

Journal Pre-proof

Omnidirectional 2.5D representation for COVID-19 diagnosis using chest CTs

Thiago L.T. da Silveira, Paulo G.L. Pinto, Thiago S. Lermen, Cláudio R. Jung



PII: S1047-3203(23)00025-1
DOI: <https://doi.org/10.1016/j.jvcir.2023.103775>
Reference: YJVC 103775

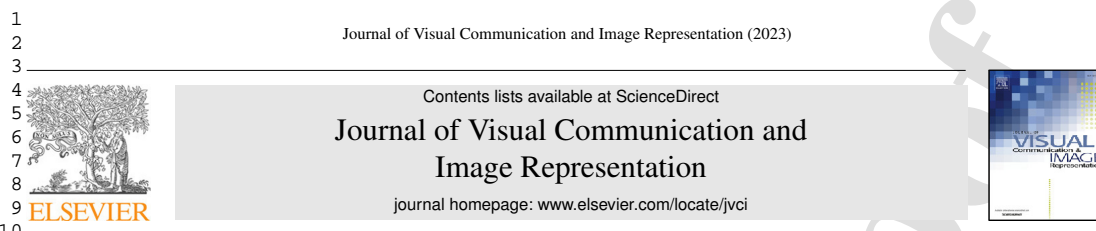
To appear in: *J. Vis. Commun. Image R.*

Received date: 24 August 2022
Revised date: 18 January 2023
Accepted date: 27 January 2023

Please cite this article as: T.L.T. da Silveira, P.G.L. Pinto, T.S. Lermen et al., Omnidirectional 2.5D representation for COVID-19 diagnosis using chest CTs, *J. Vis. Commun. Image R.* (2023), doi: <https://doi.org/10.1016/j.jvcir.2023.103775>.

This is a PDF file of an article that has undergone enhancements after acceptance, such as the addition of a cover page and metadata, and formatting for readability, but it is not yet the definitive version of record. This version will undergo additional copyediting, typesetting and review before it is published in its final form, but we are providing this version to give early visibility of the article. Please note that, during the production process, errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

© 2023 Published by Elsevier Inc.



Omnidirectional 2.5D Representation for COVID-19 Diagnosis using Chest CTs

Thiago. L. T. da Silva^{a,*}, Paulo G. L. Pinto^a, Thiago S. Lermen^a, Cláudio R. Jung^a

^aInstitute of Informatics - Federal University of Rio Grande do Sul, Porto Alegre, 91501-970, Brazil

ARTICLE INFO

Article history:

Received day August 2022

Received in final form day Month 2022

Accepted day Month 2022

Available online day Month 2022

Communicated by Someone

Keywords: 2.5D representation,
COVID-19 diagnosis, ground-glass
opacity, omnidirectional imaging

ABSTRACT

The Coronavirus Disease 2019 (COVID-19) has drastically overwhelmed most countries in the last two years, and image-based approaches using computerized tomography (CT) have been used to identify pulmonary infections. Recent methods based on deep learning either require time-consuming per-slice annotations (2D) or are highly data- and hardware-demanding (3D). This work proposes a novel omnidirectional 2.5D representation of volumetric chest CTs that allows exploring efficient 2D deep learning architectures while requiring volume-level annotations only. Our learning approach uses a siamese feature extraction backbone applied to each lung. It combines these features into a classification head that explores a novel combination of Squeeze-and-Excite strategies with Class Activation Maps. We experimented with public and in-house datasets and compared our results with state-of-the-art techniques. Our analyses show that our method provides better or comparable prediction quality and accurately distinguishes COVID-19 infections from other kinds of pneumonia and healthy lungs.

© 2023 Elsevier B. V. All rights reserved.

1. Introduction

As of December 2022 – more than two years after the Coronavirus disease 2019 (COVID-19) was declared a pandemic – the World Health Organization reported more than 650 million cases and 6.6 million confirmed deaths due to complications linked to Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2) infections¹. Different COVID-19 virus strains continue to evolve and spread over the countries, even those highly vaccinated (Kupferschmidt, 2021; Callaway, 2022). Reverse Transcription Polymerase Chain Reaction (RT-PCR) is the gold standard approach used for testing subjects for COVID-19 (Roberts et al., 2021; Xue and Abhayaratne, 2021), but complementary exams may help diagnosis, depending on the symptoms (Kwee and Kwee, 2020). RT-PCR may not be widely available, is prone to false-negative scoring (Heidarian et al., 2020; Yang et al., 2021; Li et al., 2021; Hassan et al., 2022), and imaging exams like radiography and computed tomography (CT) allow detecting advanced SARS-CoV-2 pulmonary infections (Wang et al., 2020; Roberts et al., 2021; Rahman et al., 2022).

Lung lesions associated with SARS-CoV-2 infections can be found in chest imaging exams in most hospitalized subjects (Kwee and Kwee, 2020; de Souza Portes Meirelles, 2020). The most common radiological finding associated with COVID-19 is the

*Corresponding author:

e-mail: tltsilveira@inf.ufrgs.br (Thiago. L. T. da Silva)

¹<https://covid19.who.int/>

presence of ground-glass opacity (GGO) lesions, but others include crazy paving patterns and consolidations (Chate et al., 2020; Yu et al., 2020; Chen et al., 2020; Afshar et al., 2021; Grando et al., 2020). Chest CTs may thus help patient screening and decision making (Kwee and Kwee, 2020), but require trained experts to analyze data and search for infection signs, which might include other pathologies as well (Roberts et al., 2021).

Image-based machine learning (ML) approaches have been used as diagnostic tools for many diseases and imaging modalities. Many methods have been developed for COVID-19 diagnosis using CT images (Garg et al., 2021; Chaudhary et al., 2021; Yang et al., 2021; Li et al., 2021; Bougourzi et al., 2021a; Xue and Abhayaratne, 2021; Bougourzi et al., 2021b; Wang et al., 2020; Abdel-Basset et al., 2021; Ewen and Khan, 2021; Jin et al., 2020; Li et al., 2021). Despite a large number of published papers in the past two years, there are still several challenges and open questions that are mostly inter-related, such as the definition of the ground-truth annotation (RT-PCR or diagnosis by a radiologist), the number of categories and their nature (e.g., COVID-19 vs. non-COVID-19, COVID-19 vs. healthy, COVID-19 vs. other diseases vs. healthy, etc.), dataset and annotation formats (e.g., full or partial volumes, CT slices, annotations per-volume or per-slice, etc.), and solution type (volumetric or per-slice).

An important design choice is using individual slices or the full CT volume. On the one hand, using slices involves lighter 2D networks, but generating per-slice annotations is costly. On the other hand, volumetric models require a single annotation per volume, but 3D networks are heavier and more challenging to train (Su et al., 2020). In this paper, we propose a novel 2.5D solution that explores the projection of CT volumes onto unit spheres, retrieving points with maximum GGO probabilities and their distances from the center of each lung (sphere), which are mapped to a 2D rectangular domain using the equirectangular projection (da Silveira et al., 2022). With this strategy, we combine the advantages of 2D classification models requiring only per-volume annotations. Our approach achieves very competitive results for the publicly available COVID-CT-MD/SPGC-COVID dataset (Afshar et al., 2021; Heidarian et al., 2021a) and outperforms existing methods for an in-house dataset containing CT scans from two hospitals that provides only per-volume annotations.

The rest of this paper is organized as follows. Section 2 discusses closely related state-of-the-art (SOTA) works for COVID-19 diagnosis from CT scans. The whole pipeline for the method – including volume preprocessing, GGO modeling, mapping to the sphere, and construction of the classification model – is described throughout Section 3. Comparisons with other methods and our ablation studies are presented and discussed in Sections 4 and 5. Finally, the conclusions of this paper are drawn in Section 7.

2. Related Work

In the past two years, many methods for automatically identifying COVID-19 using chest CT scans have been proposed (Garg et al., 2021; Chaudhary et al., 2021; Yang et al., 2021; Li et al., 2021; Bougourzi et al., 2021a; Xue and Abhayaratne, 2021; Bougourzi et al., 2021b; Wang et al., 2020; Abdel-Basset et al., 2021; Ewen and Khan, 2021; Jin et al., 2020; Li et al., 2021). The vast majority rely on deep learning and explore the classical supervised learning paradigm but vary concerning the input dimension (2D slice-based or 3D volume-based), the level of supervision (per-slice annotation or per-volume annotation), the number of classes (two or many), and also the choice for training/validation/test datasets. As such, a fair comparison is hard to obtain, and, as pointed out by Roberts et al. (2021), most existing approaches present methodological flaws or underlying dataset biases (caused by different factors such as using different scanners, dosage, exposure, etc.). Ewen and Khan (2021) tackled the problems of slight domain shift and weak supervision towards an online improvement of COVID-19 detection models to alleviate the bias issue, but their results are still embryonic. Next, we discuss related approaches categorized according to the input data representation.

2.1. Slice-based Classification Methods

In a couple of works, Heidarian et al. (2020, 2021b) propose to use capsule networks for COVID-19 diagnosis from chest CTs. The second method – COVID-FACT (Heidarian et al., 2020) – builds on top of the first, CT-CAPS (Heidarian et al., 2021b). Both methods rely on lung segmentation (Hofmanninger et al., 2020) and process individual 2D slices before aggregating the results for final diagnosis. The methods present two stages, where the first separates infected slices (by community-acquired pneumonia – CAP – or COVID-19) from the healthy ones through capsule networks, and the second performs a binary COVID-19 vs. non-COVID-19 classification. In phase two, pooling and voting give a final volume-level score in CT-CAPS and COVID-FACT, respectively.

