Inf-Net+FCN8s | 0.646 | 0.301 | 0.474 | | 33.1M |
| Semi-Inf-Net+MC | 0.624 | 0.458 | 0.541 | | 33.1M |
| UNet | 0.717 | 0.566 | 0.641 | | 611.7 K |
| +CCA | 0.723 | 0.596 | 0.660 | **2.84** | 847.3K |
| **DDANet** | **0.734** | **0.613** | **0.673** | **4.90** | 849.7K |

## CRediT authorship contribution statement

**Kumar T. Rajamani:** Conceptualization, Methodology, Software, Writing - original draft. **Hanna Siebert:** Data curation, Writing - review & editing. **Mattias P. Heinrich:** Writing - review & editing, Supervision.

## Declaration of Competing Interest

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

## Appendix A. Supplementary material

Supplementary data associated with this article can be found, in the online version, at https://doi.org/10.1016/j.jbi.2021.103816.

## References

[1] 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. (2019) 197–207.

[2] D. Fan, T. Zhou, G. 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.

[3] worldometers.info, Covid-19 coronavirus pandemic. [Online]. Available: https://www.worldometers.info/coronavirus/.

[4] C. Menni, A.M. Valdes, M.B. Freidin, et al., Real-time tracking of self-reported symptoms to predict potential covid-19, Nat. Med. (2020).

[5] 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 (2020).

[6] 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 in coronavirus disease 2019 (covid-19) in china: A report of 1014 cases, Radiology (2020).

[7] H.Y.F. Wong, H.Y.S. Lam, A.H.-T. Fong, S.T. Leung, T.W.-Y. Chin, C.S.Y. Lo, M.M.-S. Lui, J.C.Y. Lee, K.W.-H. Chiu, T. Chung, E.Y.P. Lee, E.Y.F. Wan, F.N.I. Hung, T.P. W. Lam, M. Kuo, M.-Y. Ng, Frequency and distribution of chest radiographic findings in covid-19 positive patients, Radiology (2020).

[8] M.-Y. Ng, E.Y. Lee, J. Yang, F. Yang, X. Li, H. Wang, M.M.-s. Lui, C.S.-Y. Lo, B. Leung, P.-L. Khong, C.K.-M. Hui, K.-y. Yuen, M.D. Kuo, Imaging profile of the covid-19 infection: Radiologic findings and literature review, Radiol. Cardiothoracic Imaging (2020).

[9] icometrix.com, https://icometrix.com/resources/the-role-of-imaging-ai-and-ct-in-covid-19.

[10] M.-Y. Ng, E.Y. Lee, J. Yang, F. Yang, X. Li, H. Wang, M.M.-S. Lui, C.S.-Y. Lo, B. Leung, P.-L. Khong, C.K.-M. Hui, K.-y. Yuen, M.D. Kuo, Imaging profile of the covid-19 infection: Radiologic findings and literature review, Radiol. Cardiothoracic Imaging (2020).

[11] 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. (2020).

[12] Y. Oh, S. Park, J.C. Ye, Deep learning covid-19 features on cxr using limited training data sets, IEEE Trans. Med. Imaging (2020) 1.

[13] D. Dong, Z. Tang, S. Wang, H. Hui, L. Gong, Y. Lu, Z. Xue, H. Liao, F. Chen, F. Yang, R. Jin, K. Wang, Z. Liu, J. Wei, W. Mu, H. Zhang, J. Jiang, J. Tian, H. Li, The role of imaging in the detection and management of covid-19: a review, IEEE Rev. Biomed. Eng. (2020) 1.

[14] J. Chen, L. Wu, J. Zhang, L. Zhang, D. Gong, Y. Zhao, S. Hu, Y. Wang, X. Hu, B. Zheng, K. Zhang, H. Wu, Z. Dong, Y. Xu, Y. Zhu, X. Chen, L. Yu, H. Yu, Deep learning-based model for detecting 2019 novel coronavirus pneumonia on high-resolution computed tomography: a prospective study, medRxiv (2020).

[15] S. Wang, B. Kang, J. Ma, X. Zeng, M. Xiao, J. Guo, M. Cai, J. Yang, Y. Li, X. Meng, B. Xu, A deep learning algorithm using ct images to screen for corona virus disease (covid-19) (2020).

