Caverns Detection and Caverns Report in Tuberculosis: lesion detection based on image using YOLO-V3 and median based multi-label multi-class classification using SRGAN

Tetsuya Asakawa[a], Riku Tsuneda[a], Kazuki Shimizu[b], Takuyuki Komoda[b] and Masaki Aono[a]

[a] *Toyohashi University of Technology, 1-1 Hibarigaoka Tenpaku, Toyohashi, Aichi, Japan*
[b] *Toyohashi Heart center, 21-1 Gobutori Tenpaku, Oyama, Toyohashi, Aichi, Japan*

Abstract

The ImageCLEF 2022 Tuberculosis Task is an example of a challenging research problem in the field of CT image analysis. The purpose of this research is to make lesion detection for tuberculosis and accurate estimates for the three labels. We describe the tuberculosis task and approach for chest CT image analysis, then perform lesion detection and multi-label classification in the CT image analysis using the task dataset. We propose a fine-tuning deep neural network model that uses inputs from multiple CNN features. In addition, this paper presents two approaches for applying mask data to the extracted 2D image data and for extracting a set of 2D projection images along multi-axis based on the 3D chest CT data. Our submissions on the task test dataset reached a mAP IOU value of about 19% at detection, reached a mean AUC value of about 66% and a minimum AUC value of about 32% in classification.

Keywords 1

Tuberculosis, Deep Learning, Image Super Resolution, Detection, Multi-label and Multi-class classification.

1. Introduction

With the spread of various diseases (e.g., tuberculosis (TB), COVID-19, and influenza), medical research has been performed to develop and implement the necessary treatments for viruses. However, there is no method currently available to identify such diseases early. An early diagnosis method is needed to provide the necessary treatment, develop specific medicines, and prevent the deaths of patients.

Accordingly, a significant amount of effort has been invested in medical image analysis research in recent years. In fact, a task dedicated to TB has been adopted as part of the ImageCLEF evaluation campaign for the five last years [1][2][3][4][5]. In ImageCLEF 2022 the main task [6], "ImageCLEFmed Tuberculosis," is treated as a computed tomography (CT) report.

The goal of the first subtask (Caverns Detection) is to detect lung cavern regions in lung CT images associated with lung caverns characteristics. And the second subtask (Caverns Report) is to predict three binary features of caverns suggested by experienced radiologists.

In this paper, we employ a new fine-tuning neural network model that uses features extracted by pre-trained convolutional neural network (CNN) models and Vision Transformer (ViT) as input. We propose a new fully connected two layers. The new contributions of this paper are the proposition of novel feature building techniques, the incorporation of features from the proposed CNN model, and the

use of several forms of pre-processing to predict TB from the images. In Section 2, we describe the conducted task and the ImageCLEF2022 dataset. In Section 3, we introduce the image pre-processing, experimental settings, and features used in this study. In Section 4, we describe the experiments we performed. In Section 5, we provide our conclusions.

2. ImageCLEF 2022 Dataset

The TB task of the ImageCLEF 2022 Challenge included partial 3D patient chest CT images [6]. The task includes two subtasks: Caverns Detection, Caverns Report.

2.1. Caverns Detection

The task dataset contains 559 train and 140 test cases. In addition, participants may also use 60 training cases from the Caverns Report task. Any other public dataset usage is also welcome.

Each case includes the CT image, two versions of automatically extracted lung masks, and information on cavern area location.

2.2. Caverns Report

The task dataset contains 60 train and 16 test cases. In this task, participants must generate automatic lung-wise reports based on CT image data. Each report should include probability scores (ranging from 0 to 1) for each of the three labels and for each of the lungs (resulting in three entries per CT). The resulting list of entries includes thick walls, calcification, and presence of foci. Table 1 lists the labels for the chest CT scan in the training dataset.

Table 1
Presence of labels for the chest CT scan in the training dataset.

Label	In Training set (total numbers)
thick walls	49
calcification	34
foci presence	30

3. Proposed Method

We propose detection and a multi-label analysis system to predict caverns characteristics from CT scan images. The first step is the input data pre-processing in both analyses. After pre-processing input data, we will describe our deep neural network model that enables the detection and multi-label outputs, given CT scan images. In addition, we add an optional step to the first step. We use a CT scan movie not CT scan images. We will detail our proposed system in the following section.