Many approaches (Bougourzi et al., 2021a,b; Chaudhary et al., 2021; Garg et al., 2021; Li et al., 2021; Xue and Abhayaratne, 2021; Yang et al., 2021; Abdel-Basset et al., 2021) address a ternary classification problem where a chest CT volume is associated with one of the classes: “normal” (i.e., with no signs of pneumonia), CAP or COVID-19. Abdel-Basset et al. (2021) introduce algorithms for segmenting lung infections and classifying CT slices for this ternary problem. The authors adapt an EfficientNet (Tan and Le, 2019) architecture for extracting and combining features in multiple scales, further aggregated, and an attention mechanism that helps the model to focus on the infections. All the assessments presented by Abdel-Basset et al. (2021) are at slice level only.

Two-step volume-level classification mechanisms are described in (Chaudhary et al., 2021; Bougourzi et al., 2021a,b; Li et al., 2021). Similar to CT-CAPS and COVID-FACT, Chaudhary et al. (2021) firstly classify infected slices and then proceed to a ternary classification. The authors use a DenseNet (Huang et al., 2017) backbone for the first task and an EfficientNet (Tan and Le, 2019) for the second one. More precisely, the first stage uses annotated slices and transfer learning to classify the CT’s unseen central slices as infected (CAP or COVID-19) or healthy. The second stage refines the previous classification and produces the ternary score. A voting scheme is used for volume-level diagnosis. Bougourzi et al. (2021a) use slice-level annotations to train a convolutional neural network (CNN) for scoring the slices into three possible categories. A second stage aggregates the slices into 20 groups and produces the class probabilities for each group. This feature vector feeds an extreme gradient boost (XGBoost) classifier (Chen and Guestrin, 2016) for the final volume-level score. Bougourzi et al. (2021a) extended their analysis in (Bougourzi et al., 2021b). Li et al. (2021) use a slice-level, semi-supervised EfficientNet (Tan and Le, 2019) classifier based on FixMatch (Sohn et al., 2020) to produce probability scores for the ternary problem. Then, a sequence classifier based on AdaBoost (Hastie et al., 2009) computes weak per-slice classifiers weights for the ensemble. K-means clusters the features into regions and produces the final patient diagnosis by voting.

Like the previous works, Garg et al. (2021) propose a three-stage classifier. The first stage uses a ResNet (He et al., 2016) backbone and classifies CT slices as “normal”, CAP or COVID-19. The second stage uses the features from the former and trains four volume-level classifiers. These classifiers produce the class probabilities, combined in the third stage by a two-layer neural network that outputs the final score. Jin et al. (2020) address a four-class problem, involving COVID-19, CAP, influenza, and non-pneumonia (“healthy” lungs) classes. Their method performs slice-level lung segmentation (Ronneberger et al., 2015), per-class diagnosis, and infection location. The slice diagnosis explores a ResNet model (He et al., 2016), and the results are fused for volume-level scoring. The same architecture is used for COVID-19 infection localization but, in this case, trained with COVID-19 slices only. Guided GradCAM (Selvaraju et al., 2017) and t-SNE (Van der Maaten and Hinton, 2008) provide some explainability.

Ren et al. (2022) propose a hybrid model that employs both fixed and trainable filters – the former built from wavelet theory, arguing that less data is required for training after wavelet filtering. Li et al. (2020) propose a ResNet-based method, called COVNet, which can extract 2D local and 3D global features from chest CTs. COVNet receives as input a series of CT slices along with segmented lungs and outputs the probability scores for COVID-19, CAP, and non-pneumonia classes. Unlike the other

approaches discussed in this section, theirs do not use slice-level annotations.

Despite the variation in the backbone, classification strategy, and the number of stages and categories, most methods mentioned above require, at some point, per-slice annotations for training. During critical moments of the pandemic, experts are on the front lines and cannot afford time-consuming per-slice annotation tasks (Wu et al., 2021). We show that we can obtain competitive results even considering a smaller supervision level, i.e., by exploring volume-level annotations only.

2.2. 3D Volume-based Classification Methods

Wang et al. (2020) introduce a 3D CNN to detect COVID-19, which they call DeCoVNet. Their method inputs a chest CT volume and its extracted lung mask using a 3D U-Net (Ronneberger et al., 2015). DeCoVNet comprises a 3D vanilla CNN and residual blocks for feature extraction and 3D convolutions, poolings, and a fully connected layer for classification. Their method produces a binary probability score indicating that the patient is infected with COVID-19 or has healthy lungs.

Other works (Xue and Abhayaratne, 2021; Yang et al., 2021) tackle the ternary classification problem involving the “normal”, CAP, or COVID-19 classes but using 3D data and 3D classifiers. Xue and Abhayaratne (2021) firstly segment the lungs in a per-slice fashion via contrast adjustment and thresholding. Their method considers only the middle range of slices for further processing since, as the authors claim, they present more salient features associated with lung infections. This sub-volume is fed to a ResNet-like architecture (He et al., 2016) with 3D convolutions and poolings, pre-trained in a magnetic resonance imaging (MRI) imagery dataset. Yang et al. (2021) extract lung segments (Hofmanninger et al., 2020) and combine them with the pre-processed volume for feeding a pseudo-3D ResNet (He et al., 2016). Their CNN alternates 2D convolutions along the axes aligned with and orthogonal to the slice planes. The authors train six models using different data subsets to cope with data imbalance. The first model (first stage) allows discriminating normal cases from the others, which are further processed via an ensemble of the remaining models (second stage). Activations for the COVID-19 class rely on the second stage, and CAP scores result from combining the two stages.

Our brief literature review suggests that the number of volume-based methods is considerably smaller than slice-based ones. In general, training 3D CNNs is more complex and challenging (even in terms of hardware requirements) and particularly more challenging when the amount of available data is scarce (Jin et al., 2020; Li et al., 2021; Su et al., 2020). On the other hand, they require per-volume (not per-slice) annotations, which are much simpler and faster to obtain than individual slices, particularly during the pandemic, when radiologists are overwhelmed with exams and cannot afford time-consuming annotation processes.

In this work, we propose a compromise solution that maps a full CT volume onto an omnidirectional 2.5D representation of each lung. The proposed representation enables the efficiency of 2D processing, but instead of per-slice, it requires volume-level annotations only. We highlight that Su et al. (2020) also propose to represent and process 3D medical signals (not in the COVID-19 detection context) in 2D – which goes in the same vein as ours. Su and colleagues propose a spiral scanning approach that samples the 3D lung volume along preset latitudinal/longitudinal coordinates of a sphere circumscribed into the volume. The samples, which may relate to nodules, are captured from the center to the sphere surface in a given spiral direction and are arranged into a column of the output image. Columns are concatenated as the spiral scanning procedure advances from the topmost to the lowest spherical sample point. This design choice results in a periodic-like 2D signal where the latitudinal/longitudinal spiral coordinates are implicitly encoded along the horizontal dimension representing the sampling movement.

3. The Proposed Method

The proposed method takes a full volumetric chest CT as input and requires per-volume class annotations. The core idea is to summarize the full CT volume as a pair of two-channel spherical images using the equirectangular projection (ERP) (da Silva

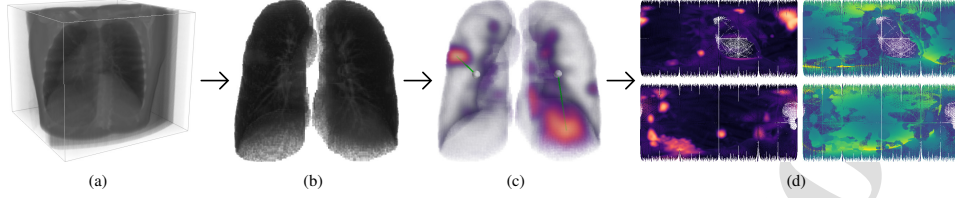


Fig. 1. Our omnidirectional imaging strategy: (a) input isotropic chest CT volume; (b) lungs segmentation; (c) estimation of GGO probabilities and positioning of the virtual cameras; (d) construction of the ERP maps with the projection of the maximum GGO probabilities and associated distances for each lung.

et al., 2022), one for each lung, as illustrated in Fig. 1. A preprocessing step is initially applied to generate isotropic voxels to account for different scanners or configurations, as shown in Fig. 1(a). Then, a 3D segmentation technique is used to identify the voxels related to each lung (Fig. 1(b)). In Figs. 1(a) and (b), transparency is applied for visualization purposes. Then, we produce a 3D “probability” map for each lung, indicating the presence of GGO lesions. We place a virtual spherical camera at the centroid of each lung, and for each camera, we cast rays outwards and retrieve the probability of the voxels mostly related to GGO and their corresponding distance from the camera. Fig. 1(c) shows the camera positioning (white sphere) and ray casting (green line segment) on the lung volumes represented by 3D GGO probabilities – low values are shown in higher transparency for better visualization. Finally, the spherical information sampled is represented in the plane using the ERP mapping. Fig. 1(d) depicts the probability (purplish) and distance (greenish) ERP images. Finally, the pair of two-channel images (registered probability and distance maps) are fed to a planar deep network (omitted in Fig. 1) that outputs the volume label. All these steps are detailed next.

3.1. Preprocessing and Lungs Segmentation

CT scanners might differ in terms of resolution (spatial and slice thickness) (Afshar et al., 2021), which impacts the voxel projection to the unit sphere. To alleviate this potential problem, we perform a simple volumetric interpolation as a preprocessing step to resample the input volume to an isotropic representation so that the voxels are 1cm^3 thick.

