



MLCA2F: Multi-Level Context Attentional Feature Fusion for COVID-19 lesion segmentation from CT scans

Ibtissam Bakkouri¹ · Karim Afdel¹

Received: 17 June 2021 / Revised: 13 May 2022 / Accepted: 16 July 2022 / Published online: 3 August 2022
© The Author(s), under exclusive licence to Springer-Verlag London Ltd., part of Springer Nature 2022

Abstract

In the field of diagnosis and treatment planning of Coronavirus disease 2019 (COVID-19), accurate infected area segmentation is challenging due to the significant variations in the COVID-19 lesion size, shape, and position, boundary ambiguity, as well as complex structure. To bridge these gaps, this study presents a robust deep learning model based on a novel multi-scale contextual information fusion strategy, called Multi-Level Context Attentional Feature Fusion (MLCA2F), which consists of the Multi-Scale Context-Attention Network (MSCA-Net) blocks for segmenting COVID-19 lesions from Computed Tomography (CT) images. Unlike the previous classical deep learning models, the MSCA-Net integrates Multi-Scale Contextual Feature Fusion (MC2F) and Multi-Context Attentional Feature (MCAF) to learn more lesion details and guide the model to estimate the position of the boundary of infected regions, respectively. Practically, extensive experiments are performed on the Kaggle CT dataset to explore the optimal structure of MLCA2F. In comparison with the current state-of-the-art methods, the experiments show that the proposed methodology provides efficient results. Therefore, we can conclude that the MLCA2F framework has the potential to dramatically improve the conventional segmentation methods for assisting clinical decision-making.

Keywords COVID-19 pneumonia · Segmentation · Context attentional features · Multi-level fusion · Contextual information · Multi-scale features

1 Introduction

According to the recent survey [1], COVID-19 is ranked among the largest global public health crisis. Due to the influence of the virus, in the past few months, a high number of mortalities have been triggered over the world [2]. Currently, many recent clinical studies have reported that COVID-19 lesion segmentation from CT images is an important step in the clinical treatment assessment [3,4] and can be beneficial in this context due to the presence of some prominent features [5]. However, finding an accurate method for COVID-19 lesion segmentation from CT images is a very challenging task due to the large variations in the COVID-19 lesion size, shape, and position, boundary ambiguity, as well as complex structure [6,7].

More recently, convolutional neural network (CNN) architectures have also been extended for medical image segmentation [8,9]. They have shown remarkable performance in segmentation tasks of medical images due to high representation power, filter sharing properties, and fast inference. However, research on COVID-19 lesion segmentation using CNN is still relatively scarce and often suffers from false positives. Following the above discussion, the main intention of this paper is to implement a segmentation algorithm for clinical decision-making based on CNN architectures to cope with this pandemic issue. The main contributions of this study can be summarized as follows:

- (1) It introduces a new framework, called MLCA2F, that integrates three fused MSCA-Net units for an effective COVID-19 lesion segmentation with a high focus on small and crucial lesion boundaries.
- (2) It proposes an effective MC2F paradigm to capture more specific and effective contextual information.
- (3) It designs a novel MCAF module that contributes to improving the performance of our segmentation system

✉ Ibtissam Bakkouri
ibtissam.bakkouri@gmail.com

¹ Laboratory of Computer Systems and Vision (LabSIV),
Department of Computer Science, Faculty of Science, Ibn
Zohr University, BP 8106, 80000 Agadir, Morocco

by forcing the MSCA-Net model to focus on COVID-19 lesions in CT images.

- (4) It presents a series of comparative experiments to validate the effectiveness of MLCA2F.

The remainder of this paper is organized as follows: In Sect. 2, we describe the novel methods applied for COVID-19 lesion segmentation. In Sect. 3, we detail the proposed approach. Then, we present the results of experiments realized on the Kaggle CT dataset in Sect. 4 and we discuss in detail the results of our proposed method in Sect. 5. Finally, we conclude this paper in Sect. 6.

2 Related works

Currently, deep learning models have been applied in various image processing tasks including medical imaging fields and showed promising performance [10]. For instance, pixel-wise classification-based CNN methods have achieved state-of-the-art performances in COVID-19 image segmentation [8]. In most cases, this is done by the U-Net algorithm which has been extensively studied over the past few years [11]. Khalifa et al. [12] proposed a standard U-Net architecture for COVID-19 lesions detection in chest CT scans. The same idea has been explored in [13], by introducing a U-Net architecture to realize accurate COVID-19 region localization in CT images. Chen et al. [5] proposed an variant of U-Net, called U-Net++ [14]. Unlike U-Net, U-Net++ consists of an encoder and decoder connecting through a series of nested dense convolutional blocks for detecting COVID-19 lesions on CT scans. Inspired by the success of attention-based CNN approaches in computer vision [8], researchers have proposed various methods to extend the use of the attention mechanism to automatically identify COVID-19-infected areas from chest CT images. Kuchana et al. [6] proposed the attention gates integrated into standard U-Net architecture for semantic segmentation of COVID-19 anomalies present in the CT scans of patients. Additionally, Zhou et al. [7] modeled the spatial and channel attention modules for automatic COVID-19 CT segmentation using U-Net to capture rich contextual relationships. Moreover, COVID-19 lesion segmentation from CT images has been also approached by Fan et al. [3] where the parallel partial decoder combined with a reverse attention module was adopted to establish the relationship between areas and boundary cues. Recently, region-based CNN methods have gained a lot of interest in medical imaging [9]. Mask R-CNN [15], which is considered as an instance of region-based CNNs, represents a significant potential in COVID-19 structure delineation. Ter-Sarkisov [2] performed COVID-19 lesion segmentation in CT images using Mask R-CNN to preserve the true region within the bounding box. Lately, multi-scale CNN-based

