DIP Covid-19 Course Practicum

Jiajie Li, Kaibin Zhou, Lai Ye, Xiaoyu Jia
Tongji University

Abstract

This is a report on a practical project for a digital image processing course. The task of the project was in classifying the set of computed tomography images of cases. We used a variety of methods to guide data analysis and model training, including exploratory data analysis(EDA), optimal transport dataset distances(OTDD), data distribution exploration etc. Data pre-processing includes data cleaning, morphological manipulation, background removal, etc. We compared and tested multiple network structures, loss functions, and optimizers, and selected the best results for integrated learning. And, we proposed a network structure in the form of gate based on our mining of the data distribution, where high contrast images are trained and predicted using an end-to-end model, while low contrast is used with a two-level classifier. And our model proved to be effective that we ranked first among all participating teams and reached SOTA.

1. Introduction

The task of the course project is to predict the diagnosis type of cases based on their sets of CT images. There are three types of diagnosis: uninfected, community-acquired pneumonia, and covid-19. The training set for the digital image processing course project is divided into two parts. The subject-level part contains a set of CT images of a series of cases, with one label for each case. The slice-level part, contains not only the labels of the cases, but also the labels of each CT image.

2. Related Work

2.1. Image classification based on deep learning

The ImageNet[2] is a largescale ontology of images built upon the backbone of the WordNet structure. The ImageNet Large Scale Visual Recognition Challenge[11] is a benchmark in object category classification and detection based on this dataset. The challenge has been run annually from 2010 to 2017, attracting participation from more than fifty institutions. In the challenge, many new neural network models for image classification are proposed and achieved good results.

Resnet[4] is a residual learning framework presented by He et al. in 2015 to ease the training of networks that are substantially deeper than those used previously. The Squeeze-and-Excitation Networks(SENet) [6] focus on the channel relationship and propose a novel architectural unit, which is termed as the "Squeeze-and-Excitation" (SE) block, that adaptively recalibrates channel-wise feature responses by explicitly modelling interdependencies between channels. The Dense Convolutional Network (DenseNet)[7] is a network presented by Huang el al. in 2018, which connects each layer to every other layer in a feed-forward fashion.

2.2. Deep learning based diagnosis of COVID-19

Since the outbreak of COVID-19, there have been increasing efforts on developing deep learning methods to perform screening of COVID-19 based on medical images such as CT scans and chest X-rays. He et al. developed sample-efficient deep learning methods to accurately diagnose COVID-19 from CT scans[5]. This method is able to judge whether the patient is infected with COVID-19 according to the CT scans and chest X-rays. Li et al. developed a 3D deep learning framework for the detection of COVID-19 using chest CT , referred to COVNet[9]. The model can distinguish whether the patient is infected with Community Acquired Pneumonia, COVID-19 or other non-pneumonia diseases.

2.3. Transfer learning

Transfer learning is normally performed by taking a standard neural architecture along with its pretrained weights on large-scale datasets such as ImageNet[2].

Self-supervised learning (SSL) aims to learn meaningful representations of input data without using human annotations. It creates auxiliary tasks solely using the input data and forces deep networks to learn highly-effective latent features by solving these auxiliary tasks. Momentum Contrast (MoCo) [3] expands the idea of contrastive learning with an additional dictionary and a momentum encoder.

He et al. investigated different strategies of transfer

learning and integrate contrastive self-supervised learning into the transfer learning process to learn powerful and unbiased feature representations for reducing the risk of overfitting[5].

2.4. Optimal Transport Dataset Distances

The notion of task similarity is at the core of various machine learning paradigms, such as domain adaptation and meta-learning. Current methods to quantify it are often heuristic, make strong assumptions on the label sets across the tasks, and many are architecture-dependent, relying on task-specific optimal parameters. David et al. propose an alternative notion of distance between datasets, Optimal Transport Dataset Distances(OTDD)[1]. This distance relies on optimal transport, which provides it with rich geometry awareness, interpretable correspondences and well-understood properties.

2.5. Semi-supervised learning

Semi-supervised learning is an approach to machine learning that combines a small amount of labeled data with a large amount of unlabeled data during training. Semi-supervised learning falls between unsupervised learning (with no labeled training data) and supervised learning (with only labeled training data).

Lee et al. proposded a simple and efficient method of semi-supervised learning for deep neural networks. Basically, the proposed network is trained in a supervised fashion with labeled and unlabeled data simultaneously. For unlabeled data, Pseudo-Labels[8], just picking up the class which has the maximum predicted probability, are used as if they were true labels. In principle, this method can combine almost all neural network models and training methods.

2.6. Attention mechanism

The attention mechanism emerged as an improvement over the encoder decoder-based neural machine translation system in natural language processing(NLP). Later, this mechanism, or its variants, was used in other applications, including computer vision, speech processing, etc. Attention not only tells where to focus, it also improves the representation of interests. Woo et al. proposed Convolutional Block Attention Module (CBAM)[12], a simple yet effective attention module for feed-forward convolutional neural networks, which increase representation power by using attention mechanism: focusing on important features and suppressing unnecessary ones.