[16] MedicalSegmentation.com, Covid-19 ct segmentation dataset. [Online]. Available: http://medicalsegmentation.com/covid19/.

[17] M.P. Heinrich, O. Oktay, N. Bouteldja, Obelisk-net: Fewer layers to solve 3d multi-organ segmentation with sparse deformable convolutions, Med. Image Anal. (2019).

[18] Z. Huang, X. Wang, L. Huang, C. Huang, Y. Wei, W. Liu, Ccnet: Criss-cross attention for semantic segmentation, in: 2019 IEEE/CVF International Conference on Computer Vision (ICCV), 2019.

[19] O. Oktay, J. Schlemper, L.L. Folgoc, M.C.H. Lee, M.P. Heinrich, K. Misawa, K. Mori, S.G. McDonagh, N.Y. Hammerla, B. Kainz, B. Glocker, D. Rueckert, Attention u-net: Learning where to look for the pancreas, CoRR, vol. abs/1804.03999, 2018. [Online]. Available: http://arxiv.org/abs/1804.03999.

[20] J. Long, E. Shelhamer, T. Darrell, Fully convolutional networks for semantic segmentation, in: 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2015.

[21] F. Yu, V. Koltun, Multi-scale context aggregation by dilated convolutions, in: International Conference on Learning Representations (ICLR), 2016.

[22] O. Ronneberger, P. Fischer, T. Brox, U-net: Convolutional networks for biomedical image segmentation, in: Medical Image Computing and Computer-Assisted Intervention (MICCAI), 2015.

[23] F. Isensee, J. Petersen, A. Klein, D. Zimmerer, P.F. Jaeger, S. Kohl, J. Wasserthal, G. Koehler, T. Norajitra, S. Wirkert, K.H. Maier-Hein, Abstract: nnu-net: Self-adapting framework for u-net-based medical image segmentation, in: H. Handels, T.M. Deserno, A. Maier, K.H. Maier-Hein, C. Palm, T. Tolxdorff (Eds.), Bildverarbeitung für die Medizin 2019, Springer Fachmedien Wiesbaden, Wiesbaden, 2019, p. 22.

[24] Y. Chen, Y. Kalantidis, J. Li, S. Yan, J. Feng, A2-nets: Double attention networks, in: Advances in Neural Information Processing Systems 31, 2018.

[25] B. Cheng, L.-C. Chen, Y. Wei, Y. Zhu, Z. Huang, J. Xiong, T.S. Huang, W.-M. Hwu, H. Shi, "Spgnet: Semantic prediction guidance for scene parsing, in: The IEEE International Conference on Computer Vision (ICCV), October 2019.

[26] C. Yu, J. Wang, C. Peng, C. Gao, G. Yu, N. Sang, Learning a discriminative feature network for semantic segmentation, in: The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June 2018.

[27] G. Lin, A. Milan, C. Shen, I.D. Reid, Refinenet: Multi-path refinement networks for high-resolution semantic segmentation, in: 2017 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017, Honolulu, HI, USA, July 21–26, 2017. IEEE Computer Society, 2017, pp. 5168–5177. [Online]. Available: doi: 10.1109/CVPR.2017.549.

[28] D. Lin, Y. Ji, D. Lischinski, D. Cohen-Or, H. Huang, Multi-scale context intertwining for semantic segmentation, in: The European Conference on Computer Vision (ECCV), September 2018.

[29] J. Dai, H. Qi, Y. Xiong, Y. Li, G. Zhang, H. Hu, Y. Wei, Deformable convolutional networks, in: 2017 IEEE International Conference on Computer Vision (ICCV), 2017, pp. 764–773.

[30] R. Zhang, S. Tang, Y. Zhang, J. Li, S. Yan, Scale-adaptive convolutions for scene parsing, in: The IEEE International Conference on Computer Vision (ICCV), Oct 2017.

[31] Z. Liu, X. Li, P. Luo, C.-C. Loy, X. Tang, Semantic image segmentation via deep parsing network, in: Proceedings of the 2015 IEEE International Conference on Computer Vision (ICCV), IEEE Computer Society, 2015.