Since the CT volume contains the lungs, trachea, and other structures in the thoracic cavity, the next step of our method is to segment the lungs. Chest CT segmentation is a well-studied problem in the medical imaging community (Sousa et al., 2019; Hofmanninger et al., 2020; Ronneberger et al., 2015), and in this work, we adopted LungMask (Hofmanninger et al., 2020). LungMask is open-source² and has shown to be robust to many thorax abnormalities related to previous diseases, trauma, and surgeries, being suitable for patients with lung complaints. It segments CT volumes into the left lung, right lung, and background, from which we retrieve only the lungs. We used the default pre-trained model with no fine-tuning (unlike (Heidarian et al., 2020, 2021b)). It is worth noting that we also tested ALTIS (Sousa et al., 2019) for lung segmentation, but it failed to recover the segments when the lungs have large consolidation lesions.

3.2. Ground-glass opacity modeling

Our approach relates to the classical maximum intensity projection (MIP) (Jabeen et al., 2019). However, instead of manipulating raw intensities, we treat GGO-related probabilities, related to COVID-19 infection. We first generate a volumetric “probability map” related to each lung and, as discussed later, retrieve the maximum probability value and the corresponding normalized distance from the lung center. Our motivation for storing the probability and the distance is based on observations that GGO occurrences

²LungMask is available at <https://github.com/JoHof/lungmask>

associated with COVID-19 are bilateral and peripheral (Chate et al., 2020; Grando et al., 2020), which are captured by the proposed sampling approach (i.e., “depth” helps to discriminate interior from peripheral lesions).

Obtaining the GGO-related probabilities can be tackled from a fuzzy 3D GGO segmentation approach such as (Deen K et al., 2017), where the fuzzy values can be viewed as probability values or out-of-the-box solutions that produce COVID-19-related lung lesion maps through slice- or volume-based SOTA methods (Yang et al., 2022; Zhang et al., 2021). However, these approaches typically require a considerable amount of manually segmented CT data, and we adopted a simpler approach based on radiological findings associating COVID-19 infections and GGO intensities (in Hounsfield units – HU). According to Yu et al. (2020), GGO intensities typically lie in the range $\mathcal{R} = [-600, -200]$ HU, and we assume a Normal distribution for these values. More precisely, we define the mean $\mu = -400$ (the central point of the range) and standard deviation $\sigma = 121.95$, so that 95% of the distribution lies in \mathcal{R} , and considered the normalized values of the Gaussian as “GGO probabilities”.

After computing the GGO probabilities, we perform a 3D Gaussian filtering to suppress scattered mid-level probabilities associated with lung vessels and reinforce those related to actual GGO clusters. We do not consider consolidation probabilities since their patterns are indistinguishable from those associated with other respiratory diseases, as claimed by Laghi (2020). It is worth mentioning that we may not expect significant lesion self-occlusions since GGO tends to appear on the periphery of the lungs.

3.3. Omnidirectional image representation

The next step of our method is to summarize the volumetric information of each lung as a spherical image. The spherical imaging model allows capturing the information from the camera surrounding directions at once (Benseddik et al., 2016), and we adopt the ERP format to represent the spherical capture on a 2D rectangular domain. Such mapping of 3D data to a planar domain allows the use of conventional 2D deep architectures, as done in (Wang et al., 2018; Wu et al., 2018) for processing LiDaR data. We scan each pixel of the ERP image, which corresponds to a point on the unit sphere, and relates to a ray going from the sphere center (i.e., the lung center) outwards. Since these rays can cross several voxels, an adequate sampling strategy is required.

In our experiments, we set the ERP image resolution to 256×512 pixels, often found in other recent works combining spherical media and CNNs (Yang et al., 2019; Eder et al., 2019; Fernandez-Labrador et al., 2020). In fact, 512×1024 pixel images are often considered “high-resolution” in learning-based omnidirectional processing (Sun et al., 2021). Our sampling strategy produces a pair of probability-distance registered images, as shown in Fig. 1(d). Some of the rays do not hit any volume voxel during the sampling procedure, which leads to “missing renders”. Missing renders often appear near the poles (due to oversampling) and in the areas where the trachea attaches to the lungs, as shown in white in Fig. 1(d). We emphasize that missing renders may occur in other visual computing problems, such as depth capturing and scene modeling using panoramas (Zioulis et al., 2018; Stommel et al., 2014). Defining a preset value to encode these pixels using deep neural networks has shown promising results in related tasks (Zioulis et al., 2018). Here, we set the value zero for both probability and distance at pixel locations without corresponding voxels. We also experimented with 128×256 and 512×1024 ERP images. However, we noted that some GGO lesions were too small in the former and that the number of “missing renders” was too high in the latter. Additionally, processing the higher-resolution images imposes strong memory requirements for the GPU we have to train our model. An example of sampling with different ERP image resolutions is depicted in Fig. 2. Section 5.1 revisits this discussion.

We propose capturing more than one spherical image per subject to account for the imperfections in lung segmentation and tackling the irregular sampling in ERP format. We produce additional full-FoV virtual views by adding a small perturbation on the lungs’ centroid computation before imaging – we experimentally set the perturbation to 5% of the signal amplitude in each dimension. We found that the centroids computed from our two databases’ CT volumes do not vary more than 5% around the

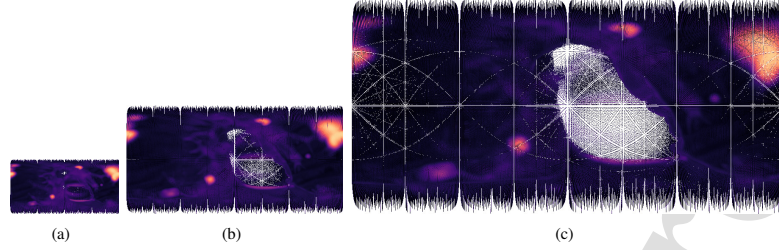


Fig. 2. The impact of varying the ERP image resolution during sampling: (a) 128×256 , (b) 256×512 , and (c) 512×1024 pixels. Example images store the maximum GGO probabilities associated with one lung.

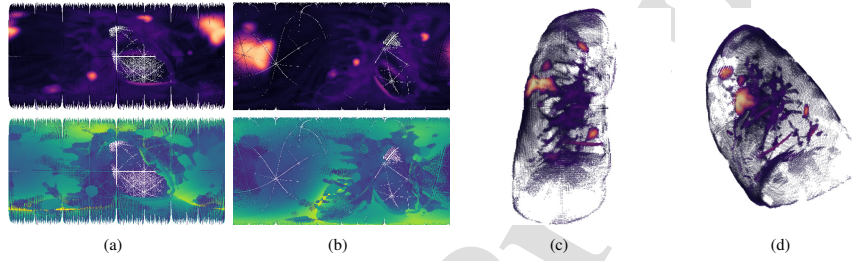


Fig. 3. Visualization of the same lung using different virtual cameras: (a)-(b) ERP images with GGO probabilities and distances; and (c)-(d) point clouds with the GGO probabilities in their 3D spatial location.

mean centroid position. We also apply random virtual camera rotations before imaging to encourage rotation invariance in our classification procedure since any spherical camera orientation produces the same information (da Silveira et al., 2022). This procedure helps tackling the irregular sampling of ERP images, network regularization, and alleviates overfitting. Hence, we capture N_s different images for the same volume by randomly varying the virtual camera center and orientation.

An example of our volume-based augmentation process is shown in Fig. 3. Figs. 3(a) and 3(b) depict the ERP images with the GGO probabilities and distances for the same volume (related to a single lung) but with different virtual camera parameters. For better visualization, the same information is also shown as a colored point cloud in Figs. 3(c) and 3(d), where low probabilities are displayed with higher transparency. Note that moving the virtual camera entails another sampling and underlying sparse 3D representation of the same lung information. Our ablation study, detailed in Section 5, indicates that $N_s = 5$ is a good choice. These N_s images are used for training. Only the unrotated (“canonical”) capture is used for inference.

3.4. The Classification Model

The approach described so far summarizes a full CT chest volume in a pair of two-channel 256×512 ERP images. Each image relates to a different lung, and the pixel values contain the maximum GGO probability and the corresponding normalized distance for the 3D ray associated with the pixel.

Since the two input image pairs relate to the lungs (left and right), it is natural to apply the same feature extractor to both images for retrieving features that relate (or not) to COVID-19 patterns. Hence, we explored a siamese backbone in our approach (yellow block in Fig. 4), where two identical blocks share the same weights. The siamese backbone is applied to each lung in parallel, and the feature maps are vertically concatenated in the end. By concatenating the feature maps in the vertical dimension (and not along the channels), we preserve the spatial location of the feature maps related to both lungs, which is important due to the bilateral

nature of GGO lesions in COVID-19 patients (Chate et al., 2020; Grando et al., 2020). Furthermore, it allows the visualization of activation maps for each lung separately, as we will describe next.

The classification head is fed by the concatenated feature map produced by the siamese backbone, and it produces as output a C -channel tensor with the class-related probabilities, where $C = 3$ for the ternary problem (normal, CAP, and COVID-19) or $C = 2$ for the binary problem (normal and COVID-19). Since visualizing the per-class activations is relevant in terms of explainability, we adopted an approach based on Class Activation Maps (CAMs) (Zhou et al., 2016) instead of using a fully connected layer as in Tan and Le (2019). However, we also explored the basic concept in Squeeze-and-Excite (SE) layers Hu et al. (2018), in which each channel's global "importance" is used to re-weight the feature map. Our secondary contribution is a novel classification head, named SE-CAM, that branches off from the joint feature map produced by the backbone. One branch performs Global Average Pooling (GAP) (Y. Lecun et al., 2015) across the spatial dimensions of the map, fed to two Fully Connected (FC) (Y. Lecun et al., 2015) layers that produce C neurons at the end with sigmoid activation to produce normalized responses. The second branch performs $C \times 3 \times 3$ convolutions with ReLU activation (Y. Lecun et al., 2015), yielding a C -channel CAM-like feature map. The outputs of both branches are multiplied in a channel-wise fashion, producing the SE-CAM layer. Finally, a GAP layer with softmax activation (Y. Lecun et al., 2015) is used to obtain the per-class probabilities, noting that each channel of the SE-CAM provides the spatial per-class activation maps: the top half relates to the left lung and the bottom half to the right lung. The SE-CAM block scheme is shown (in green) in Fig. 4.