3. Network architecture design

3.1. Exploratory data analysis

Exploratory data analysis is an approach to analyzing data sets to summarize their main characteristics.

In our work, we obtains the mean and variance of CT images of normal, covid-19, CAP subject and visualizes the differences between them, as shown in figure 1.

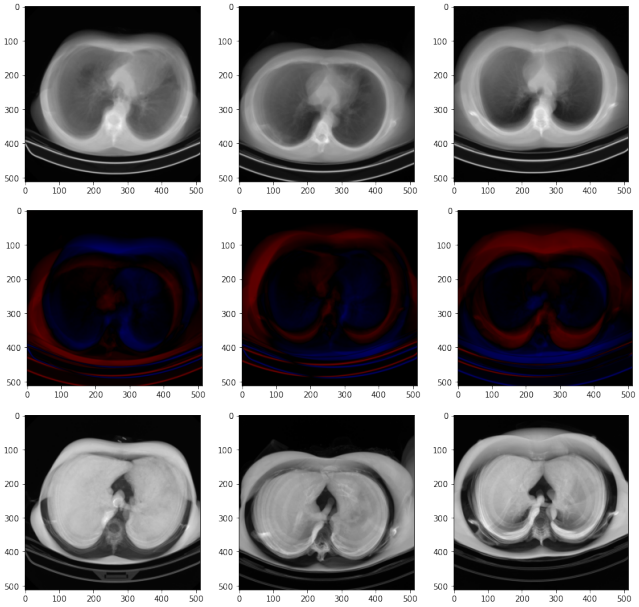


Figure 1. The first line shows the mean of normal, Cap, Covid-19 slices. The third line shows the standard deviation of them. The second line shows the difference of mean of Cap & normal, Covid-19 & normal, Cap & Covid-19 slides.

We find that the image noise was concentrated in the water stains below the pathological section, and the images also have varying degrees of faint noise around the pathological section.

3.2. Data cleaning and preprocessing

3.2.1 Confusing data set and image similarity

We find many errors in the image dataset. The images in the computed tomography (CT) image set of a medical case should be listed in strict order, however, many of the images in the dataset appeared in the wrong order, which forced us to delete them.

To clean the dataset, we calculate the similarity of the images, which is measured by calculating the hash value of the image. When the difference between the hash values of two images exceeds the threshold (around 5), we consider them to be dissimilar. If an image is dissimilar to both its previous and subsequent images, then it is in the wrong order and we remove it.

3.2.2 Removing the background

We use morphological operations and Unet to remove noise from the images.

Image morphological operations are a collection of shape-based image processing operations, mainly based on the mathematics of morphology based on set theory. There are four main operations: expansion, erosion, opening, and closing.

Expansion is the process of overlapping the origin of a structure element with a 1 in the binary image and changing the value of the overlapping part of the binary image that is not a 1 to a 1.

A and B are two sets, and A is defined by B expansion as :

$$A \oplus B = \left\{ z \mid (\hat{B})_z \cap A \neq \phi \right\} \tag{1}$$

Corrosion is the overlay of the origin of the structure element on the 1 of each binary image. As long as there is a 0 overlapping the 1 of the structure element on the binary image, then the value overlapping the origin is 0.

A and B are two sets, and A is corrupted by B defined as :

$$A \Theta B = \{ z \mid (B)_z \subseteq A \} \tag{2}$$

Figure 2. original image,corroded image and expanded image

The open operation refers to erosion followed by expansion, which can remove small objects.

The open operation refers to expansion followed by erosion, which can fill small objects.

We perform intensity thresholding and morphological operations, then estimate the lung lobe field of view, select the largest connected component and save the image to display the segmentation results.

In addition, we try to use Unet for image segmentation.Unet networks have been used extensively for segmentation of medical images since they were proposed. The basic structure of Unet is shown in Figure 3.

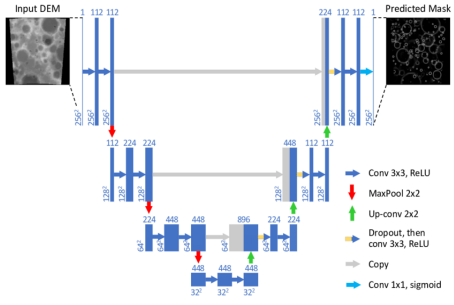


Figure 3. Unet

The example result is shown in Figure 4.

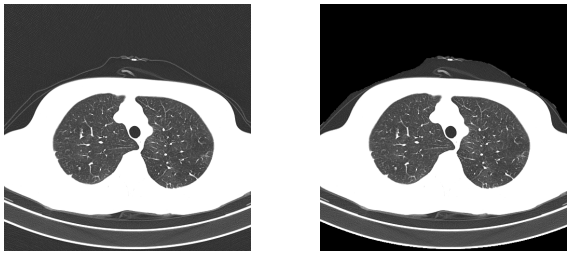


Figure 4. result of image segmentation

3.2.3 Image Enhancement

We try various histogram equalization methods to enhance the image. The result of histogram equalization(HE) is shown in Figure 5 and the result of contrast-limited adaptive histogram equalization(CLAHE) is shown in Figure 6. The comparison of the original image, the result of HE and the result of CLAHE is shown in Figure 7.