3.1. Input data pre-processing
3.1.1. Input data pre-processing for Caverns Detection

CT scans in the training and test datasets are provided in compressed Nifti format. We decompressed the files and extracted the slices along the z-axis of the 3D image, as shown in fig. 1. For each Nifti image, we obtained several slices, according to the dimensions, ranging from 110 to 250 images for the z-dimension. After extracting the slices along the z-axis, we filtered the slices of each patient using mask1 and mask2 data [7][8]. The mask1 data provide more accurate masks but tend to miss large abnormal regions of the lungs in the most severe TB cases. The mask2 data provide more rough bounds

but behave more stably in terms of including lesion areas. We extracted the filtered CT scan images. We noticed that all slices contain relevant information, including bone, space, fat, and skin, in addition to the lungs that could help classify the samples. Therefore, we added a step to the filter and selected several slices per patient. We call this data the **Applying mask CT data**.
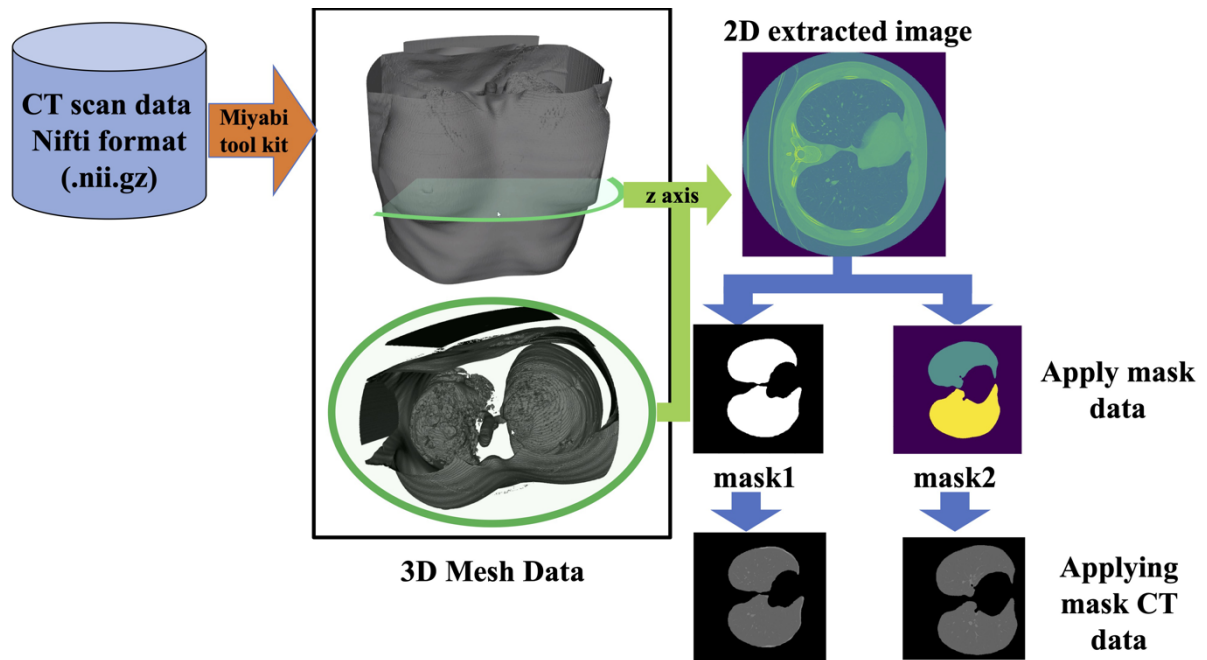

Figure 1: Pre-processing of the input data applying mask data.

3.1.2. Input data pre-processing in Caverns Report

The 3D CT scans in the training and test datasets are provided in compressed Nifti format. We decompressed the files and extracted the slices along the z-axis of the 3D image, as shown in fig. 1. For each Nifti image, we obtained several slices, according to the dimensions, ranging from 110 to 250 images for the z-dimension. After extracting the slices along the z-axis, we filtered the slices of each patient using mask data [8]. The mask data provide more rough bounds but behave more stably in terms of including lesion areas. We extracted the filtered CT scan images. We noticed that all slices contain relevant information, including bone, space, fat, and skin, in addition to the lungs that could help classify the samples. Therefore, we added a step to the filter and selected several slices per patient.