Unlike large generic-purpose classification datasets such as ImageNet (Deng et al., 2009), the number of training samples is typically limited in CT-based COVID-19 datasets. Also, our input is a pair of two-channel images containing GGO probabilities and distances mapped to an ERP image. Since the nature of our data differs from most common datasets used for pre-training backbones (such as the number of channels, visual characteristics and representation in equirectangular format), we need to train it from scratch. Training details are provided in Section 4.2. To select the backbone of our siamese network, we started with an EfficientNet model (Tan and Le, 2019), which presents a good compromise between the number of parameters and classification accuracy in generic datasets, and in particular has been used for image-based COVID-19 classification (Abdel-Basset et al., 2021; Chaudhary et al., 2021; Li et al., 2021). We started with the smallest model (EfficientNet-B0), and the model produced low train accuracy values, which indicates problems with convergence. We also tested EfficientNet-B1 and had convergence issues, which led us to investigate smaller backbones. We then proceeded to a simple Neural Architecture Search (NAS) (Elsken et al., 2019) based on block pruning on top of EfficientNet-B0. We noted that the validation accuracy peaked when we used the first four blocks (details on the architecture pruning are given in Section 5). Since the output of the fourth block in EfficientNet-B0 presents only 80 channels (compared to 1,280 in the full model), we added a convolutional layer with $256 \times 3 \times 3$ kernels in the end for feature expansion to accommodate for training-time data augmentation, as explained in Section 4.2. A visual representation of the backbone is provided in the top-right of Fig. 4.

Recall that each input volume (CT scan from a patient) produces a set of N_s spherical image pairs that differ because of random rotations and small virtual camera displacements. When training the network, all these augmented versions are used to improve generalization and achieve better invariance to rotations and sampling. Only the canonical view obtained directly from the CT volume is used in the inference phase. However, we explored a *snapshot ensemble* (Huang et al., 2017), which consists of computing the average of the softmax outputs of the model using the weights at different stages of the training process. More precisely, we used a set of five weights to compute the ensemble related to epochs 60, 55, 50, 45, and 40 – more training details are provided in Section 4.2.

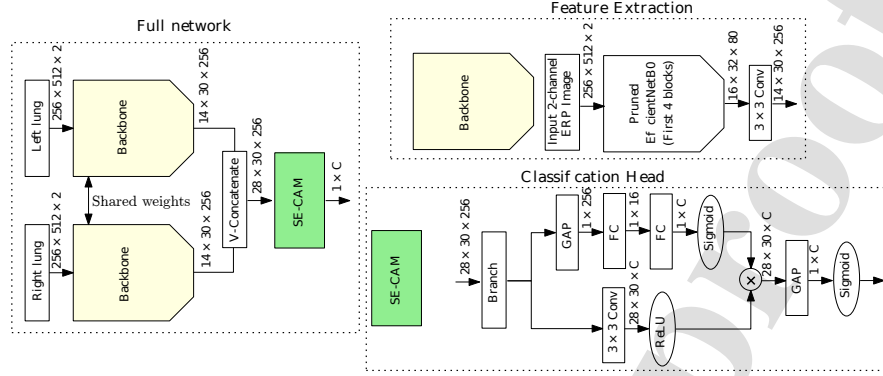


Fig. 4. Overview of our two-input deep network

4. Results and Discussion

4.1. Datasets

This work considers a publicly available dataset and an in-house dataset (CTs from two hospitals) for training and testing. These two datasets are described next.

4.1.1. COVID-CT-MD/SPGC-COVID dataset

For a comprehensive comparison with the SOTA, we consider the dataset associated with the Signal Processing Grand Challenge (SPGC) on COVID-19 Diagnosis (Mohammadi et al., 2021) held by the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP) in 2021. The dataset contains volumetric chest CT scans of “normal” patients and subjects infected with CAP and COVID-19 and divides into two: the COVID-CT-MD dataset (Afshar et al., 2021) and the SPGC-COVID dataset (Heidarian et al., 2021a). The former provides the train and validation sets, whereas the latter delivers three test sets. The validation set cannot be used for training purposes. We adopt in our work the benchmark proposed for the SPGC on COVID-19 Diagnosis (Mohammadi et al., 2021).

The COVID-CT-MD dataset (Afshar et al., 2021) contains CT scans of 76 normal patients, 60 patients with CAP infections, and 169 with COVID-19 collected in the Babak Imaging Center, Tehran, Iran. All chest CTs were acquired with a Siemens Somatom Scope scanner with 2mm slices, 50mAs, 110-130kV, and roughly 7mSv provided in DICOM format. Slice-level annotations are provided for a subset of the CT scans, but our approach explores only the CT volume label. We consider the preset test and validation dataset splits that contain 207 and 98 samples, respectively.

Three subsets of the SPGC-COVID dataset (Heidarian et al., 2021a) were used in the SPGC on COVID-19 Diagnosis. The first test set has 15 scans of normal and 15 COVID-19 cases from the Babak Imaging Center, Tehran, Iran. Scanning parameters such as reference exposure, kVp, and radiation differ from the training set. The second test set contains ten normal cases, ten CAP cases, and ten COVID-19 cases. The data is captured by a Siemens Somatom Emotion 16 scanner in Tehran Heart Center, Tehran, Iran, with varying slice thickness, reference exposure, and radiation regarding the training set. Some CT scans in the second test set have imaging findings related to cardiovascular diseases or surgeries which are not available in the training set. Finally, the third test has ten normal cases, ten CAP cases, and ten COVID-19 cases and uses the same scanning protocols as the training set. The results are averaged from the three test sets as suggested in the Challenge (Mohammadi et al., 2021).

Although the published results linked to the SPGC on COVID-19 Diagnosis focus on the ternary classification problem, the Challenge organizers also promote a binary – normal vs. COVID-19 – analysis (Mohammadi et al., 2021). We refer the reader to the papers that release the datasets (Afshar et al., 2021; Heidarian et al., 2021a) for more information. Here, we analyze the performance of our approach in a ternary classification problem, adopted in many related works (Chaudhary et al., 2021; Garg et al., 2021; Yang et al., 2019; Xue and Abhayaratne, 2021; Bougourzi et al., 2021a,b; Ewen and Khan, 2021; Li et al., 2021).

4.1.2. In-house dataset

We also consider an in-house private dataset composed of CT scans from two hospitals in Porto Alegre, RS, Brazil: Hospital Moinhos de Ventos (HMV) and Hospital de Clínicas de Porto Alegre (HCPA). The studies involving the in-house HCPA and HMV sub-datasets have ethics board protocol numbers 32314720.8.0000.5327 and 32314720.8.3001.5330 and were approved on 16 June 2020 and 22 July 2020, respectively. Informed consent was waived due to the retrospective nature of the studies.

HMV chest CTs were collected using different scanners, namely Siemens Emotion 16, Siemens Somatom Drive, and Siemens Sensation 64, and the files are provided in DICOM format. Scanning parameters are set to 0.75–1.0mm slices, 46–171mAs of exposure, 100–130kV peak kilovoltage (kVp), spiral pitch factor 0.4–1.5, and collimation width 0.60mm. The dataset is well-balanced in terms of gender: 51.93% of the patients are male, and 48.07% female. From the total of 181 chest CTs, 90 were labeled as “typical”, and 91 as “negative” for COVID-19 pneumonia based on radiological findings according to the Radiological Society of North America (RSNA) expert consensus (Simpson et al., 2020; Grando et al., 2020). Patients infected with COVID-19 may not present lung-related imaging findings (they can even be asymptomatic), which justifies the adoption of the RSNA consensus classes instead of pure RT-PCR scores.

Chest CTs from HCPA were captured using the GE Medical Systems BrightSpeed S and Toshiba Aquilion scanners. Slices thickness and exposure vary between 1.25–2.5mm, and 5–275mAs, respectively, spiral pitch factor is set to 1.35–1.67, collimation width is 1.25–2.50mm, kVp varies in 100–140kV, and the data is also provided in DICOM format. From a total of 45 scans, 30 were labeled as “typical” and 15 as “negative” for COVID-19 pneumonia. The chest CTs correspond to 21 male and 24 female patients aged 56.22 ± 15.10 years.

Unlike the analysis using the COVID-CT-MD/SPGC-COVID dataset, this one considers the binary problem where the classes are defined as “typical” and “negative”, meaning COVID-19 positive and negative diagnoses. This classification scenario is helpful for initial patient screening and is also addressed in many related works (Zhang et al., 2020; Wang et al., 2020; Heidarian et al., 2020, 2021b).

4.2. Training Details

As mentioned in Section 3.4, our model is trained from scratch. We used the exponential moving average variant of Adam (Reddi et al., 2019) with $\beta_1 = 0.9$ and $\beta_2 = 0.999$. We used a learning rate scheduler that starts with an exponential warm-up in the first quarter of training, reaching a value of 10^{-3} . It is kept constant during the second quarter and reduces to 10^{-4} in the second half of training. The classifier is trained by 60 epochs, noting that the validation accuracy starts stabilizing by epoch 40.