approaches are broadly used for medical image segmentation [16] which can capture complementary contextual information [17]. In the COVID-19 lesion segmentation context, multi-scale CNN schemes have currently been addressed by only a few research works. Zheng et al. [4] presented a new multi-scale discriminative approach, called MSD-Net. In the MSD-Net, the authors proposed a pyramid convolution block, channel attention block, and residual refinement block for the accurate segmentation of COVID-19 lesions from CT images into three categories.

As a matter of fact, multi-scale contextual information learning is a crucial factor in effectively recovering accurate boundaries, extracting complex regions, and fulfilling better semantic segmentation performance. However, in the literature, only a restricted number of researches are available using multi-scale CNN for COVID-19 lesion segmentation from CT images. In view of the scarce existing studies on this domain, a detailed plan will be described in a fair amount of detail to capture fine details without losing any spatial information.

3 Proposed approach

As illustrated in Fig. 1, in this section, we first introduce the whole network architecture and then present the detail of each block. The proposed MLCA2F is developed in three levels each containing one MSCA-Net fused with the original input to improve the segmentation performance. Every MSCA-Net block is built upon the six modules of MC2F and MCAF to better delineate the COVID-19 regions. The specific ideas of the proposed framework are described as follows.

3.1 Data pre-processing

Pre-processing of the digital CT data is performed in two stages, including data size normalization and intensity correction to facilitate the application of MLCA2F for COVID-19 lesion segmentation. The resolution of CT images varies from 401×630 to 630×630 , which requires a high cost of computation. To overcome this limitation, it is necessary to rescale the CT images. We first cropped the center area of the lung region and then adopt the pyramidal down-sampling process as indicated in [18] to resize all images to 256×256 pixels. The second pre-processing technique is based on adaptive bi-histogram equalization [19] algorithm to boost image brightness.

3.2 Multi-scale context-attention network (MSCA-Net)

The main component of our proposed framework is the MSCA-Net block that localizes the area of COVID-19 lesions