3.2.   Proposed deep neural network model
3.2.1. Proposed deep neural network model for Caverns Detection

To conduct this detection, we propose annotation-based YOLO-V3 [9] that allow inputs coming from medical images. We used Caverns Report Train cavern bounding boxes as input. Cavern area location information includes a cavern area bounding box and cavern area centroid. As illustrated in fig. 2, we annotate from the image, and predict cavern area using YOLO-V3.
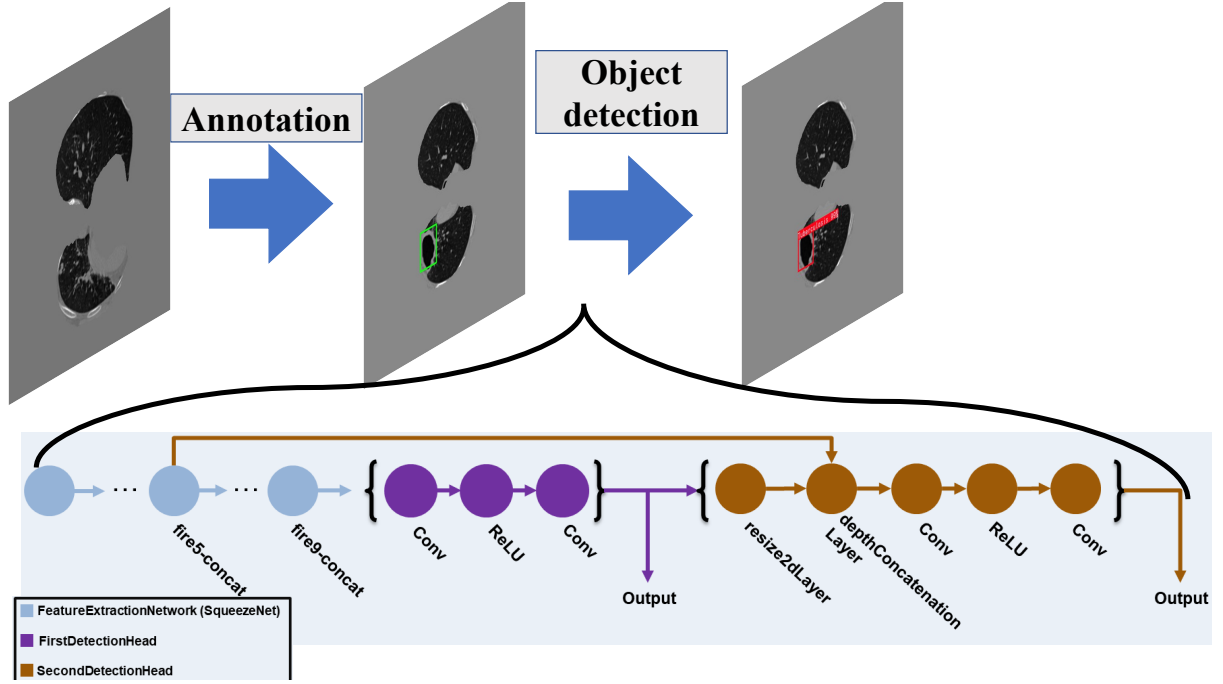
Figure 2: Process for proposed deep neural network model at Caverns Detection.

3.2.2. Proposed deep neural network model at Caverns Report

To solve our multi-label problem, we propose new combined neural network models which allow inputs coming from End-to-end (CNN) features.

We perform Super-Resolution using SRGAN(Generative Adversarial Network for Super-Resolution). The SRGAN allows the model to achieve an upscaling factor of almost 4x for most image visuals. Thus, the resolution pixel size of 512 X 512 of dataset provided by ImageCLEF has pixel size of 2048 X 2048.

Here, we divided the training dataset at random into training and validation datasets with a ratio of 8:2. The CNN features were extracted using pre-trained CNN-based neural networks, including EfficientNet B07. To deal with the above features, we propose a deep neural network architecture.