Even using N_s rotated versions of each volume, the number of training samples in existing COVID-19 datasets with CT volumes is limited. Hence, we relied on training-time data augmentation besides the 3D translated and rotated versions described in Section 3.3. As augmentation primitives, we used random circular horizontal shifts and added random noise and blur. The former relates to rotations in the spherical image in the direction of the equator, whereas the latter emulates uncertainties in the GGO probability and distance computations. More precisely, we used Gaussian noise and Gaussian blur, where both σ_{noise} and σ_{blur} are selected based on a uniform distribution $\mathcal{U}(0, 3)$. We also explore cutout (DeVries and Taylor, 2017) to alleviate overfitting and

Table 1. Average results in the COVID-CT-MD/SPGC-COVID dataset

Method	Sensitivity (%)			Accuracy (%)
	COVID-19	CAP	Normal	
Chaudhary et al. (2021)	85.71	90.00	94.29	90.00
<i>Our technique</i>	94.29	75.00	91.43	88.89
Garg et al. (2021)	88.57	90.00	88.57	88.89
Yang et al. (2021)	80.00	90.00	100.00	87.78
Xue and Abhayaratne (2021)	82.86	80.00	91.43	85.56
Bougourzi et al. (2021a,b)	91.43	45.00	91.43	81.11
Ewen and Khan (2021)	88.57	35.00	97.14	80.00
Li et al. (2021)	88.87	35.00	97.14	80.00

improve regularization, which consists of “erasing” rectangular patches in the input images. Patch sides are sized to 35% of the input image dimensions, set empirically. Each of these four primitives is applied separately at training time for each lung, with individual probabilities of 50%.

To deal with potential class imbalance that could bias the classifier, we explored a weighted categorical cross-entropy loss (Cui et al., 2019; Ho and Wookey, 2020). This issue occurs mainly in the COVID-CT-MD/SPGC-COVID dataset, where the number of COVID-19 samples is more than twice the amount of CAP and normal samples. The weight of each class is given by N_i/N_c , where N_i is the total number of training samples, and N_c is the number of samples in class c (Cui et al., 2019). All experiments related to our model were conducted in modest hardware: an i7 3.2GHz PC with 16GB RAM and an NVIDIA RTX 2070 GPU, using a batch size of eight.

4.3. Comparison with the SOTA

The COVID-CT-MD/SPGC-COVID dataset provides per-slice annotations, but our approach used only higher-level information (the per-volume/patient label). Our in-house dataset is labeled for a binary classification problem and presents only per-volume annotations, limiting the comparison with competitive techniques requiring per-slice annotations. We present our results and comparative analysis for each dataset next.

4.3.1. COVID-CT-MD/SPGC-COVID dataset

COVID-CT-MD/SPGC-COVID dataset allows a comprehensive comparison of our and peering approaches for the ternary classification problem. More precisely, we compare the proposed approach with SOTA techniques (Chaudhary et al., 2021; Garg et al., 2021; Yang et al., 2019; Xue and Abhayaratne, 2021; Bougourzi et al., 2021a; Ewen and Khan, 2021; Li et al., 2021; Bougourzi et al., 2021b) that recently published their results on the SPGC on COVID-19 Diagnosis. All the methods compared in this section were trained, validated, and tested with the same dataset and splits as ours. We emphasize that only the works (Yang et al., 2019; Xue and Abhayaratne, 2021) can deal with volume-level annotations, while the remaining approaches (Chaudhary et al., 2021; Garg et al., 2021; Bougourzi et al., 2021a; Ewen and Khan, 2021; Li et al., 2021; Bougourzi et al., 2021b) explore at some point in their pipeline slice-level annotations. We consider the same training, validation, and testing protocol as the Challenge (Mohammadi et al., 2021).

Class sensitivity and accuracy for the ternary problem involving normal, CAP, and COVID-19 classes are summarized in Table 1, and the approaches are ranked based on the global accuracy as suggested by Mohammadi et al. (2021). No other performance measurement is publicly available for these methods. According to the accuracy metric, our approach performs better than Yang et al. (2019); Xue and Abhayaratne (2021), which process 3D data and are more expensive in terms of computational resources. Our technique outperforms several SOTA methods (Ewen and Khan, 2021; Li et al., 2021; Bougourzi et al., 2021a,b; Garg et al.,

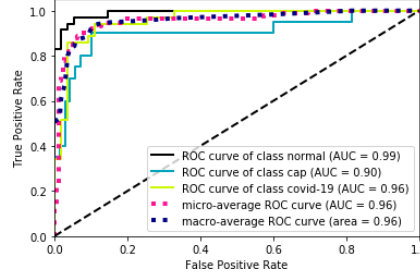


Fig. 5. Per-category ROC curves for the COVID-CT-MD/SPGC-COVID dataset

Table 2. Per-fold and average test results for the in-house dataset

Method	Sensitivity (%)		Accuracy (%)
	Typical	Negative	
<i>Our technique</i>	Fold 1	91.67	100.00
	Fold 2	96.00	95.65
	Fold 3	100.00	100.00
	Fold 4	100.00	100.00
	Fold 5	92.00	93.75
	Mean	95.93	96.99
	Std.	3.65	2.55
DeCoVNet	Fold 1	96.00	93.47
	Fold 2	96.00	95.65
	Fold 3	100.00	91.30
	Fold 4	100.00	91.11
	Fold 5	88.00	88.88
	Mean	96.00	92.08
	Std.	4.38	2.30
COVNet	Fold 1	88.00	90.70
	Fold 2	75.00	81.98
	Fold 3	79.00	83.72
	Fold 4	88.00	89.13
	Fold 5	75.00	96.00
	Mean	81.00	86.12
	Std.	5.89	3.28

2021) and presents slightly lower accuracy results than Chaudhary et al. (2021), noting that all these approaches require per-slice annotations. We highlight that our approach outperformed all techniques in COVID-19 sensitivity, with 33 out of 35 COVID-19 CTs correctly predicted. High COVID-19 sensitivity is paramount for diagnosis since patients with a false negative diagnosis might not be adequately treated/observed. Our performance for the CAP class is inferior to top-ranking methods, which might be explained since we explicitly explore GGO probabilities as input to our network – and these lesions might also appear in patients infected with CAP (Di et al., 2021). As a final result for this dataset, we show the per-category receiver operating characteristic (ROC) curves (Melo, 2013) along with the corresponding areas under the curve (AUC) in Fig. 5.

4.3.2. In-house dataset

Our dataset does not present preset train/validation/test sets. Hence, we adopted a 5-fold cross-validation strategy for all evaluated approaches, reporting the per-fold test and overall accuracies as in the COVID-CT-MD/SPGC-COVID dataset. Considering

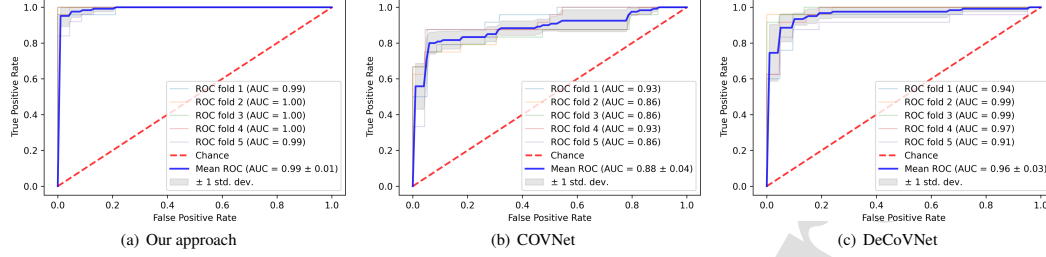


Fig. 6. ROC curves for the typical class in the in-house dataset.

existing methods that provide source code and work with volume-level annotation, we identified COVNet (Li et al., 2020)³ and DeCoVNet (Wang et al., 2020) as the leading competitors. We emphasize that the methods that report results for the COVID-CT-MD/SPGC-COVID dataset do not have public code or require slice-level annotations. We followed the training guidelines provided in the original papers (Li et al., 2020; Wang et al., 2020) to train COVNet and DeCoVNet in our dataset. For training these two models, we had to use another hardware with a better GPU due to memory constraints: an Intel Core i5-9400F 2.90GHz PC with 16GB RAM and a 24GB NVIDIA RTX 3090 GPU. Per-fold average training time for COVNet (50 epochs) and DeCoVNet (200 epochs) were 3.04h and 18.3h, respectively. Although Wang et al. (2020) suggest using 100 epochs for training, we noted that using additional 100 epochs improved both the training loss and the test accuracy for DeCoVNet. For comparison, the proposed model takes 10 minutes to train each fold (60 epochs) using a modest GPU (NVIDIA RTX 2070). COVNet and DeCoVNet take 248.48 and 35.61 GFLOPs, respectively, against 2.03 GFLOPs of our approach.

The results produced by our method and peering approaches are summarized in Table 2. The best results are highlighted in bold. Our approach presented better results than COVNet for almost all folds in both classes. Compared to DeCoVNet, our approach produced similar per-fold and average sensitivity for typical samples but considerably better values for negative samples. Our approach achieves a better average accuracy compared to COVNet and DeCoVNet, and a Friedman test (Friedman, 1937) indicates that our approach significantly outperforms COVNet and DeCoVNet across the folds (p -value < 0.007). As a final quantitative analysis, we show the ROC curves for the typical class across the five test folds for all three methods in Fig. 6. We can observe that the AUC values produced by our approach were higher than competitive approaches for all folds, with an average AUC of 0.99 and a low standard deviation of 0.01.