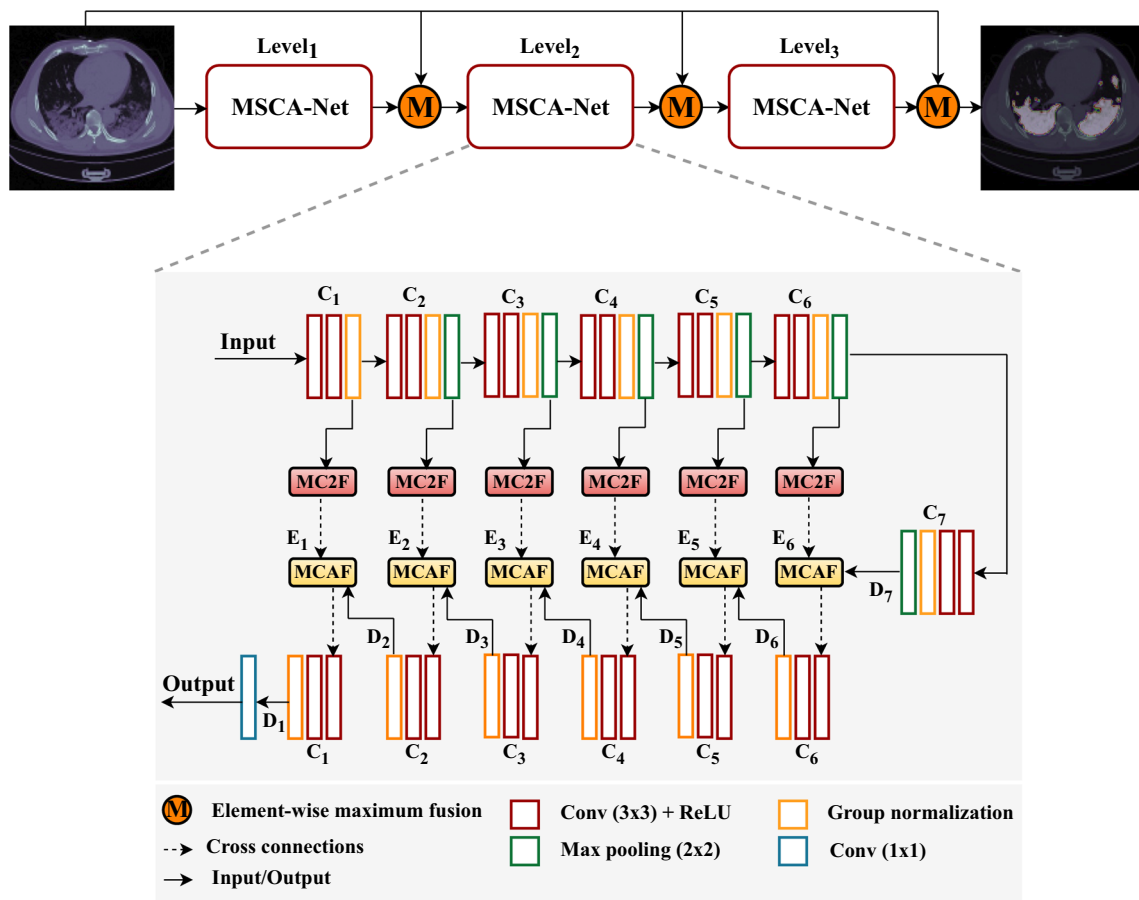


Fig. 1 The overview of the proposed MLCA2F framework

in CT images. One of the key advantages of MSCA-Net is that it can be trained to directly focus on the target regions, resolve object ambiguities, and capture contextual information. It adopts the encoder E_i and decoder D_i structures, where $i \in \{1,2,3,4,5,6\}$, for semantic segmentation, equipped with MC2F and MCAF modules. The encoder and decoder components consist of six convolution blocks, C_i , each containing two 3×3 convolution layers followed by a ReLU activation function, group normalization [20], while the seventh convolution block is a transitional block between encoder and decoder. The output shape of each convolutional group is shown in Table 1. To progressively reduce the resolution of feature maps, the group normalization output of the encoder component is fed into the 2×2 max-pooling layer. Broadly, the encoder is the process of capturing the latent feature representation from input images, whereas the decoder path is designed in a way to upsample the feature maps and force the model to focus on infected areas in CT images through the cross-connections [21]. For accurate COVID-19 lesion segmentation, our framework mainly has three innovative mechanisms, the first is integrating the MC2F component to learn more COVID-19 lesion details and capture contex-

Table 1 Output shape of each convolutional group

Layers	Output shape
Input	$256 \times 256 \times 1$
C_1	$256 \times 256 \times 8$
C_2	$128 \times 128 \times 16$
C_3	$64 \times 64 \times 24$
C_4	$32 \times 32 \times 32$
C_5	$16 \times 16 \times 128$
C_6	$8 \times 8 \times 256$
C_7	$4 \times 4 \times 512$

tual information, the second is using the MCAF module to highlight the salient features, while the third is applying the fused MSCA-Net blocks through the element-wise maximum layer to significantly boost the performance.

3.3 Multi-scale contextual feature fusion (MC2F)

Considering that a single-scale feature model cannot highlight the contextual information and learn more COVID-19 lesion details, due to the large variations in the COVID-19

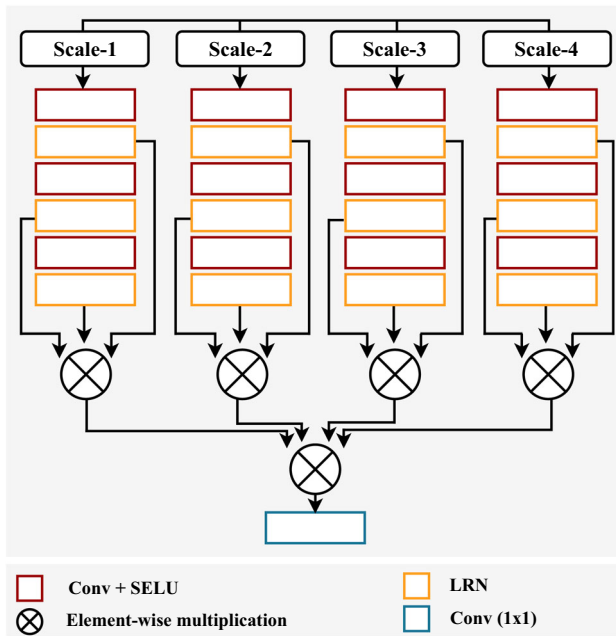


Fig. 2 The overview of the proposed MC2F architecture

lesion size and shape, boundary ambiguity, as well as complex structure, we design a Multi-Scale Contextual Feature Fusion (MC2F). Taking advantage of the fact that the multi-scale features provide semantically different information, MC2F captures the detailed features from different receptive fields that are useful for locating boundaries. Specifically, in each MC2F block, four scales of convolutional layers with the size of $[3 \times 3]$, $[5 \times 5]$, $[7 \times 7]$, and $[9 \times 9]$ are involved. Within each scale structure, we integrate three convolutional layers such that the output of each layer is passed to the Scaled Exponential Linear Unit (SELU) activation function [22]. After that, each feature is fed to a Local Response Normalization (LRN) [23]. The output of the three LRN layers is later fused through the element-wise multiplication layer [24]. The four

parallel scale paths are connected via a global element-wise multiplication layer. Finally, the output of the element-wise multiplication layer is fed into a 1×1 convolution layer to get the same MC2F input depth. The arrangement and description of the MC2F block are shown in Fig. 2. The constructed scale blocks are implemented in a similar structure and built by the same number of kernels (n) and stride (S), and they differ essentially from the kernel shape $[\text{width} \times \text{height}] \times n$, padding (P), LRN and SELU parameters. Table 2 shows the parameters specification of MC2F block.