Our system incorporates CNN features, which can be extracted using deep CNNs pre-trained on ImageNet [10] such as EffcientNet B07[11]. Because of the lack of datasets in visual sentiment analysis, we adopted transfer learning for the feature extraction to prevent overfitting. We decreased the dimensions of the fully connected layers used in the CNN models. In addition, we extracted the vector to 2048 dimensions.

We employ from three CNNs models. As illustrated in Fig. 2, the CNN feature is combined and represented by an integrated feature as a linearly weighted average, where weights are w3 for CNN features, respectively. CNN feature is passed out on "Fusion" processing to generate the integrated features, followed by "softmax" activation function.

Algorithm 1 Predicting multi hot vector for an image

Input: Image data i including K kinds of disease for Lungs
Output: Multi hot vector S_i
  1: **for** k **do** range (K):
  2:    $Prob_{i,k} = FeatureExtraction_{i,k}$
  3:    $T_i^k = \text{median}(Prob_{i,k})$
  4: **end for**

We propose a method illustrated in Algorithm 1. The input is a collection of features extracted from each image with K kinds of sentiments, while the output is a K-dimensional multi-hot vector.

In Algorithm 1, we assume that the extracted CNN feature is represented by their probabilities. For each caverns characteristics, we sum up the features, followed by median of the result, which is denoted by T_k^i in Algorithm 1. In short, the vector S_i represents the output multi-hot vector. We repeat this computation until all the test (unknown) images are processed.
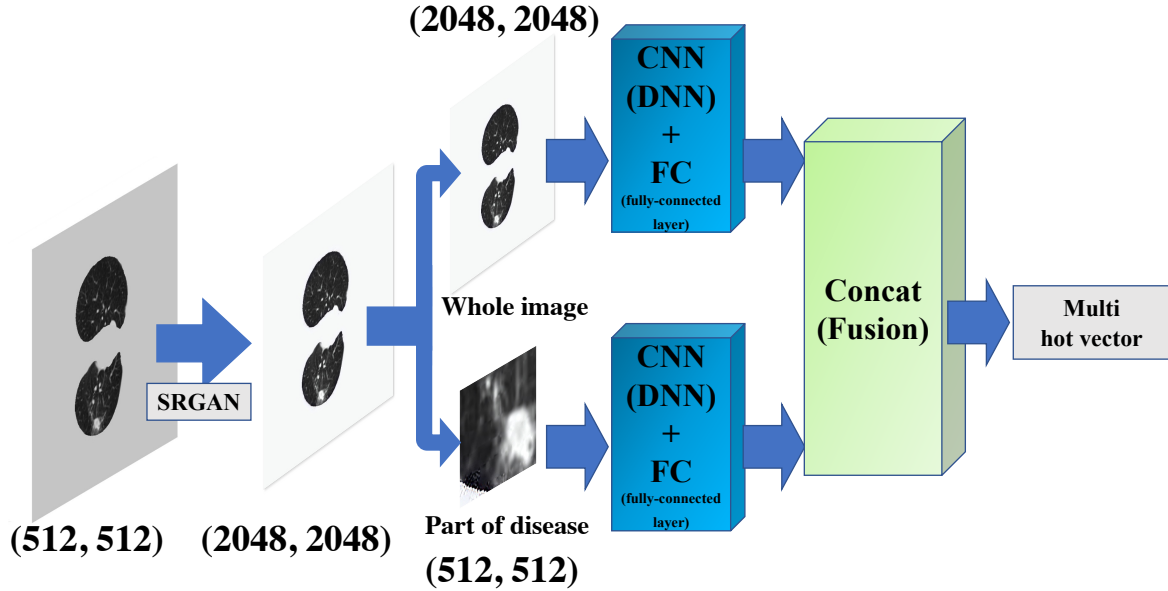

Figure 3: Our proposed method for feature extraction.