4.4. Visualizing Class Activations

As explained in Section 3.4, the SE-CAM module in our network provides per-lung Class Activation Maps, which allow the visualization of the spatial regions that contribute to each class. Fig. 7 shows examples of input images as point clouds (GGO probabilities mapped to the corresponding distances to generate a 3D point cloud) along with the per-lung CAMs for four patients mapped to the same 3D locations. The first column shows a true negative, and the CAM presents a solid and spatially-distributed response in both lungs for the negative class. The second shows a true positive result, and the CAM presents localized activations related to larger GGO probabilities. The last two columns illustrate errors: in Fig. 7(c), the CAM for class typical did respond to GGO on the left lung, but the overall contribution of the negative-related CAM on the right lung is higher; in Fig. 7(d) the typical-related CAM responds strongly for a small GGO region on the top of the right lung, generating a false positive. Fig. 7 shows a

³We adapted the COVNet architecture to the binary classification problem.

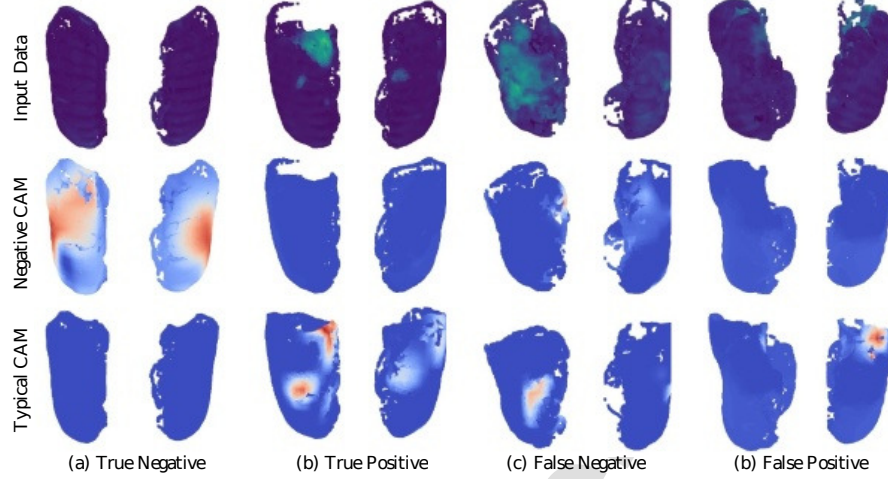


Fig. 7. Examples of CAMs for the in-house dataset. Top: input data. Bottom: CAMs for negative and typical classes.

canonical view of the CAMs, but the physician can freely rotate the 3D point clouds for better visualization. Although the example was given for the binary classification problem, it also applies to the ternary.

5. Ablation Study

In this section, we provide more details to justify the selection of the proposed network and training parameters. All parameters were set based on experiments with the COVID-CT-MD dataset (validation subset), and the same parameters were used in the in-house dataset. Our ablation studies explore changing the backbone, the number of views N_s in our offline augmentation, and the classification head. In all the experiments, we considered the snapshot ensemble approach explained earlier with the model weights trained at epochs 40, 45, 50, 55, and 60.

5.1. Backbone Selection and Input Resolution

As mentioned in Section 3.4, we started with an EfficientNet-B0 backbone and trained it with the COVID-CT-MD training set (using SE-CAM head and $N_s = 1$), leading to low train and validation accuracies that indicate poor convergence, which we hypothesize is due to a large number of parameters w.r.t. the size of the training set. We then progressively reduced the number of inverted bottleneck MBConv blocks in the model and noticed that both train and validation accuracy increased up to a point and started to decrease. We performed this experiment using the selected ERP resolution of 256×512 and a reduced version with 128×256 pixels. Table 3 shows the results of this experiment with the corresponding model size – number of parameters – and we can see that selecting four blocks yields a peak in the validation accuracy chosen as the baseline. We also note that using lower-resolution ERPs yields considerably smaller training accuracies, but the gap is smaller when considering the validation accuracy. Nevertheless, the default resolution achieves the highest validation accuracy. For the sake of comparison, using a traditional backbone such as ResNet50 generates a model with 25B parameters, which also does not converge. As mentioned earlier, we also faced convergence issues using the EfficientNet-B1 model.

Our ablation studies regarding the backbone size corroborate the expected relationship between the number of training data and model size. Although the network design experiments were performed using the COVID-CT-MD dataset, results for the in-house

Table 3. Train and validation accuracy for COVID-CT-MD/SPGC-COVID dataset varying the number of blocks in EfficientNet-B0 and the input size: 256×512 (default, or D), and 128×256 (reduced, or R).

No. of blocks	No. of params.	Train accuracy (D/R)	Validation accuracy (D/R)
2	115,768	78.00/37.00	71.42/55.93
3	200,968	88.00/83.00	76.53/76.25
4	541,338	91.50/85.50	79.59/78.12
5	1,166,182	95.50/81.00	77.55/77.95
6	3,394,898	82.50/73.00	71.42/71.77
7	7,038,850	26.50/25.00	13.26/24.58

Table 4. Validation accuracy for COVID-CT-MD/SPGC-COVID dataset when varying N_s in the pruned EfficientNet-B0

N_s	Validation accuracy
1	77.55
5	81.63
9	79.59

Table 5. Validation accuracy for COVID-CT-MD/SPGC-COVID dataset when varying the classification head in the pruned EfficientNet-B0

Classification head	Validation accuracy
FC	79.59
CAM	81.63
SE-CAM	81.63

dataset shown in Section 4.3.2 indicate that the same architecture is suitable for smaller datasets.

5.2. Size of the Offline Data Augmentation

Different views from the same volume entail different sampling in the image equator and poles, as shown in Fig. 3, and might make GGO lesions more or less evident during training. We consider the optimal backbone from the previous experiment, a four-block pruned EfficientNet-B0, with the SE-CAM classification head and varied the number of captures in the offline augmentation (N_s). Increasing the value of N_s helps with model generalization, but large values might preclude convergence and increase training time. Table 4 shows that $N_s = 5$ is a compromise solution.

5.3. Classification Head Selection

This experiment uses the pruned EfficientNet-B0 and $N_s = 5$ and evaluates the effect of changing the classification head from the original FC layer of the EfficientNet family with the proposed SE-CAM and the CAM layer (Zhou et al., 2016). Table 5 shows that CAM and SE-CAM perform the same and better than FC in the validation set. However, we noticed that SE-CAM performs better in the test sets assessments (88.89% vs. 87.78% in overall accuracy). It also occurs in the experiments with the in-house dataset, where the average accuracy using CAM was $96.92 \pm 2.71\%$, compared to $96.99 \pm 2.55\%$ using SE-CAM, which shows a slight advantage for SE-CAM.

5.4. Dataset size

Inspired by Su et al. (2020), this experiment aims to assess the impact of the training set size on the performance of the chosen classifier in previous experiments. Note that our initial training set is already small (containing only 207 training samples), whereas the baseline training set in (Su et al., 2020) was considerably larger.

We reduce our original training set size by a half, a quarter, and an eighth in a stratified manner (i.e., maintaining the proportion of each class), while keeping the test set fixed. Since this experiment propositionally reduces the dataset size, we show in Table 6 the results using the proposed architecture but considering $N_s = 1$. One can note that the classifier accuracy drops compared to the

Table 6. Train and validation accuracy for COVID-CT-MD/SPGC-COVID dataset reducing the training set size

Reduction factor	Train accuracy	Validation accuracy
1/2	77.88	70.41
1/4	77.08	68.37
1/8	75.00	24.49

entire dataset size (91.50% and 79.59% of train and validation accuracies, as shown in the third row of Table 3), as expected. In particular, the reduction factor of 1/8 is very extreme, leading to low validation accuracy.

6. Challenges and Limitations

One challenge we might face when dealing with CT scans from multiple sources is the variable volume resolution. Indeed, only for the COVID-CT-MD/SPGC-COVID dataset, we find CT volumes with sizes ranging from $276 \times 276 \times 242$ to $366 \times 366 \times 595$ voxels. As explained in Section 3.1, we try to alleviate this problem by performing a volumetric interpolation to the input volume so that it becomes isotropic and its voxels are 1cm^3 thick. However, volumes with a small number of slices might not benefit from the proposed spherical projection.

A potential limitation is that the proposed approach depends on lung segmentation. Perfect lung segmentation is not crucial for our method, and we add a small camera translation in the offline data augmentation procedure to cope with possible imperfect segmentations. However, as argued in Section 3.1, some lung segmentation approaches may fail in highly compromised lungs, such as those with large consolidation lesions. We found that ALTIS fails to recover the lungs from CTs in 3.25% of the train/validation set and 8.89% of the test set from the COVID-CT-MD/SPGC-COVID dataset. For this dataset, when considering ALTIS for lung segmentation, our approach reaches an overall accuracy of 85.56% compared to 88.89% using LungMask. In our experiments with LungMask, however, we did not suffer from this problem.

Another possible limitation of our approach is relying on a specific range (in Hounsfield units) for the GGO. Although radiologic finds on COVID-19 are linked to GGO, other diseases, including CAP, can also present similar lesions. Fortunately, the spatial distribution of these lung lesions (they are bilateral and peripheral when caused by COVID-19) might help the classifier identify the diseases (Chate et al., 2020; Grando et al., 2020). Our results in both the binary and ternary problems show that our choice of modeling GGO probabilities and distances from the lesions to the lungs' center leads to accurate classification results.

7. Conclusion

This work introduced a 2.5D omnidirectional representation for volumetric chest CTs applied to COVID-19 diagnosis, enabling efficient 2D deep learning mechanisms and requiring only volume-level annotation. Our approach segments the lungs from the volumetric chest CTs, obtaining the voxels related to each one. We produce a 3D probability map of GGO infections for the lungs and retrieve the omnidirectional maximum probability projection and the associated distance to the lung center. This process generates a two-channel spherical image per lung, mapped to the plane by ERP projection. The pair of two-channel ERP images feed a deep classifier based on a siamese backbone and a novel SE-CAM classifier head.