3.4 Multi-context attentional feature (MCAF)

Since the current COVID-19 lesion segmentation methods are facing challenges posed by varying shapes, sizes, and positions of the infected area [6,7], we propose a multi-context attentional feature mechanism, called MCAF, which guides the MLCA2F model to learn more discriminative features for separating the COVID-19 and non-COVID-19 pixels, and then, to estimate the position of the more complex regions in the whole CT image. As shown in Fig. 3, the proposed MCAF adopts three typical fusion layers, namely L_2 normalization fusion approach [25], and element-wise multiplication and addition layers [24]. Specifically, in the first, the output of E_i is fed to the first convolutional layer. To generate the layers with the same E_i depth, the D_{i+1} goes through 1×1 convolution layer. Moreover, a bilinear interpolation [26] is introduced to upsample the decoded features, D_{i+1} , and obtain the same size resolution as the output of E_i . The output of the bilinear interpolation layer is passed to the second convolutional layer. The fused E_i and upsampled D_{i+1} features fed into the third convolutional layer. Note that, to keep the same predefined depth, the convolutional layers are used with the same kernel shape as C_i . After that, the obtained features are directly passed through an element-wise sigmoid function to generate multi-contextual attentive

Table 2 Structure parameters of MC2F block

Scale	Layer	Kernel shape	Specifications
Scale-1	Conv	$[3 \times 3] \times 512$	$(P = 1, S = 1)$
	SELU	–	$(\lambda = 1.04, \alpha = 1.58)$
	LRN	–	$(k = 1, \alpha = 10^{-2}, \beta = 0.62)$
Scale-2	Conv	$[5 \times 5] \times 512$	$(P = 2, S = 1)$
	SELU	–	$(\lambda = 1.02, \alpha = 1.63)$
	LRN	–	$(k = 3, \alpha = 10^{-5}, \beta = 0.95)$
Scale-3	Conv	$[7 \times 7] \times 512$	$(P = 3, S = 1)$
	SELU	–	$(\lambda = 1.09, \alpha = 1.72)$
	LRN	–	$(k = 2, \alpha = 10^{-4}, \beta = 0.84)$
Scale-4	Conv	$[9 \times 9] \times 512$	$(P = 4, S = 1)$
	SELU	–	$(\lambda = 1.08, \alpha = 1.35)$
	LRN	–	$(k = 2, \alpha = 10^{-5}, \beta = 0.73)$

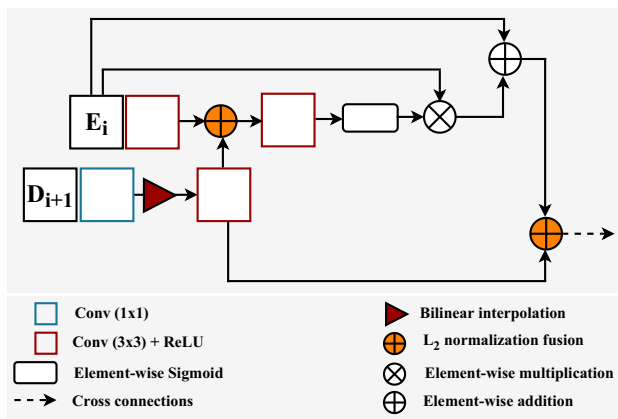


Fig. 3 The overview of the proposed MCAF architecture

feature maps. At the end of MCAF, the second L_2 normalization fusion concatenates the output of the multi-contextual attentive features with the feature maps extracted from the upsampled decoding path to recover the COVID-19 lesion details.

4 Experimental results

In this section, we validate the proposed method on a publicly available CT dataset. First, we introduce the CT images used to assess the segmentation performance and the implementation details. Then, we highlight the details of the findings and report the performance comparison with the state-of-the-art methods.

4.1 CT image acquisition

In this study, the chest axial CT dataset is collected from Kaggle [27] which is a combination of [28,29] and [30]. It consists of 20 full-chest 3D CT scans from different patients diagnosed with COVID-19. We extracted 3520 2D CT axial slices from the 3D volumes. All CT slices came with a corresponding ground truth defined by the radiologists and physicians. The CT dataset was divided into 60% for training, 20% for validation, and 20% for testing.

