

CNN Filter Learning from Drawn Markers for the Detection of Suggestive Signs of COVID-19 in CT Images

Azael M. Sousa¹, Fabiano Reis², Rachel Zerbini², João L. D. Comba³ and Alexandre X. Falcão¹

Abstract—Early detection of COVID-19 is vital to control its spread. Deep learning methods have been presented to detect suggestive signs of COVID-19 from chest CT images. However, due to the novelty of the disease, annotated volumetric data are scarce. Here we propose a method that does not require either large annotated datasets or backpropagation to estimate the filters of a convolutional neural network (CNN). For a few CT images, the user draws markers at representative normal and abnormal regions. The method generates a feature extractor composed of a sequence of convolutional layers, whose kernels are specialized in enhancing regions similar to the marked ones, and the decision layer of our CNN is a support vector machine. As we have no control over the CT image acquisition, we also propose an intensity standardization approach. Our method can achieve mean accuracy and kappa values of 0.97 and 0.93, respectively, on a dataset with 117 CT images extracted from different sites, surpassing its counterpart in all scenarios.

I. INTRODUCTION

The world is currently facing a public health crisis caused by COVID-19, an infectious disease named after the new coronavirus SARS-CoV-2. According to the World Health Organization (WHO), over 100 million people have been infected worldwide by COVID-19. An early diagnosis can significantly reduce the transmission rate by keeping the infected under lockdown. Although the reverse-transcriptase polymerase chain reaction (RT-PCR) is considered the gold-standard for the diagnosis of COVID-19, computed tomography (CT) images are also important to support the diagnosis, monitor the progression of the disease, and detect possible complications [7], [12], [14]. Studies have reported cases with negative RT-PCR and CT exams with abnormal characteristics that became positive RT-PCR later on [22], [8]. As consequence, laboratories are employing both RT-PCR and chest CT for patients with suspect of COVID-19 infection [12].

In CT images, common findings of COVID-19 include (i) peripheral, bilateral, ground-glass opacities (GGO) with or without consolidation, (ii) multifocal GGO of rounded morphology, and (iii) reverse halo sign or other findings of organizing pneumonia [11], as shown in Figure 1. As these findings may appear in other pulmonary conditions, they can only be suggestive of COVID-19.

Due to the urgency of the problem, several deep learning methods, mostly Convolutional Neural Networks (CNN), have been presented to detect suggestive patterns of COVID-19 [15], [16], [17], [13], [9], [2], [10]. Wang *et al.* [21]

employ a 3D CNN, called DeCovNet, that can achieve an accuracy of approximately 90% on an in-house dataset with 630 CT images. Polsinelli *et al.* [13] uses a light design of the 2D SqueezeNet CNN. The authors report an accuracy of 83% on two public datasets [23], [18]. Silva *et al.* [17] propose a 2D CNN called EfficientCovidNet. The authors employ a voting mechanism that determines the patient's outcome based on the number of slices classified as COVID-19. This method can achieve an accuracy of 89% when tested on new images of the same dataset used for training, but presents an accuracy drop to 56% when tested on images from a different dataset. Other studies [9], [2], [10] have addressed the problem by first identifying lung lesions and then giving the diagnosis of the patient based on the classification of those lesions.

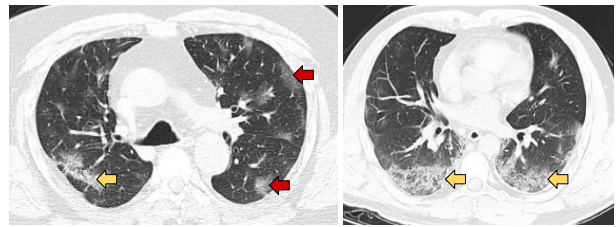


Fig. 1. CT images from patients with confirmed diagnosis of COVID-19. The red arrows indicate ground-glass opacities whereas yellow arrows indicate consolidations.

One of the main drawbacks of deep learning is the number of annotated images required for training. Typically, it requires large annotated datasets to learn representative features from different abstraction levels by backpropagation, making the trained model robust to real-world scenarios. However, such datasets can be challenging to gather. Due to its novelty, publicly available datasets of COVID-19 with standardized, annotated, and volumetric CT exams are scarce. Another obstacle faced when dealing with deep learning models is the architecture choice. Many methods employ off-the-shelf architectures for applications of entirely different domains, leading to poor computational resources management, too many training samples, and still non-explainable solutions.

We involve the specialist in the training process to address the above issues. By taking advantage of the expert's knowledge about the problem, we can build a more efficient model with less trainable parameters and considerably reduced number of annotated training images. Our approach explores a recently proposed method for feature extraction (i.e., a network formed by a sequence of convolutional

¹Laboratory of Image Data Science (LIDS), Institute of Computing, University of Campinas

²Faculty of Medical Sciences, University of Campinas

³Institute of Informatics, Federal University of Rio Grande do Sul

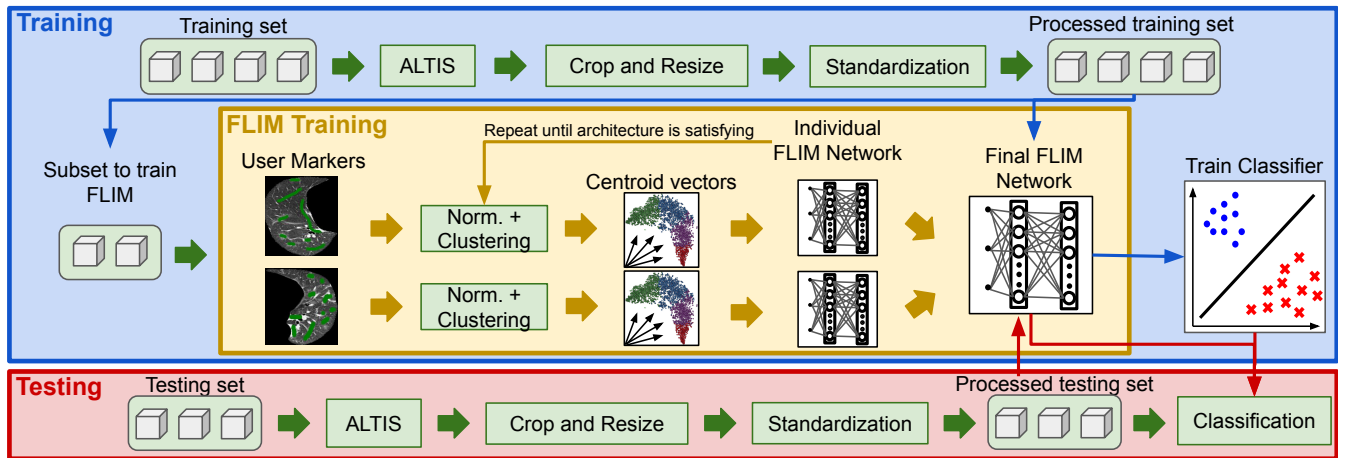


Fig. 2. Pipeline of the proposed method. A subset with few images is selected from a preprocessed training set to train the FLIM network. After the FLIM training, the entire training set is passed through the network for feature extraction, and a SVM classifier is trained. The classification of the test set is performed over the activations of the FLIM network.

layers only), named *Feature Learning from Image Markers* (FLIM) [4], [5]. FLIM kernels are estimated based on a set of markers drawn by the user in very few training images (e.g., two or three images per class), with no backpropagation. These markers indicate pertinent image regions for feature extraction, creating kernels that enhance regions similar to the marked ones.

In this work, we extend the 2D FLIM network to detect suggestive signs of COVID-19 in volumetric CT images. In addition to that, we adopt an interactive architecture learning approach to construct the FLIM architecture. The network designer incorporates his/her knowledge in the training loop by studying the impact that the selected hyperparameters have on the network’s accuracy over a small validation set and adapting them accordingly.

In Section II we describe the proposed method, Section III depicts the experimental procedure and assessment of our method in comparison with another approach from the state-of-the-art, and Section IV states the conclusion of our work.

II. METHOD

The proposed method is divided into four stages: lung segmentation, pre-processing, feature extraction, and classification. Figure 2 shows the entire pipeline.

A. Lung Segmentation

Image segmentation consists of defining the spatial extension of one or multiple objects of interest. The segmentation stage is of paramount importance in deep learning as it reduces the regions to be processed and consequently the computational effort. In this work, we employed the method *ALTIS* (Automatic Lungs and Trachea Image Segmentation) [20] for the segmentation of the lungs. ALTIS is a graph-based method that explores the Image Foresting Transform (IFT) framework [6] to estimate object seeds and delineate the lungs and trachea-bronchi as separated objects.

B. Preprocessing

Each lung from the CT image is first cropped according to the mask generated by ALTIS. To guarantee feature vectors of the same size, the cropped-lung images are resized to a lower common dimension of $200 \times 200 \times 200$ voxels; wherein ground-glass opacities can still be clearly seen.

The lack of control during image acquisition and reconstruction creates a disparity problem regarding the intensity feature that may hampers deep learning methods [1]. For that reason, we applied a standardization method to keep each lung image within a fixed intensity range. This method identifies the first two summits closest to each edge of the lung image’s intensity histogram. The first summit represents the darker and typical parenchyma region, while the second one represents the bright veins/arteries and part of the mediastinal region. Then, the histogram is shifted to a previously established spectrum. Figure 3 shows an example of the standardization method.

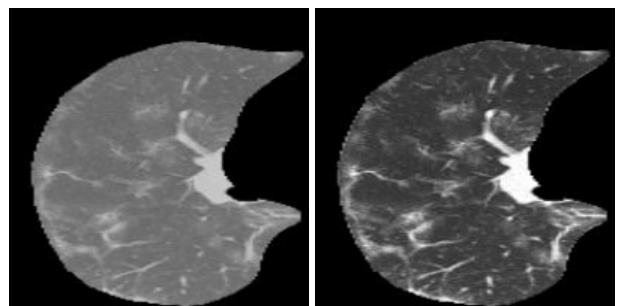


Fig. 3. Standardization method. The left image is the original right lung cropped and resized, whereas the right one is the cropped and resized right lung image after standardization.

C. Feature Extraction

Feature extraction is an essential part of any machine learning algorithm. CNN’s have powerful feature extractors that can learn representative patterns from training samples.

In our case, they consist of a sequence of convolutional layers, each containing marker-based normalization, convolution with a filter bank, ReLU activation and, optionally, a max-pooling operation. Let \mathcal{F}_l be the set of convolutional filters (kernels) for a layer $l \in \{1, 2, \dots, L\}$. The FLIM network estimates \mathcal{F}_l by analyzing the markers drawn by the user, with no backpropagation. It works as an attention mechanism for relevant regions.

Let \mathcal{D} be the training set containing CT images from both classes, with m channels. The convolution operation of an image $I \in \mathcal{D}$ with a kernel $f \in \mathcal{F}_l$ highlights local patterns that matches f . It is denoted as $P_{I(v)} * f$, where $P_{I(v)}$ is a patch of size $h \times w \times d \times m$, extracted from the image I around the voxel v . A unit-norm kernel f may be seen as an orthogonal vector to a hyperplane at the origin of the patch feature space. The convolution with f measures the distance to that hyperplane and the ReLU function eliminates regions with negative distances (non-activated regions). A kernel f is said discriminative for a given class $c \in \{1, 2, \dots, C\}$ when the regions activated by f are different from those it activates for other classes.

Let \mathcal{M}_I be the set of markers drawn by the user on image $I \in \mathcal{D}$ (green engravings on “User Markers” at Figure 2). Every voxel $v \in \mathcal{M}_I$ has a patch $P_{I(v)}$ centered on it. Let \mathcal{P}_I be the set of patches from image I . Marker-based normalization is needed to eliminate bias in the FLIM network. It shifts all patches by moving their geometric center to the origin of the patch feature space. By grouping the normalized patches \mathcal{P}_I (e.g., using K-Means) such that \mathcal{G}_I is the resulting set of groups, each unit vector pointing towards the centroid of each group $g \in \mathcal{G}_I$ is a potential kernel f_g for \mathcal{F}_l in the FLIM architecture. Therefore, the complete set of filters for layer l is $\mathcal{F}_l = \cup_{g \in \mathcal{G}_I, I \in \mathcal{D}} f_g$.

To remove redundant kernels from \mathcal{F}_l , whenever necessary, we use a methodology similar to the one described in [19]. The idea is to make all kernels linearly independent. For that reason, the Principal Component Analysis (PCA) method is applied over \mathcal{F}_l generating a set of orthogonal eigenvectors and eigenvalues. The k_l eigenvectors associated with the highest eigenvalues are selected as the final set of kernels, where k_l is the number of desired kernels for layer l . The kernel estimation procedure is repeated for every subsequent layer $l + 1$ by extracting patches from marker voxels at the output of layer l , until the network designer is satisfied with the architecture. However, the user only draws markers over the original images (at the input layer).

1) *Interactive Architecture Learning*: To define the FLIM architecture, we explore the maximum class separability that a layer can achieve on its own. For that purpose, each layer is built individually by the network designer. Its hyperparameters, e. g. number of kernels, kernel size, and dilation rate (in case of atrous convolution) are intuitively optimized according to the model’s effectiveness on a small validation set. Once a layer has reached its maximum separability, the network designer decides to either build another layer and repeat the process or finish the model.

D. Classification

Once the network architecture and respective weights have been computed, all training data, both normal CT and with COVID-19, are passed forward through the network. The output of the last convolutional layer is then used to train a Support Vector Machine (SVM) classifier [3].

III. EXPERIMENTS AND DISCUSSION

A. Datasets

The Faculty of Medical Sciences at UNICAMP, Brazil, has provided an annotated dataset containing 66 volumetric CT images of patients with RT-PCR positive for COVID-19. Additionally, we have gathered 51 volumetric CT images with normal parenchyma from different sites. All images were interpolated to an isotropic voxel dimension of $1 \times 1 \times 1$ mm³. We conducted experiments on five random stratified splits where 50% were used to train the SVM classifier and 50% for testing. From the SVM training set of every split, 10% of the images with COVID-19 (i.e. about 6 images) were separated to train the FLIM network (i.e., 3 images for marker selection and 3 images for hyperparameter selection based on validation, see Section III-B). The splits were performed in a patient-wise manner to ensure learning integrity, where images from the same patient are either in the training set or test set but never in both.

B. FLIM architecture

The FLIM architecture was determined by analyzing its accuracy and kappa scores on the validation split. To avoid skewing the interactive architecture learning process, we selected four different sets of images to train FLIM, extracted from the training/validation split. The final model was the one with the highest mean accuracy and kappa over all four sets. It has two convolutional layers with two kernels of size $3 \times 3 \times 3$, dilation rate of 3, and pooling with a stride of 4. The ReLU activation function was used in all layers.

C. Compared Method

The method used for comparison is known as deCoVNet [21]. It is a 3D deep CNN composed of a convolutional layer with a kernel of size $5 \times 7 \times 7$, batch normalization and pooling, followed by 2 3D residual blocks. The last part is a progressive classifier with 3 convolutional layers and a fully connected one with the softmax function. Henceforth we call deCoVNet* the deCoVNet model trained from scratch with the same 50%-50% data splits described in Section III-A.

D. Results and Discussion

The accuracy and kappa scores of the proposed method and the deCoVNet* for the classification of COVID-19 patterns are presented in Table I. Note that our method surpassed the deCovNet* in all five splits by a considerable margin. The data imbalance between the classes makes kappa more sensitive than accuracy. The inferior results achieved by deCovNet* might be due to two reasons: (i) the size of the dataset available in this study is five times smaller than the in-house dataset used in the deCovNet’s original

TABLE I
ACCURACY (ACC.) AND KAPPA SCORES FOR THE CLASSIFICATION OF COVID-19

Method	Split 1		Split 2		Split 3		Split 4		Split 5		Mean and Stdev	
	Acc.	Kappa	Acc.	Kappa	Acc.	Kappa	Acc.	Kappa	Acc.	Kappa	Acc.	Kappa
deCovNet*	0.87	0.73	0.84	0.69	0.87	0.74	0.92	0.83	0.94	0.89	0.89 ± 0.04	0.77 ± 0.08
FLIM layer 1	0.95	0.90	0.94	0.87	0.96	0.92	0.95	0.90	0.91	0.81	0.94 ± 0.02	0.88 ± 0.04
FLIM layer 2	0.96	0.92	0.97	0.95	0.96	0.92	0.95	0.89	0.97	0.95	0.97 ± 0.01	0.93 ± 0.02

paper [21], and (ii) the number of training samples (about 58 images) is considerably lower than the number used in the original paper. As a result, backpropagation might have been affected by overfitting. Since our method is not restrained to backpropagation, we successfully trained the FLIM network with a small set of CT images, generating more representative features to train the SVM classifier.

The standardization method plays an essential role in the entire process. Without it, marker-based normalization does not work properly and the performances of the FLIM network and SVM are seriously affected. That is, images with normal parenchyma might present activated regions by kernels specialized in enhancing ground-glass opacities.

IV. CONCLUSION

In this work, we extended from 2D to 3D a feature extraction method, named FLIM, based on image markers and a sequence of convolutional layers. Together with a SVM classifier, the method was used to detect suggestive signs of COVID-19 in CT images. The proposed method is backpropagation-free and computes its kernel weights by analyzing markers drawn by the user on very few images. The whole process requires segmenting the lungs and trachea from CT exams, preprocess the images, train a FLIM network for feature extraction, and train a SVM classifier for the detection of abnormal images with ground-glass opacities. The experiments showed that the proposed method can surpass deCovNet*, a state-of-the-art approach, suggesting that FLIM could generate more representative image features for the given problem. We intend to make our model available for public use and design another FLIM network for ground-glass segmentation.

V. ACKNOWLEDGEMENTS

The authors thank the financial support from CAPES, CNPq (303808/2018-7 grant) and FAPESP (2014/12236-1 and 2017/03940-5 grants).

REFERENCES

- [1] R. Berenguer, M. D. R. Pastor-Juan, J. Canales-Vázquez, M. Castro-García, M. V. Villas, F. M. Legorburo, and S. Sabater, "Radiomics of CT features may be nonreproducible and redundant: influence of CT acquisition parameters," *Radiology*, vol. 288, pp. 407–415, 2018.
- [2] J. Chen, L. Wu, J. Zhang, L. Zhang, D. Gong, Y. Zhao, Q. Chen, S. Huang, M. Yang, X. Yang, *et al.*, "Deep learning-based model for detecting 2019 novel coronavirus pneumonia on high-resolution computed tomography," *Scientific reports*, vol. 10, pp. 1–11, 2020.
- [3] C. Cortes and V. Vapnik, "Support-Vector Networks," *Machine Learning*, vol. 20, 1995.
- [4] I. E. de Souza, B. C. Benato, and A. X. Falcao, "Feature learning from image markers for object delineation," in *2020 33rd SIBGRAPI Conf. on Graphics, Patterns and Images*, 2020, pp. 116–123.
- [5] I. E. de Souza and A. X. Falcão, "Learning cnn filters from user-drawn image markers for coconut-tree image classification," *IEEE Geoscience and Remote Sensing Letters*, 2020.
- [6] A. X. Falcão, J. Stolfi, and R. de Alencar Lotufo, "The image foresting transform: Theory, algorithms, and applications," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 26, pp. 19–29, 2004.
- [7] Y. Fang, H. Zhang, J. Xie, M. Lin, L. Ying, P. Pang, and W. Ji, "Sensitivity of chest CT for COVID-19: comparison to RT-PCR," *Radiology*, 2020.
- [8] P. Huang, T. Liu, L. Huang, H. Liu, M. Lei, W. Xu, X. Hu, J. Chen, and B. Liu, "Use of chest ct in combination with negative rt-pcr assay for the 2019 novel coronavirus but high clinical suspicion," 2020.
- [9] C. Jin, W. Chen, Y. Cao, Z. Xu, Z. Tan, X. Zhang, L. Deng, C. Zheng, J. Zhou, H. Shi, *et al.*, "Development and evaluation of an artificial intelligence system for covid-19 diagnosis," *Nature Comms.*, 2020.
- [10] S. Jin, B. Wang, H. Xu, C. Luo, L. Wei, W. Zhao, X. Hou, W. Ma, Z. Xu, Z. Zheng, *et al.*, "AI-assisted CT imaging analysis for covid-19 screening: Building and deploying a medical AI system in four weeks," *MedRxiv*, 2020.
- [11] T. C. Kwee and R. M. Kwee, "Chest CT in COVID-19: what the radiologist needs to know," *RadioGraphics*, vol. 40, 2020.
- [12] 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, *et al.*, "Imaging profile of the covid-19 infection: radiologic findings and literature review," *Radiology: Cardiothoracic Imaging*, vol. 2, no. 1, p. e200034, 2020.
- [13] M. Polsinelli, L. Cinque, and G. Placidi, "A light cnn for detecting covid-19 from ct scans of the chest," *Pattern Recognition Letters*, 2020.
- [14] M. E. E. Rosa, M. J. R. d. Matos, R. S. O. d. P. Furtado, V. M. Brito, L. T. W. Amaral, G. L. Beraldo, E. K. U. N. Fonseca, R. C. Chate, R. B. D. Passos, G. B. d. S. Teles, *et al.*, "COVID-19 findings identified in chest computed tomography: a pictorial essay," *Einstein (São Paulo)*, 2020.
- [15] F. Shi, J. Wang, J. Shi, Z. Wu, Q. Wang, Z. Tang, K. He, Y. Shi, and D. Shen, "Review of artificial intelligence techniques in imaging data acquisition, segmentation and diagnosis for covid-19," *IEEE reviews in biomedical engineering*, 2020.
- [16] F. Shi, L. Xia, F. Shan, D. Wu, Y. Wei, H. Yuan, H. Jiang, Y. Gao, H. Sui, and D. Shen, "Large-scale screening of covid-19 from community acquired pneumonia using infection size-aware classification," *arXiv preprint arXiv:2003.09860*, 2020.
- [17] P. Silva, E. Luz, G. Silva, G. Moreira, R. Silva, D. Lucio, and D. Menotti, "Covid-19 detection in ct images with deep learning: A voting-based scheme and cross-datasets analysis," *Informatics in Medicine Unlocked*, vol. 20, p. 100427, 2020.
- [18] S. Societa italiana di Radiologia Medica e Interventistica, "Covid-19 dataset," <https://www.sirm.org/category/senza-categoria/covid-19/>, accessed: 15-02-2021.
- [19] A. M. Sousa, C. Castelo-Fernández, D. Osaku, E. Bagatin, F. Reis, and A. X. Falcão, "An approach for asbestos-related pleural plaque detection," in *2020 42nd International Conf. of the IEEE Engineering in Medicine & Biology Society (EMBC)*, 2020, pp. 1343–1346.
- [20] A. M. Sousa, S. B. Martins, A. X. Falcao, F. Reis, E. Bagatin, and K. Irion, "ALTIS: A fast and automatic lung and trachea ct-image segmentation method," *Medical physics*, vol. 46, pp. 4970–4982, 2019.
- [21] X. Wang, X. Deng, Q. Fu, Q. Zhou, J. Feng, H. Ma, W. Liu, and C. Zheng, "A weakly-supervised framework for covid-19 classification and lesion localization from chest ct," *IEEE transactions on medical imaging*, vol. 39, no. 8, pp. 2615–2625, 2020.
- [22] X. Xie, Z. Zhong, W. Zhao, C. Zheng, F. Wang, and J. Liu, "Chest CT for typical coronavirus disease 2019 (COVID-19) pneumonia: relationship to negative RT-PCR testing," *Radiology*, vol. 296, 2020.
- [23] J. Zhao, Y. Zhang, X. He, and P. Xie, "Covid-ct-dataset: a ct scan dataset about covid-19," *arXiv preprint arXiv:2003.13865*, 2020.