We tested our approach using publicly available and in-house datasets. For the COVID-CT-MD/SPGC-COVID dataset, our results are competitive with SOTA methods that require more elaborate per-slice annotations, yielding the highest recall rate for COVID-19 samples. Our method is more straightforward than the peering approaches tested in our in-house dataset and produces better results.

Although this work focused on chest CTs for COVID-19 diagnosis, we believe that a similar approach can also be explored in other problems, such as lung nodule detection, and imaging modalities where dense 3D volumes can be reduced to representative samples and mapped to ERP maps, such as in MRI and magnetoencephalography. We also plan to explore other neural architectures that can better handle the distortions in ERP images, such as transformer-based layers (Liu *et al.*, 2022).

Acknowledgments

We thank the financial support from Fundação de Amparo à Pesquisa do Estado do Rio Grande do Sul (FAPERGS), Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), and Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) – Finance Code 001, Brazil.

References

- 1 K. Kupferschmidt, New coronavirus variants could cause more reinfections, require updated vaccines, *Science* (2021).
- 2 E. Callaway, Fast-evolving COVID variants complicate vaccine updates, *Nature* 607 (2022) 18–19.
- 3 M. Roberts, D. Driggs, M. Thorpe, J. Gilbey, M. Yeung, S. Ursprung, A. I. Aviles-Rivero, C. Etmann, C. McCague, L. Beer, *et al.*, Common pitfalls and recommendations for using machine learning to detect and prognosticate for COVID-19 using chest radiographs and CT scans, *Nature Machine Intelligence* 3 (2021) 199–217.
- 4 S. Xue, C. Abhayaratne, COVID-19 diagnostic using 3D deep transfer learning for classification of volumetric computerised tomography chest scans, in: *IEEE International Conference on Acoustics, Speech and Signal Processing*, 2021, pp. 8573–8577.
- 5 T. C. Kwee, R. M. Kwee, Chest CT in COVID-19: What the radiologist needs to know, *RadioGraphics* 40 (2020) 1848–1865.
- 6 S. Heidarian, P. Afshar, N. Enshaei, F. Naderkhani, A. Oikonomou, S. F. Atashzar, F. B. Fard, K. Samimi, K. N. Plataniotis, A. Mohammadi, M. J. Rafiee, Covid-fact: A fully-automated capsule network-based framework for identification of COVID-19 cases from chest CT scans, *Frontiers Media SA* 4 (2020) 1–16.
- 7 Z. Yang, Y. Hou, Z. Chen, L. Zhang, J. Chen, A multi-stage progressive learning strategy for COVID-19 diagnosis using chest computed tomography with imbalanced data, in: *IEEE International Conference on Acoustics, Speech and Signal Processing*, 2021, pp. 8578–8582.
- 8 B. Li, Q. Zhang, Y. Song, Z. Zhao, Z. Meng, F. Su, Diagnosing Covid-19 from CT Images Based on an Ensemble Learning Framework, in: *IEEE International Conference on Acoustics, Speech and Signal Processing*, 2021, pp. 8563–8567.
- 9 H. Hassan, Z. Ren, H. Zhao, S. Huang, D. Li, S. Xiang, Y. Kang, S. Chen, B. Huang, Review and classification of AI-enabled COVID-19 CT imaging models based on computer vision tasks, *Computers in Biology and Medicine* 141 (2022) 105123.
- 10 X. Wang, X. Deng, Q. Fu, Q. Zhou, J. Feng, H. Ma, W. Liu, C. Zheng, A weakly-supervised framework for COVID-19 classification and lesion localization from chest CT, *IEEE Transactions on Medical Imaging* 39 (2020) 2615–2625.
- 11 M. F. Rahman, Y. Zhuang, T.-L. B. Tseng, M. Pokojovy, P. McCaffrey, E. Walser, S. Moen, A. Vo, Improving lung region segmentation accuracy in chest x-ray images using a two-model deep learning ensemble approach, *Journal of Visual Communication and Image Representation* 85 (2022) 103521.
- 12 G. de Souza Portes Meirelles, COVID-19: a brief update for radiologists, *Radiologia Brasileira* 53 (2020) 320–328.
- 13 R. C. Chate, E. K. U. N. Fonseca, R. B. D. Passos, G. B. d. S. Teles, H. Shoji, G. Szarf, Presentation of pulmonary infection on CT in COVID-19: initial experience in Brazil, *Jornal Brasileiro de Pneumologia* 46 (2020).
- 14 Q. Yu, Y. Wang, S. Huang, S. Liu, Z. Zhou, S. Zhang, Z. Zhao, Y. Yu, Y. Yang, S. Ju, Multicenter cohort study demonstrates more consolidation in upper lungs on initial ct increases the risk of adverse clinical outcome in covid-19 patients, *Theranostics* 10 (2020) 5641–5648.
- 15 J. Chen, S. Peng, B. Zhang, Z. Liu, L. Liu, W. Zhang, An uncommon manifestation of COVID-19 pneumonia on CT scan with small cavities in the lungs, *Medicine* 99 (2020) e21240.
- 16 P. Afshar, S. Heidarian, N. Enshaei, F. Naderkhani, M. J. Rafiee, A. Oikonomou, F. B. Fard, K. Samimi, K. N. Plataniotis, A. Mohammadi, COVID-CT-MD, COVID-19 computed tomography scan dataset applicable in machine learning and deep learning, *Scientific Data* 8 (2021) 1–8.
- 17 R. D. Grando, V. B. Brentano, A. P. Zanardo, F. T. Hertz, L. C. A. Júnior, J. F. P. dos Santos, G. S. Galvão, A. P. Zavascki, M. B. Gazzana, Clinical usefulness of tomographic standards for COVID-19 pneumonia diagnosis: Experience from a brazilian reference center, *The Brazilian Journal of Infectious Diseases* 24 (2020) 524–533.
- 18 P. Garg, R. Ranjan, K. Upadhyay, M. Agrawal, D. Deepak, Multi-Scale Residual Network for COVID-19 Diagnosis Using CT-Scans, in: *IEEE International Conference on Acoustics, Speech and Signal Processing*, 2021, pp. 8558–8562.
- 19 S. Chaudhary, S. Sadbhawna, V. Jakhethiya, B. N. Subudhi, U. Baid, S. C. Guntuku, Detecting COVID-19 and Community Acquired Pneumonia Using Chest CT Scan Images With Deep Learning, in: *IEEE International Conference on Acoustics, Speech and Signal Processing*, 2021, pp. 8583–8587.
- 20 F. Bougourzi, R. Contino, C. Distant, A. Taleb-ahmed, C. Lille, U. M. R. Iemn, F. Valenciennes, CNR-IEMN: a deep learning based approach to recognize COVID-19 from CT-scan, *IEEE International Conference on Acoustics, Speech and Signal Processing* (2021a) 8568–8572.
- 21 F. Bougourzi, R. Contino, C. Distant, A. Taleb-Ahmed, Recognition of covid-19 from ct scans using two-stage deep-learning-based approach: CNR-IEMN, *Sensors* 21 (2021b) 5878.
- 22 M. Abdel-Basset, H. Hawash, N. Moustafa, O. M. Elkomy, Two-stage deep learning framework for discrimination between COVID-19 and community-acquired pneumonia from chest CT scans, *Pattern Recognition Letters* 152 (2021) 311–319.
- 23 N. Ewen, N. Khan, Online unsupervised learning for domain shift in COVID-19 CT scan datasets, in: *IEEE International Conference on Autonomous Systems*, 2021, pp. 1–5.
- 24 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 Communications* 11 (2020) 1–14.
- 25 R. Su, W. Xie, T. Tan, 2.75d: Boosting learning efficiency and capability by representing 3d features in 2d, 2020. ArXiv:1904.00906.
- 26 T. L. T. da Silveira, P. G. L. Pinto, J. Murrugarra-Llerena, C. R. Jung, 3D scene geometry estimation from 360° imagery: A survey, *ACM Computing Surveys* (2022). (Just Accepted).
- 27 S. Heidarian, P. Afshar, N. Enshaei, F. Naderkhani, M. J. Rafiee, A. Oikonomou, A. Shafiee, F. B. Fard, K. N. Plataniotis, A. Mohammadi, Robust automated framework for COVID-19 disease identification from a multicenter dataset of chest CT scans (2021a). ArXiv 2109.09241.