4. Experiments
4.1. Submission at several models
4.1.1. Submission at Caverns Detection

The training, validation, and test dataset consists of Applying mask1 and mask2 CT data. Here, we have divided the filtering data into training and validation datasets with a ratio of 8:2 to all slices from the same patient. We determined the following hyper-parameters: the optimizer function is SGD with a learning rate of 0.001 and a momentum of 0.9, decay of 0.0005, a batch size of 4, and the number of epochs is 200. We implement the "map_iou" as a loss. For the implementation, we employed PyTorch as our deep learning framework. These experiments were performed using PyTorch on Ubuntu 20.04. The workstation has an Intel Xeon 6242RXeon(20core/3.10GHz/TDP:205W) CPU with 16GB of 6 RAM and an NVIDIA RTX A6000 GPU.

For the evaluation of the Caverns Detection, Table2 shows the results. Finally, we employed YOLO-V3 for the training and validation datasets and the test data. The results are given in Section (4.2.1).

Table 2
Our submission for the Caverns Detection.

Mask	Model	map_iou
Mask1	YOLO-V3	0.178
Mask2	YOLO-V3	0.185

4.1.2. Submission for Caverns Report

Here, we have divided the filtering data into training and validation datasets with a ratio of 8:2. We determined the following hyper-parameters: the batch size is 256, the optimization function is stochastic gradient descent with a learning rate of 0.001 and a momentum of 0.9, and the number of epochs is 200 using early-stopping. For the implementation, we employed Tensorflow[12] as our deep learning framework. These experiments were performed using Tensorflow 2.6 on Ubuntu 20.04. The workstation has an Intel Xeon 6242RXeon(20core/3.10GHz/TDP:205W) CPU with 16GB of 6 RAM and an NVIDIA RTX A6000 GPU.

For the evaluation of the multi-label classification, we employed mean_auc and min_auc. Table 3 shows the results. Finally, we employed ViT-L/16, DenseNet201, EfficientNet B07 for the training and validation datasets and the test data. The results are given in Section (4.2.2).

Table 3
Our submission for the Caverns Report.

Model	Parameter	mean_auc	min_auc
ViT-L/16	1000	0.598	0.508
DenseNet201	1920	0.559	0.460
EfficientNet B07	2560	0.658	0.317

4.2. The result of Caverns Detection and Report
4.2.1. Results for the training and validation datasets and the test data using our proposed model for Caverns Detection

The results of the other participants' submissions with the map_iou are shown in Table 4. Here, we compare the results in terms of the map_iou. For our team, KDE-lab, our proposed YOLO-V3 has the best score only this time. However, this is not the latest method.

We need to apply these datasets to other methods such as YOLO-X and YOLO-V5 in the future. The results achieved by our submissions are well ranked compared to those at the top of the list given in Table 4. In terms of the map_ioc, our model ranks 3rd.

Table 4
The best participants' runs submitted for the Caverns Detection.

Group name	Rank	map_iou
CSIRO	1	0.504
SenticLab.UAIC	2	0.295
KDE-Lab	**3**	**0.185**
SDVA-UCSD	4	0.000

4.2.2. Results for the training and validation datasets and the test data using our proposed model at Caverns Report

We show the results in Table 5 with the terms of mean_auc and min_auc, and we compare the results in terms of the mean_auc and min_auc to the results of the other participants' submissions.

For our team, KDE-lab, our proposed CNN model has the best mean_auc and min_auc. The results achieved by our submissions are well ranked compared to those at the top of the list given in Table 5. In terms of the mean_auc model ranks 2nd and in terms of the min_auc it ranks 3rd.

Table 5
The best participants' runs submitted for the Caverns Report.

Group name	Rank	mean_auc	min_auc
SDVA-UCSD	1	0.687	0.513
KDE-Lab	**2**	**0.658**	**0.317**
klssncse	3	0.536	0.413
SSN_Dheepak	4	0.461	0.256

5. Conclusions

In this study, we proposed image pre-processing and a CNN model to detect lung cavern regions in lung CT images associated with lung caverns characteristics and to predict three binary features of caverns suggested by experienced radiologists.

We performed a lung CT image analysis in which we proposed a deep neural network model that enabled the inputs to be derived from the CNN features. To predict the three labels, we introduced a median-based multi-label prediction algorithm.

Specifically, after training our deep neural network using the pre-processed images, we were able to predict the categories of the three types of TB cases from unknown CT scan images.

The experimental results demonstrate that our proposed models out-perform some models in terms of the mean_auc and the min_auc. For the mean_auc and min_auc, our model achieved a good value. Therefore, we believe that using pre-processed images is effective.

In the future, given an arbitrary X-ray, CT, echo, or magnetic resonance imaging image might include the optimal weights for the neural networks. Moreover, we hope our proposed model will encourage further research into the early detection of diseases (such as TB, COVID-19, and influenza) or unknown diseases.

6. Acknowledgements

7. References

[1] Y. Dicente Cid, A. Kalinovsky, V. Liauchuk, V. Kovalev, H. Müller, Overview of ImageCLEF-tuberculosis 2017 - predicting tuberculosis type and drug resistances, in: CLEF2017 Working Notes, CEUR Workshop Proceedings, CEUR-WS.org <http://ceur-ws.org>, Dublin, Ireland, 2017.
[2] Y. Dicente Cid, V. Liauchuk, V. Kovalev, , H. Müller, OverviewofImageCLEFtuberculosis2018 - detecting multi-drug resistance, classifying tuberculosis type, and assessing severity score, in: CLEF2018 Working Notes, CEUR Workshop Proceedings, CEUR-WS.org <http://ceur-ws.org>, Avignon, France, 2018.
[3] Y. Dicente Cid, V. Liauchuk, D. Klimuk, A. Tarasau, V. Kovalev, H. Müller, Overview of Im-ageCLEFtuberculosis 2019 - Automatic CT-based Report Generation and Tuberculosis Severity As-sessment, in: CLEF2019 Working Notes, CEUR Workshop Proceedings, CEUR-WS.org <http://ceur-ws.org>, Lugano, Switzerland, 2019.
[4] S. Kozlovski, V. Liauchuk, Y. Dicente Cid, A. Tarasau, V. Kovalev, H. Müller, Overview of Im-ageCLEFtuberculosis 2020 - automatic CT-based report generation, in: CLEF2020 Working Notes, CEUR Workshop Proceedings, CEUR-WS.org <http://ceur-ws.org>, Thessaloniki, Greece, 2020.

[5] S. Kozlovski, V. Liauchuk, Y. Dicente Cid, V. Kovalev, H. Müller, Overview of ImageCLEFtuberculosis 2021 - CT-based tuberculosis type classification, in: CLEF 2021 Working Notes, CEUR Workshop Proceedings, CEUR-WS.org <http://ceur-ws.org>, Bucharest, Romania, 2021.

[6] S. Kozlovski, Y. DicenteCid, V. Kovalev, H. Müller, OverviewofImageCLEFtuberculosis2022 - CT-based caverns detection and report, in: CLEF2022 Working Notes, CEUR Workshop Proceedings, CEUR-WS.org <http://ceur-ws.org>, Bologna, Italy, 2022.

[7] Y. Dicente Cid, O. A. Jiménez del Toro, A. Depeursinge, H. Müller, Efficient and fully automatic segmentation of the lungs in ct volumes, in: O. Goksel, O. A. Jiménez del Toro, A. Foncubierta-Rodríguez, H. Müller (Eds.), Proceedings of the VISCERAL Anatomy Grand Challenge at the 2015 IEEE ISBI, CEUR Workshop Proceedings, CEUR-WS.org <http://ceur- ws.org>, 2015, pp. 31–35.

[8] V. Liauchuk, V. Kovalev, Imageclef 2017: Supervoxels and co-occurrence for tuberculosis ct image classification, in: CLEF2017 Working Notes, CEUR Workshop Proceedings, CEUR-WS.org <http://ceur-ws.org>, Dublin, Ireland, 2017.

[9] R. Joseph, F Ali, YOLOv3: An Incremental Improvement, Computer Vision and Pattern Recognition (CVPR) 2018, arxiv:1804.02767.

[10] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, A. C. Berg, and L. Fei-Fei, ImageNet Large Scale Visual Recognition Challenge, International Journal of Computer Vision (IJCV), 2015, pp. 211–252.

[11] M Tan and Q. V. Le, Efficientnet: Rethinking model scaling for convo-lutional neural networks, ICML 2019, 05 2019.