[32] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A.N. Gomez, u. Kaiser, I. Polosukhin, Attention is all you need, in: Proceedings of the 31st International Conference on Neural Information Processing Systems, ser. NIPS'17, Curran Associates Inc., Red Hook, NY, USA, 2017, pp. 6000–6010.

[33] I. Bello, B. Zoph, A. Vaswani, J. Shlens, Q.V. Le, Attention augmented convolutional networks, in: The IEEE International Conference on Computer Vision (ICCV), October 2019.

[34] X. Wang, R.B. Girshick, A. Gupta, K. He, Non-local neural networks, in: 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2018, pp. 7794–7803.

[35] N. Parmar, P. Ramachandran, A. Vaswani, I. Bello, A. Levskaya, J. Shlens, Stand-alone self-attention in vision models, in: Advances in Neural Information Processing Systems 32, 2019.

[36] Y. Tang, Y. Tang, J. Xiao, and R. Summers, "Xlsor: A robust and accurate lung segmentor on chest x-rays using criss-cross attention and customized radiorealistic abnormalities generation," in Medical Imaging with Deep Learning, 04 2019.

[37] Z. Zhou, M.M. Rahman Siddiquee, N. Tajbakhsh, J. Liang, Unet++: A nested u-net architecture for medical image segmentation, in: D. Stoyanov, Z. Taylor, G. Carneiro, T. Syeda-Mahmood, A. Martel, L. Maier-Hein, J.M.R. Tavares, A. Bradley, J.P. Papa, V. Belagiannis, J.C. Nascimento, Z. Lu, S. Conjeti, M. Moradi, H. Greenspan, A. Madabhushi (Eds.), Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support, Springer International Publishing, Cham, 2018, pp. 3–11.

[38] T. Mahmud, M.A. Rahman, S.A. Fattah, Covxnet: A multi-dilation convolutional neural network for automatic covid-19 and other pneumonia detection from chest x-ray images with transferable multi-receptive feature optimization, Comput. Biol. Med. 122 (2020) 103869.

[39] M.G. Linguraru, R. Sanchez-Jacob, J. Zember, J. Molto, C.T. Diez, H. Roth, D. Yang, N. Rieke, A. Harouni, Z. Xu, W. Li, M. Flores, D. Xu, B. Wood, B. Turkbey, S. Harmon, S. Xu, E. Turkbey, M. Blain, M. Kassin, N. Varble, A. Amalou, COVID-19-20 challenge: COVID-19 Lung CT Lesion Segmentation Challenge, 2020, https://covid-segmentation.grand-challenge.org/.

[40] S. Chaganti, A. Balachandran, et al., Quantification of tomographic patterns associated with covid-19 from chest ct, arxiv, 2020.

[41] G. Chassagnon, M. Vakalopoulou, et al., Ai-driven ct-based quantification, staging and short-term outcome prediction of covid-19 pneumonia, medRxiv, 2020.

[42] Z. Tang, W. Zhao, X. Xie, Z. Zhong, F. Shi, J. Liu, D. Shen, Severity assessment of coronavirus disease 2019 (covid-19) using quantitative features from chest ct images, 2020.

[43] F. Shi, L. Xia, F. Shan, D. Wu, Y. Wei, H. Yuan, H. Jiang, Y. Gao, H. Sui, D. Shen, Large-scale screening of covid-19 from community acquired pneumonia using infection size-aware classification, 2020.

[44] F. Shan, Y. Gao, J. Wang, W. Shi, N. Shi, M. Han, et al, Lung infection quantification of covid-19 in ct images with deep learning, 2020.

[45] X. Liu, K. Wang, K. Wang, T. Chen, K. Zhang, G. Wang, "Kiseg: A three-stage segmentation framework for multi-level acceleration of chest ct scans from covid-19 patients, in: International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer, 2020, pp. 25–34.

[46] K. Gao, J. Su, Z. Jiang, L.-L. Zeng, Z. Feng, H. Shen, P. Rong, X. Xu, J. Qin, Y. Yang, et al., Dual-branch combination network (dcn): Towards accurate diagnosis and lesion segmentation of covid-19 using ct images, Med. Image Analysis 67 (2020) 101836.

[47] L.N. Smith, Cyclical learning rates for training neural networks, in: 2017 IEEE Winter Conference on Applications of Computer Vision (WACV), 2017.