- S. Heidarian, P. Afshar, A. Mohammadi, M. J. R. MD, A. O. MD, K. N. Plataniotis, F. Naderkhani, CT-CAPS: Feature extraction-based automated framework for COVID-19 disease identification from chest CT scans using capsule networks (2021b).
- J. Hofmanninger, F. Prayer, J. Pan, S. Röhrich, H. Prosch, G. Langs, Automatic lung segmentation in routine imaging is primarily a data diversity problem, not a methodology problem, *European Radiology Experimental* 4 (2020).
- M. Tan, Q. Le, EfficientNet: Rethinking model scaling for convolutional neural networks, in: *International Conference on Machine Learning*, 2019, pp. 6105–6114.
- G. Huang, Z. Liu, L. Van Der Maaten, K. Q. Weinberger, Densely connected convolutional networks, in: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2017, pp. 4700–4708.
- T. Chen, C. Guestrin, Xgboost: A scalable tree boosting system, in: *International Conference on Knowledge Discovery and Data Mining*, Association for Computing Machinery, 2016, p. 785–794.
- K. Sohn, D. Berthelot, N. Carlini, Z. Zhang, H. Zhang, C. A. Raffel, E. D. Cubuk, A. Kurakin, C.-L. Li, FixMatch: Simplifying semi-supervised learning with consistency and confidence, *Advances in Neural Information Processing Systems* 33 (2020) 596–608.
- T. Hastie, S. Rosset, J. Zhu, H. Zou, Multi-class AdaBoost, *Statistics and its Interface* 2 (2009) 349–360.
- K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition, in: *IEEE Conference on Computer Vision and Pattern Recognition*, 2016, pp. 770–778.
- O. Ronneberger, P. Fischer, T. Brox, U-Net: Convolutional networks for biomedical image segmentation, in: *Lecture Notes in Computer Science*, Springer International Publishing, 2015, pp. 234–241.
- R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, D. Batra, Grad-cam: Visual explanations from deep networks via gradient-based localization, in: *IEEE International Conference on Computer Vision*, 2017, pp. 618–626.
- L. Van der Maaten, G. Hinton, Visualizing data using t-sne., *Journal of Machine Learning Research* 9 (2008).
- Q. Ren, B. Zhou, L. Tian, W. Guo, Detection of COVID-19 with CT images using hybrid complex shearlet scattering networks, *IEEE Journal of Biomedical and Health Informatics* 26 (2022) 194–205.
- L. Li, L. Qin, Z. Xu, Y. Yin, X. Wang, B. Kong, J. Bai, Y. Lu, Z. Fang, Q. Song, K. Cao, D. Liu, G. Wang, Q. Xu, X. Fang, S. Zhang, J. Xia, J. Xia, Using artificial intelligence to detect COVID-19 and community-acquired pneumonia based on pulmonary CT: Evaluation of the diagnostic accuracy, *Radiology* 296 (2020) E65–E71.
- X. Wu, C. Chen, M. Zhong, J. Wang, J. Shi, COVID-AL: The diagnosis of COVID-19 with deep active learning, *Medical Image Analysis* 68 (2021) 101913.
- A. M. Sousa, S. B. Martins, A. X. Falcão, F. Reis, E. Bagatin, K. Irion, ALTIS: A fast and automatic lung and trachea CT-image segmentation method, *Medical Physics* 46 (2019) 4970–4982.
- N. Jabeen, R. Qureshi, A. Sattar, M. Baloch, Diagnostic accuracy of maximum intensity projection in diagnosis of malignant pulmonary nodules, *Cureus* (2019).
- J. Deen K, G. R. M. A. Fuzzy-C-means clustering based segmentation and CNN-classification for accurate segmentation of lung nodules, *Asian Pacific Journal of Cancer Prevention* 18 (2017).
- S. Yang, G. Wang, H. Sun, X. Luo, P. Sun, K. Li, Q. Wang, S. Zhang, Learning COVID-19 pneumonia lesion segmentation from imperfect annotations via divergence-aware selective training, *IEEE Journal of Biomedical and Health Informatics* (2022) 1–1.
- Y. Zhang, Q. Liao, L. Yuan, H. Zhu, J. Xing, J. Zhang, Exploiting shared knowledge from non-COVID lesions for annotation-efficient COVID-19 CT lung infection segmentation, *IEEE Journal of Biomedical and Health Informatics* 25 (2021) 4152–4162.
- A. Laghi, Cautions about radiologic diagnosis of COVID-19 infection driven by artificial intelligence, *The Lancet Digital Health* 2 (2020) e225.
- H.-E. Benseddik, H. Hadj-Abdelkader, B. Cherki, S. Bouchafa, Direct method for rotation estimation from spherical images using 3d mesh surfaces with spharm representation, *Journal of Visual Communication and Image Representation* 40 (2016) 708–720.
- Y. Wang, T. Shi, P. Yun, L. Tai, M. Liu, PointSeg: Real-time semantic segmentation based on 3d lidar point cloud (2018). [ArXiv:1807.06288](https://arxiv.org/abs/1807.06288).
- B. Wu, A. Wan, X. Yue, K. Keutzer, SqueezeSeg: Convolutional neural nets with recurrent CRF for real-time road-object segmentation from 3D lidar point cloud, in: *IEEE International Conference on Robotics and Automation*, 2018, pp. 1887–1893.
- S.-T. Yang, F.-E. Wang, C.-H. Peng, P. Wonka, M. Sun, H.-K. Chu, DuLa-Net: A Dual-Projection Network for Estimating Room Layouts from a Single RGB Panorama, in: *IEEE Conference on Computer Vision and Pattern Recognition*, 2019, pp. 3363–3372.
- M. Eder, P. Moulon, L. Guan, PanoPopups: Indoor 3D Reconstruction with a Plane-Aware Network, in: *International Conference on 3D Vision*, 2019, pp. 76–84.
- C. Fernandez-Labrador, J. M. Facil, A. Perez-Yus, C. Demonceaux, J. Civera, J. Guerrero, Corners for layout: End-to-end layout recovery from 360 images, *IEEE Robotics and Automation Letters* (2020) 1–1.
- C. Sun, M. Sun, H. Chen, Hohonet: 360 indoor holistic understanding with latent horizontal features, 2021.
- N. Zioulis, A. Karakottas, D. Zarpalas, P. Daras, OmniDepth: Dense Depth Estimation for Indoors Spherical Panoramas, in: *Proceedings of the European Conference on Computer Vision*, 2018, pp. 453–471.
- M. Stommel, M. Beetz, W. Xu, Inpainting of missing values in the kinect sensor's depth maps based on background estimates, *IEEE Sensors Journal* 14 (2014) 1107–1116.
- B. Zhou, A. Khosla, A. Lapedriza, A. Oliva, A. Torralba, Learning deep features for discriminative localization, in: *IEEE Conference on Computer Vision and Pattern Recognition*, 2016, pp. 2921–2929.
- J. Hu, L. Shen, G. Sun, Squeeze-and-excitation networks, in: *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2018, pp. 7132–7141.
- Y. Lecun, Y. Bengio, G. Hinton, Deep learning, *Nature* 521 (2015) 436–444.
- J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, L. Fei-Fei, ImageNet: A large-scale hierarchical image database, in: *IEEE Conference on Computer Vision and Pattern Recognition*, 2009.
- T. Elsken, J. H. Metzen, F. Hutter, Neural architecture search: A survey, *Journal of Machine Learning Research* 20 (2019) 1–21.
- G. Huang, Y. Li, G. Pleiss, Z. Liu, J. E. Hopcroft, K. Q. Weinberger, Snapshot ensembles: Train 1, get m for free, *arXiv preprint arXiv:1704.00109* (2017).
- A. Mohammadi, F. Naderkhani, K. N. Plataniotis, A. Oikonomou, K. Farahani, J. Kirby, M. J. Rafiee, F. B. Fard, A. Shafiee, F. Atashzar, P. Tyrrell, IEEE ICASSP 2021 Signal Processing Grand Challenge (SPGC) on COVID-19 Diagnosis, 2021. <http://i-sip.encs.concordia.ca/2021SPGC-COVID19/index.html>, Last accessed on 2021-17-08.
- S. Simpson, F. U. Kay, S. Abbasa, S. Bhalla, J. H. Chung, M. Chung, T. S. Henry, J. P. Kanne, S. Kligerman, J. P. Ko, H. Litt, Radiological society of north america expert consensus document on reporting chest CT findings related to COVID-19: Endorsed by the society of thoracic radiology, the american college of radiology, and RSNA, *Radiology: Cardiothoracic Imaging* 2 (2020) e200152.
- J. Zhang, Y. Chu, N. Zhao, Supervised framework for COVID-19 classification and lesion localization from chest CT, *Ethiopian Journal of Health Development* 34 (2020) 235–242.
- S. J. Reddi, S. Kale, S. Kumar, On the convergence of adam and beyond (2019). [ArXiv:1904.09237](https://arxiv.org/abs/1904.09237).
- T. DeVries, G. W. Taylor, Improved regularization of convolutional neural networks with cutout (2017). [ArXiv:1708.04552](https://arxiv.org/abs/1708.04552).
- Y. Cui, M. Jia, T.-Y. Lin, Y. Song, S. Belongie, Class-balanced loss based on effective number of samples, in: *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2019, pp. 9268–9277.
- Y. Ho, S. Wookey, The real-world-weight cross-entropy loss function: Modeling the costs of mislabeling, *IEEE Access* 8 (2020) 4806–4813.
- D. Di, F. Shi, F. Yan, L. Xia, Z. Mo, Z. Ding, F. Shan, B. Song, S. Li, Y. Wei, Y. Shao, M. Han, Y. Gao, H. Sui, Y. Gao, D. Shen, Hypergraph learning for

- 1
2
3
4 identification of COVID-19 with CT imaging, *Medical Image Analysis* 68 (2021) 101910.
5 F. Melo, Area under the ROC curve, in: *Encyclopedia of Systems Biology*, Springer New York, 2013, pp. 38–39. URL: https://doi.org/10.1007/978-1-4419-9863-7_209. doi:10.1007/978-1-4419-9863-7_209.
6 M. Friedman, The use of ranks to avoid the assumption of normality implicit in the analysis of variance, *Journal of the American Statistical Association* 32 (1937) 675–701.
7 Z.-S. Liu, M.-P. Cani, W.-C. Siu, See360: Novel panoramic view interpolation, *IEEE Transactions on Image Processing* 31 (2022) 1857–1869.
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

- We propose a 2.5D omnidirectional CT lung representation for COVID-19 classification.
- We combine Squeeze-and-Excite with Class Activation Maps for explainability.
- We explore the efficiency of 2D processing while requiring patient-level annotations.
- We obtain state-of-the-art results in binary and ternary COVID-19 diagnosis from CTs.

Declaration of interests

☒ 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.

☐ The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: