Semi-supervised Multi-Label Classification with 3D CBAM Resnet for Tuberculosis Cavern Report

Xing Lu[2], An Yan[2], Eric Y Chang[1,2], Chun-Nan Hsu[1,2], Julian McAuley[2], Jiang Du[2] and Amilcare Gentili[1,2]

[1]San Diego VA Health Care System, San Diego, CA, USA
[2]University of California, San Diego, CA, USA

Abstract
Detection and characterization of tuberculosis and the evaluation of lesion characteristics are challenging. To provide a solution for a multi-label classification task of tuberculosis cavern report task, we performed a deep learning study with backbones of 3D Resnet. Semi-supervised learning strategy was implied in this study to leverage the unlabeled dataset from cavern detection task. A convolutional block attention model (CBAM) was used to add an attention mechanism in each block of the Resnet to further improve the performance of the convolutional neural network (CNN). Our solution is ranked the 1st place with submissions obtained Mean_AUC of 0.687 and 0.681 for this task.

Keywords
Tuberculosis Cavern, 3D Convolutional Neural Network, Semi-supervised Learning, Attention Mechanism

1. Introduction

Tuberculosis (TB) is a bacterial infection caused by the germ Mycobacterium tuberculosis, and is a leading cause of death from infectious disease worldwide. An epidemic in many developing regions, such as Africa and Southeast Asia, it was responsible for 1.6 million deaths in 2017 alone. There are different manifestations of TB which require different treatments, making the detection and characterization of TB disease and the evaluation of lesion characteristics critically important tasks in the monitoring, control, and treatment of this disease. An accurate and automated method for classification of TB from CT images may be especially useful in regions of the world with few radiologists.

The ImageCLEF 2022 Tuberculosis task [1, 2] includes two sub-tasks. The first sub-task is lung cavern regions detection,participants must detect lung cavern regions in lung CT images associated with lung tuberculosis; The second sub-task is caverns classification problem. Participants must predict 3 binary features of caverns suggested by experienced radiologists.

CLEF 2022: Conference and Labs of the Evaluation Forum, September 5–8, 2022, Bologna, Italy
lvxingvir@gmail.com (X. Lu); ayan@eng.ucsd.edu (A. Yan); e8chang@health.ucsd.edu (E. Y. Chang); chunnan@health.ucsd.edu (C. Hsu); jmcauley@eng.ucsd.edu (J. McAuley); jiangdu@health.ucsd.edu (J. Du); agentili@ucsd.edu (A. Gentili)
0000-0001-6517-7497 (X. Lu); 0000-0002-0820-1355 (A. Yan); 0000-0003-3633-5630 (E. Y. Chang); 0000-0002-5240-4707 (C. Hsu); 0000-0003-0955-7588 (J. McAuley); 0000-0002-9203-2450 (J. Du); 0000-0002-5623-7512 (A. Gentili)

We tried both tasks, but the detection task is too tight in a limited time and we didn't get very reasonable results, so we focused on report our efforts for the second task, which we used our preciously developed 3D CBAM Resnet model and a semi-supervised training strategy to leverage the uncategorized cavern regions that provided in the detection task.

2. Methods

2.1. Semi-supervised training strategy

For this task, the most challenge comes from the small dataset provided by the organizer for training, which only included 60 in total. But luckily, they also provided a relative larger dataset for the detection task, so we wondered if we could leverage this detection dataset for the report task. As shown in Figure 1, we adopted a semi-supervised training strategy for this task to use both datasets of detection and the report task. Firstly, we randomly split the dataset of report task into train/validation cohort with a ratio of 4:1. Then we use the train cohort to train the model and obtain a best performance model m using the validation cohort. This model m was then used to infer on the unlabeled lesion that obtained from the detection task dataset to generate a pseudo-label for each lesion. Finally, the model m will be trained on the combined dataset to generate a final model $\mathbf{M}$. Then the model $\mathbf{M}$ will be used to do the inference on the Test dataset that provided by the organizer.

2.2. Model and Training

The dataset provided for the report task training set contained a total of 60 patients, with labeling provided for 3 categories: thick walls, calcification, foci, on a patient level. Also the bounding box(bbox) of the carven lesion is provided along with two type of lung masks are provided. To prepare the data to feed into our classification model, the original NIFTI-formatted dataset was transformed into image data using the NiBabel package as the first step. Then, the reformatted images were adjusted to three different window levels, namely baseline, lung, and soft tissue, and then normalized. For baseline window level, the foreground was obtained via the Otsu thresholding algorithm provided in openCV package; for lung and soft tissue, the image levels were set as [-600,1500] and [50,350], respectively. Then, images were normalized to [0,1] with their mean and std value. Finally, the bbox provided were used to crop the lesion area and all three windows and levels of data were saved, and annotation file were rearranged for use in further training.

In this study, a 3D convolutional block attention module (CBAM)-Resnet was employed to train the model for 3-class multi-label classification based on the PyTorch framework. For the Resnet, same as last year [3, 4], a standard 3D-resnet34 [5] was used as the convolutional neural network backbone, with three fc layers to be the classifier. CBAM [6] was used to implement channel and spatial attention mechanisms for each block of the Resnet. Sigmoid was used as the activation function for binary classification.

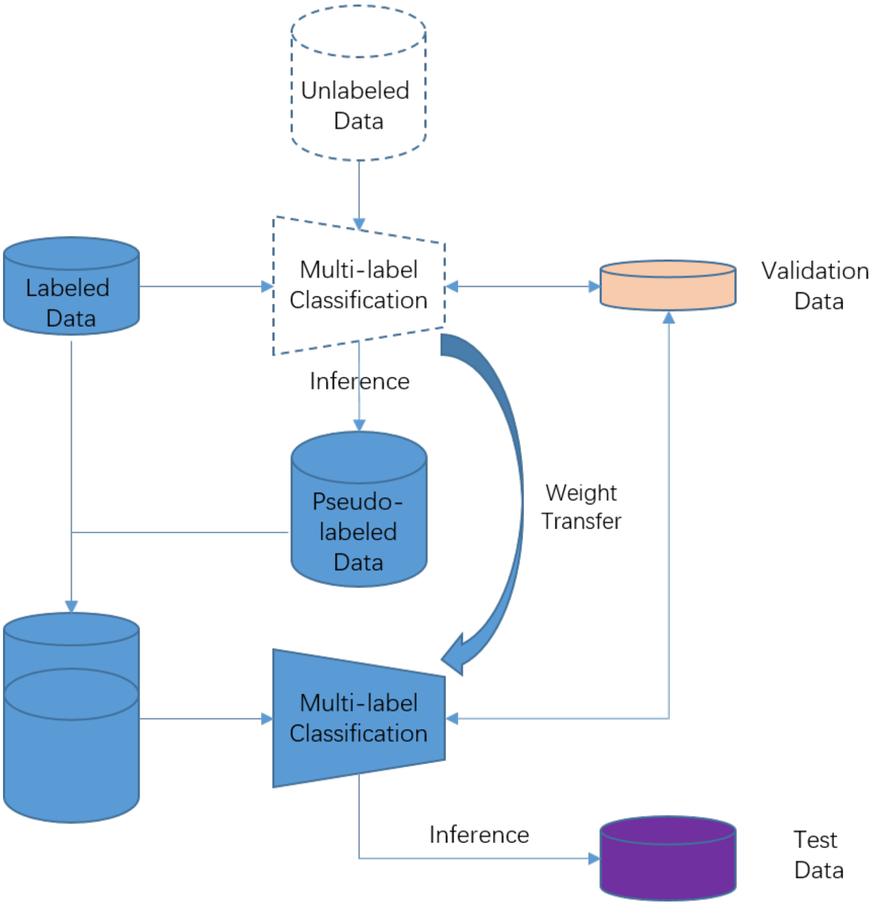

Figure 1: Semi-supervised training strategy for the cavern report task

2.3. Network and Training

In this study, a 3D convolutional block attention module (CBAM)-Resnet and a 3D EfficientNet were employed to train the model for 5-class classification based on the PyTorch framework. Similar to our last year's work [3], a standard 3D-resnet34 [5] was used as the convolutional neural network (CNN) backbone, with three fc layers as the classifier. CBAM [6]was used to implement channel and spatial attention mechanisms for each block of the Resnet, and sigmoid was used as the activation function for binary classification.

To train the neural networks, we used a workstation with 4 Nvidia GTX 1080 Ti video cards, 128 GB RAM, and a 1 TB solid state drive. During the training process, to avoid overfitting, image augmentation and a balanced sampler were implemented in each batch. For the image augmentation, traditional data augmentation methods, including brightness, shear, scale, and flip, were applied. The balanced sampler strategy, which equalized the data sampled from all five classes for each batch, was adopted during the training process.

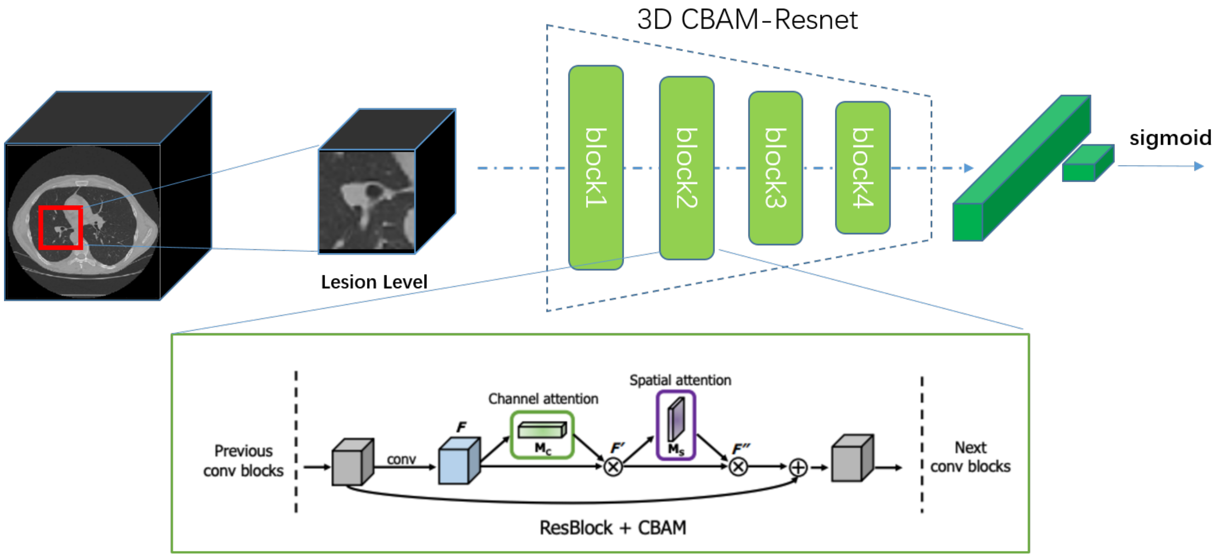

Figure 2: Data prepare and 3D CBAM Resnet for multi-label classification

2.4. Experiments and Model Selection

As the semi-supervised training process was implemented, three model selection experiments were conducted for final submissions. The first is to choose the best mean AUC that evaluated on the validation dataset. The second experiment is choose models saved at every improvement of the AUC for each category during training. The third experiment is choosing models as ensemble of previous models.

3. Results and Submissions

The provided TST dataset including 15 patient level data with lesion bbox provided. With our pre-processing pipeline, the TST data were cropped according to the provided bbox to generate calibrated image files. After evaluation of the trained model, the results were rearranged according to the requirement and saved as the .txt file to be submitted. As mentioned in methods, for three different model choosing strategies, we have 9 saved model for evaluating the TST datasets, the performance could be seen in Table 1.

From the results, the best Mean_AUC was achieved of 0.687 by slightly correction based on the epoch 10 model, while the epoch 10 model obtained a second Mean_AUC of 0.681. The best Min_AUC was obtained by the epoch 35 model of 0.571 and epoch 51 model obtained the second high Min_AUC score of 0.524.

Table 1
Submission model types and results

Submission name	Model Description	Mean AUC	Min AUC
182843	Best on Validation	**0.576**	0.444
182852	Epoch 60	0.612	0.413
182853	Epoch 35	0.593	**0.571**
182854	Epoch 20	0.651	0.476
182893	Epoch 51	0.595	0.524
182894	Epoch 10	0.681	0.492
182896	Visual Adjustment	**0.687**	0.513
182897	Ensemble Nodules	0.660	0.513
182900	Ensemble Calcification	0.581	0.513

4. Discussion and Conclusion

To provide a deep learning solution for a multi-label classification task of tuberculosis carven report with a small training dataset, we did experiments with semi-supervised 3D CBAM Resnet. There are several challenges for this task, such as the extremely small dataset that provided and 3D dimensions of CT images, so we tried several techniques to improve the model performance. First, semi-supervised training strategy was implied to get fully usage of the detection dataset that without the category labeling provided. Second, CBAM was used to add an attention mechanism in each block of the Resnet to further improve the performance of the CNN. Third, different windowing of the CT images were concatenated to make the CNN more focus on the illness features according to radiologist's experience. Using all the aforementioned techniques, we achieved a Mean_AUC of 0.687 and 0.681 in the evaluation of the test dataset, and placed 1st place in this competition task.

5. Acknowledgments

This work was supported in part by the Office of the Assistant Secretary of Defense for Health Affairs through the Accelerating Innovation in Military Medicine Program under Award No. (W81XWH-20-1-0693).

References

[1] B. Ionescu, H. Müller, R. Peteri, J. Rückert, A. Ben Abacha, A. G. S. de Herrera, C. M. Friedrich, L. Bloch, R. Brüngel, A. Idrissi-Yaghir, H. Schäfer, S. Kozlovski, Y. D. Cid, V. Kovalev, L.-D. Ştefan, M. G. Constantin, M. Dogariu, A. Popescu, J. Deshayes-Chossart, H. Schindler, J. Chamberlain, A. Campello, A. Clark, Overview of the ImageCLEF 2022: Multimedia retrieval in medical, social media and nature applications, in: Experimental IR Meets Multilinguality, Multimodality, and Interaction, Proceedings of the 13th International Conference of the CLEF Association (CLEF 2022), LNCS Lecture Notes in Computer Science, Springer, Bologna, Italy, 2022.

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

[3] X. Lu, E. Y. Chang, Z. Liu, C. Hsu, J. Du, A. Gentili, Imageclef2020: Laterality-reduction three-dimensional cbam-resnet with balanced sampler for multi-binary classification of tuberculosis and CT auto reports, in: L. Cappellato, C. Eickhoff, N. Ferro, A. Névéol (Eds.), Working Notes of CLEF 2020 - Conference and Labs of the Evaluation Forum, Thessaloniki, Greece, September 22-25, 2020, volume 2696 of *CEUR Workshop Proceedings*, CEUR-WS.org, 2020. URL: http://ceur-ws.org/Vol-2696/paper_70.pdf.

[4] X. Lu, E. Y. Chang, C. Hsu, J. Du, A. Gentili, Multi-classification study of the tuberculosis with 3d cbam-resnet and efficientnet, in: G. Faggioli, N. Ferro, A. Joly, M. Maistro, F. Piroi (Eds.), Proceedings of the Working Notes of CLEF 2021 - Conference and Labs of the Evaluation Forum, Bucharest, Romania, September 21st - to - 24th, 2021, volume 2936 of *CEUR Workshop Proceedings*, CEUR-WS.org, 2021, pp. 1305–1309. URL: http://ceur-ws.org/Vol-2936/paper-107.pdf.

[5] K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition, 2015. arXiv:1512.03385.

[6] S. Woo, J. Park, J.-Y. Lee, I. S. Kweon, Cbam: Convolutional block attention module, 2018. arXiv:1807.06521.