4.2 Implementation details

The proposed MLCA2F system was trained using Adaptive Moment Estimation (Adam) algorithm [31] with learning rate $\alpha = 10^{-5}$, first moment-decay $\beta_1 = 0.9$ and second moment-decay $\beta_2 = 0.999$ using he_normal algorithm [32] for weight initialization. The optimization ran for 200 epochs with a batch size of 64. All experiments were conducted on a desktop with an Intel Xeon E5-1603 v4 Quad-Core, 2.80 GHz processor base frequency, 10 MB cache, 16 GB RAM, and a single NVIDIA GTX Titan X GPU with 12 GB of installed memory. Our network was implemented using the TensorFlow framework and Keras as backend with Python 3 programming language. The training took 21 h and 46 min. For testing, on average, it took 6 h and 33 min.

4.3 Segmentation results

In this section, a series of comparative experiments are conducted to select the optimal network structure. To evaluate the segmentation results, we follow previous works [33,34]. Then, we validated the MLCA2F framework through the following metrics: Accuracy (Acc), Sensitivity (Sen), Specificity (Spe), Dice Coefficient (DC), also called F1-measure, and Jaccard score (JC) compared with the corresponding ground truths. The experiments are mainly divided into two parts. The first part determines the best quality segmentation network by the absence of the MC2F module, indicated by the sign “-”, while the second part is evaluating the proposed approach by integrating the MC2F block, which is denoted by the sign “+”. Each part is evaluated as a function of the $Level_n$, where $n \in \{1,2,3\}$. During this process, the use of the MC2F blocks improves the segmentation performance, as shown in Table 3. The gain of the MLCA2F at $Level_3$ with MC2F blocks compared to the segmentation results of the model without the MC2F modules at $Level_1$ is 8.20%, 3.70%, 12.70%, 7.72%, and 14.04% in terms of Acc, Sen, Spe, DC, and JC, respectively. Thus, the MC2F block makes an important contribution to the proposed method. Moreover, we also notice that the segmentation results of the MLCA2F at $Level_3$ are more precise than the MLCA2F at $Level_1$ and $Level_2$. To evaluate the visual results, we displayed in Fig. 4

Table 3 Quantitative results of the different MLCA2F components

-/+ MC2F	$Level_n$	Acc (%)	Sen (%)	Spe (%)	DC (%)	JC (%)
-	$Level_1$	90.63	96.06	85.21	91.12	83.68
	$Level_2$	90.80	88.96	92.65	90.63	82.87
	$Level_3$	91.45	92.27	90.64	91.52	84.38
+	$Level_1$	95.71	94.47	96.96	95.66	91.69
	$Level_2$	96.58	99.71	93.46	96.69	93.59
	$Level_3$	98.83	99.76	97.91	98.84	97.72

The bold values signify the results obtained by the proposed approach

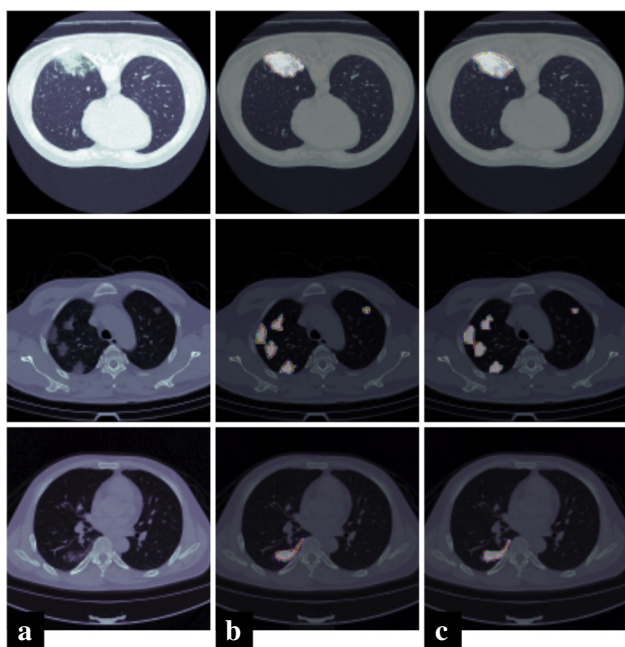


Fig. 4 Examples of proposed COVID-19 lesion segmentation on three representative images. **a** Input images. **b** Ground truth images, **c** Proposed method

the COVID-19 lesion segmentation results of our approach on three representative CT images. It is remarkable that the network learned the detailed features from different receptive fields in the MLCA2F framework. Hence, we can conclude that these three cooperative MSCA-Net models are performing better in segmenting COVID-19 lesions.

4.4 Comparison with the state-of-the-art

Since the research on the Kaggle CT dataset is still relatively scarce, we have implemented the recent state-of-the-art image segmentation methods to demonstrate the effectiveness of our solution. Therefore, we compare our method against six segmentation approaches, namely U-Net [11], U-Net++ [14], Attention U-Net (AU-Net) [8], Mask R-CNN [15], Multi-Scale Discriminative Network (MSD-Net) [4], and Multi-scale Context-attention Network (MC-Net) [16].

In Table 4, we provide the quantitative results with respect to the five evaluation metrics. We observe from the results that MLCA2F outperforms the recent state-of-the-art methods by a large margin over all five metrics. Figure 5 shows visual segmentation results on the lung dataset obtained by MLCA2F and the other state-of-the-art methods. We can see from the figure that the proposed MLCA2F framework is more precise than the other methods, which can capture detailed boundary information.

5 Discussion

In this paper, we present an optimal data representation framework for COVID-19 lesion segmentation from CT images. Motivated by the outstanding performance of multi-scale CNN-based methods, the MLCA2F architecture is proposed to cope with the large variations in the COVID-19 lesion size and shape, differences in infected area position, boundary ambiguity, and complex structure problem. To better demonstrate the effectiveness of the proposed framework, two main study axes are presented. In the first axis, we have analyzed the performance results with the change in the number of MSCA-Net blocks, denoted by L_{level_n} . The second study axis is conducted to demonstrate the potentiality of the MC2F module. As shown in Table 3, the MC2F approach has demonstrated outstanding performance for the COVID-19 lesion segmentation. In addition, by integrating the MC2F module, as the number of MSCA-Net blocks increases, the improvements of MLCA2F increment significantly.

To confirm the quality of the MLCA2F algorithm, the experimental results are compared with six previous works. These methods can be broadly categorized into four classes, namely pixel-wise classification-based CNN methods [11, 14], attention-based CNNs [8], region-based CNN models [15], and multi-scale CNN methods [4, 16]. As clearly shown in Table 4, the proposed MLCA2F achieves the best overall performance and consistently outperforms all the other approaches on the Kaggle CT dataset by a large margin. Furthermore, as can be seen from Fig. 5, the COVID-19 lesions segmented by U-Net [11] and U-Net++ [14] have

Table 4 Comparison between the proposed method and other algorithms using the Kaggle CT image dataset

Methods	Acc (%)	Sen (%)	Spe (%)	DC (%)	JC (%)
U-Net [11]	86.90	88.19	85.62	87.07	77.10
U-Net++ [14]	87.45	86.05	88.85	87.28	77.43
AU-Net [8]	88.16	85.44	90.89	87.83	78.31
Mask R-CNN [15]	92.07	90.00	94.14	91.90	85.03
MSD-Net [4]	94.25	92.01	96.50	94.12	88.90
MC-Net [16]	95.10	94.68	95.53	95.08	90.63
MLCA2F (Ours)	98.83	99.76	97.91	98.84	97.72

The bold values signify the results obtained by the proposed approach

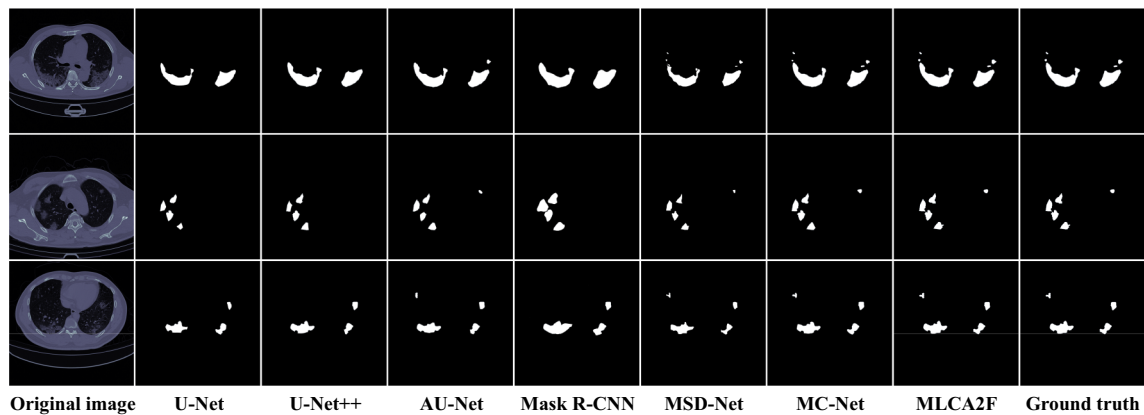


Fig. 5 Visual segmentation comparisons of MLCA2F with six state-of-the-art methods on the Kaggle CT image dataset

a poor segmentation effect and the small regions are not identified, while the CT images segmented by AU-Net [8] possess an incomplete shape of the lesion. Moreover, it can be observed that Mask R-CNN [15] completely fails to segment small tissues. MSD-Net [4] and MC-Net [16] can detect most COVID-19 areas. However, the protruded parts in the boundaries of the COVID-19 regions are not covered correctly and some fine details are missed. By contrast, our approach successfully captures fine structures, extracts more complex regions, and detects the complete shapes.

According to these experimental outcomes, we gain three significant reports. Firstly, it is noticeable that the single MSCA-Net stream without fusion is not sufficient for accurate infected area segmentation. Secondly, these findings reaffirm the superiority of the MC2F module which is more appropriate to preserve information. Finally, the experiments on the CT dataset reveal that using multi-level fusion combined with multi-scale contextual features can automatically focus on critical information, harvest discriminatory features, and strengthen the discriminative representation for COVID-19 lesion segmentation compared with the existing conventional segmentation frameworks.

6 Conclusion

In this study, we propose a novel deep learning model, called MLCA2F, which integrates three fused MSCA-Net units for COVID-19 lesion segmentation from CT scans. The MSCA-Net consists of two robust blocks: MC2F and MCAF. The MC2F module is implemented to capture smooth lesion details, while the MCAF block is modeled to force the MSCA-Net model to focus on COVID-19 lesions in CT images. Furthermore, the repetitive MSCA-Net blocks are introduced to handle the complex structure of COVID-19 lesions and significantly boost the performance. To demonstrate the effectiveness of the proposed MLCA2F framework,

several experiments are conducted on the Kaggle CT dataset. The results of the MLCA2F segmentation experiment show that our method has better semantically consistent segmentation than the existing methods. In conclusion, the proposed segmentation method is beneficial for the automatic COVID-19 lesion segmentation study. In the future, we will further collect more challenging images, improve the proposed method by taking into consideration multi-modal analysis, and evaluate our approach on more datasets.

Acknowledgements This work is supported in part by the PPR2-2015 project under grant number 14UIZ2015, in part by the Scientific and Technological Research Support Program related to COVID-19 under grant number COV/2020/73, and in part by the AI Khawarizmi project under grant number ALKHAWARIZMI/2020/02 financed by the Moroccan government through the CNRST funding program.

Data Availability The datasets analyzed during the current study are available in the Kaggle repository, <https://www.kaggle.com/datasets/andrewmvd/covid19-ct-scans>.

References

1. Murakami, M., Miura, F., Kitajima, M., Fujii, K., Yasutaka, T., Iwasaki, Y., Ono, K., Shimazu, Y., Sorano, S., Okuda, T., Ozaki, A., Katayama, K., Nishikawa, Y., Kobashi, Y., Sawano, T., Abe, T., Saito, M., Tsubokura, M., Naito, W., Imoto, S.: COVID-19 risk assessment at the opening ceremony of the Tokyo 2020 Olympic Games. *Microb. Risk Anal.* **19**, 00162 (2021)
2. Ter-Sarkisov, A.: One shot model for the prediction of COVID-19 and lesions segmentation in chest CT scans through the affinity among lesion mask features. *Appl. Soft Comput.* **116**, 108261 (2022)
3. Fan, D., Zhou, T., Ji, G., Zhou, Y., Chen, G., Fu, H., Shen, J., Shao, L.: Inf-Net: automatic COVID-19 lung infection segmentation from CT images. *IEEE Trans. Med. Imaging* **39**, 2626–2637 (2020)
4. Zheng, B., Liu, Y., Zhu, Y., Yu, F., Jiang, T., Yang, D., Xu, T.: MSD-Net: multi-scale discriminative network for COVID-19 lung infection segmentation on CT. *IEEE Access.* **8**, 185786–185795 (2020)

5. Chen, J., Wu, L., Zhang, J., Zhang, L., Gong, D., Zhao, Y., Chen, Q., Huang, S., Yang, M., Yang, X., Hu, S., Wang, Y., Hu, X., Zheng, B., Zhang, K., Wu, H., Dong, Z., Xu, Y., Zhu, Y., Chen, X., Zhang, M., Yu, L., Cheng, F., Yu, H.: Deep learning-based model for detecting 2019 novel coronavirus pneumonia on high-resolution computed tomography. *Sci. Rep.* **10**, 1–11 (2020)
6. Kuchana, M., Srivastava, A., Das, R., Mathew, J., Mishra, A., Khatter, K.: AI aiding in diagnosing, tracking recovery of COVID-19 using deep learning on Chest CT scans. *Multimed. Tools Appl.* **80**, 9161–9175 (2020)
7. Zhou, T., Canu, S., Ruan, S.: Automatic COVID-19 CT segmentation using U-Net integrated spatial and channel attention mechanism. *Int. J. Imaging Syst. Technol.* **31**, 16–27 (2020)
8. Oktay, O., Schlemper, J., Folgoc, L.L., Lee, M.C.H., Heinrich, M.P., Misawa, K., Mori, K., McDonagh, S., Hammerla, N., Kainz, B.: Attention U-Net: learning where to look for the pancreas. [arXiv:1804.03999](https://arxiv.org/abs/1804.03999) (2018)
9. Liu, Z., Yang, C., Huang, J., Liu, S., Zhuo, Y., Lu, X.: Deep learning framework based on integration of S-Mask R-CNN and Inception-v3 for ultrasound image-aided diagnosis of prostate cancer. *Futur. Gener. Comput. Syst.* **114**, 358–367 (2021)
10. Alalwan, N., Abozeid, A., ElHabshy, A., Alzahrani, A.: Efficient 3D deep learning model for medical image semantic segmentation. *Alex. Eng. J.* **60**, 1231–1239 (2021)
11. Ronneberger, O., Fischer, P., Brox, T.: U-Net: convolutional networks for biomedical image segmentation. In: *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pp. 234–241 (2015)
12. Khalifa, N., Manogaran, G., Taha, M., Loey, M.: A deep learning semantic segmentation architecture for COVID-19 lesions discovery in limited chest CT datasets. *Expert Syst.* **39**, e12742 (2021)
13. Peyvandi, A., Majidi, B., Peyvandi, S., Patra, J.: Computer-aided-diagnosis as a service on decentralized medical cloud for efficient and rapid emergency response intelligence. *New Gener. Comput.* **39**, 677–700 (2021)
14. Zhou, Z., Rahman Siddiquee, M., Tajbakhsh, N., Liang, J.: UNet++: a nested U-Net architecture for medical image segmentation. In: *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support*, pp. 3–11 (2018)
15. He, K., Gkioxari, G., Dollár, P., Girshick, R.: Mask R-CNN. In: *Proceedings of 2017 IEEE International Conference on Computer Vision (ICCV)*, pp. 2980–2988 (2017)
16. Xia, H., Ma, M., Li, H., Song, S.: MC-Net: multi-scale context-attention network for medical CT image segmentation. *Appl. Intell.* **52**, 1508–1519 (2021)
17. Sang, H., Wang, Q., Zhao, Y.: Multi-scale context attention network for stereo matching. *IEEE Access.* **7**, 15152–15161 (2019)
18. Bakkouri, I., Afdel, K.: Multi-scale CNN based on region proposals for efficient breast abnormality recognition. *Multimed. Tools Appl.* **78**, 12939–12960 (2018)
19. Tang, J., Mat Isa, N.: Adaptive image enhancement based on bi-histogram equalization with a clipping limit. *Comput. Electr. Eng.* **40**, 86–103 (2014)
20. Crytzer, T., Keramati, M., Anthony, S., Cheng, Y., Robertson, R., Dicianno, B.: Exercise prescription using a group-normalized rating of perceived exertion in adolescents and adults with spina bifida. *PM&R* **10**, 738–747 (2018)
21. Trebing, K., Stańczyk, T., Mehrkanoon, S.: SmaAt-UNet: precipitation nowcasting using a small attention-UNet architecture. *Pattern Recogn. Lett.* **145**, 178–186 (2021)
22. Klambauer, G., Unterthiner, T., Mayr, A., Hochreiter, S.: Self-normalizing neural networks. In: *Proceedings of the 31st International Conference on Neural Information Processing Systems 2017*, pp. 972–981 (2017)
23. Krizhevsky, A., Sutskever, I., Hinton, G.: ImageNet classification with deep convolutional neural networks. *Commun. ACM* **60**, 84–90 (2017)
24. Ren, S., Han, C., Yang, X., Han, G., He, S.: TENet: triple excitation network for video salient object detection. In: *European Conference on Computer Vision. Computer Vision: ECCV 2020*, pp. 212–228 (2020)
25. Jin, X., Jiang, Q., Chu, X., Lang, X., Yao, S., Li, K., Zhou, W.: Brain medical image fusion using L2-norm-based features and fuzzy-weighted measurements in 2-D Littlewood–Paley EWT domain. *IEEE Trans. Instrum. Meas.* **69**, 5900–5913 (2020)
26. Zhou, R., Cheng, Y., Liu, D.: Quantum image scaling based on bilinear interpolation with arbitrary scaling ratio. *Quantum Inf. Process.* **18**, 1–19 (2019)
27. COVID-19 CT scans in Kaggle. <https://www.kaggle.com/andrewmvd/covid19-ct-scans>. Accessed 19 June 2020
28. Paiva, O.: CT scans of patients with COVID-19 from Wenzhou Medical University. <https://coronacases.org/>. Accessed 19 June 2020
29. Glick, Y.: COVID-19 Pneumonia. <https://radiopaedia.org/playlists/25887>. Accessed 19 June 2020
30. Jun, M., Cheng, G., Yixin, W., Xingle, A., Jiantao, G., Ziqi, Y., Mingqing, Z., Xin, L., Xueyuan, D., Shucheng, C., Hao, W., Sen, M., Xiaoyu, Y., Ziwei, N., Chen, L., Lu, T., Yuntao, Z., Qiongjie, Z., Guoqiang, D., Jian, H.: COVID-19 CT lung and infection segmentation dataset. <https://zenodo.org/record/3757476>. Accessed 19 June 2020
31. Kingma, D., Ba, J.: Adam: a method for stochastic optimization. [arXiv:1412.6980](https://arxiv.org/abs/1412.6980)
32. He, K., Zhang, X., Ren, S., Sun, J.: Delving deep into rectifiers: surpassing human-level performance on ImageNet classification. In: *2015 IEEE International Conference on Computer Vision (ICCV)* (2015)
33. Zhao, X., Wang, S., Zhao, J., Wei, H., Xiao, M., Ta, N.: Application of an attention U-Net incorporating transfer learning for optic disc and cup segmentation. *Signal Image Video Process.* **15**, 913–921 (2020)
34. Al-Shamasneh, A., Jalab, H., Shivakumara, P., Ibrahim, R., Obaidallah, U.: Kidney segmentation in MR images using active contour model driven by fractional-based energy minimization. *SIViP* **14**, 1361–1368 (2020)

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.