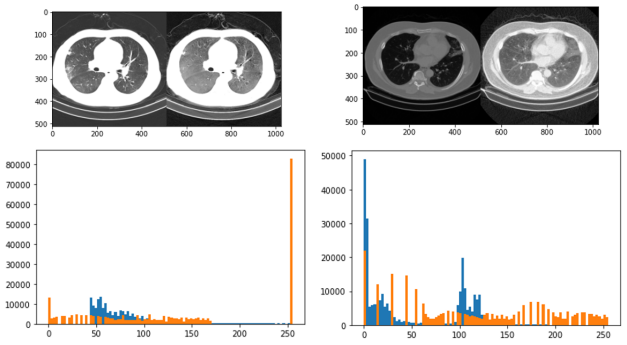


Figure 5. images and histograms with HE

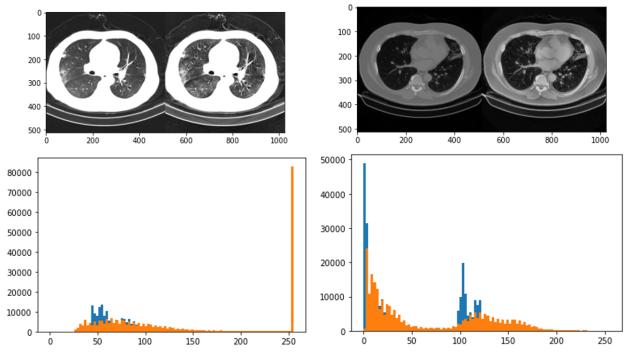


Figure 6. images and histograms with CLAHE

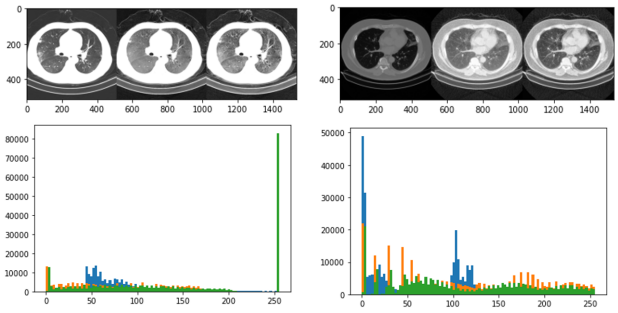


Figure 7. images and histograms with HE and CLAHE

After comparing the results of different processing methods, we used global histogram equalization (HE) followed by contrast-limited adaptive histogram equalization (CLAHE) for image enhancement.

3.2.4 OTDD

We use Optimal Transport Dataset Distance(OTDD) to measure the distances between classification datasets. There are two purposes to do this, the first is to guide us how to select the dataset for pre-training, and the second is to guide us how to do image enhancement.

OTDD is an approach to defining and computing similarities, or distances, between classification datasets. The approach is able to compute distances between two different kinds of probability distributions, those corresponding to the labels and those corresponding to the entire datasets. Using optimal transport to compare two probability distributions requires defining a distance between points sampled from those distributions. Thanks to optimal transport, we have a distance between distributions over feature-label pairs—that is, datasets—which is our Optimal Transport Dataset Distance:

$$\mathrm{OTDD}\left(\mathcal{D}_A, \mathcal{D}_B\right) = \min_{\pi \in \Pi(P_A, P_B)} \int_{\mathcal{Z} \times \mathcal{Z}} d\left(z, z'\right) \mathrm{d}\pi\left(z, z'\right) \tag{3}$$

With the method to compute the OTDD, we do two experiments to compare the distances between two datasets: the low contrast image dataset and the high contrast image dataset, the original image dataset and the image dataset with histogram equalization. The result is shown in Figure 8.

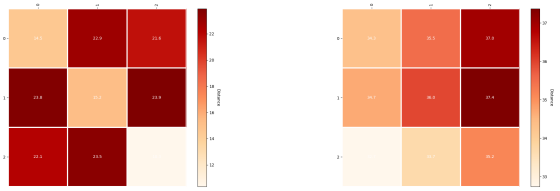


Figure 8. OTDD (first is the distance between original image and image with he, second is the distance between low contrast and high contrast image)

We can see from the figure that between the original image and the image with histogram equalization, the distance between the images of the same category remains the minimum, while the distance between the images of different categories is larger, which indicates that histogram equalization will not make the categories between images become chaotic, and the images of different categories can still be clearly distinguished.

By comparing the distance between high contrast images and low contrast images, we can see that the distance between different categories of images in two datasets is chaotic. One category of images in one dataset is not necessarily the same category in another dataset. Therefore, if a model is used to classify high contrast images and low contrast images at the same time, it will not work well. This leads us to use two different models to classify two kinds of contrast images.

3.3. Network architecture

3.3.1 Decision fusion model

Decision fusion is a form of data fusion that combines the decisions of multiple classifiers into a common decision.

3.3.2 Semi-supervised learning and pseudo-labeling

We apply semi-supervised learning and pseudo-labeling techniques to train the first and second level models.

We train the first-level classifier with slice-level data, input a single slice image, and output the probability vector that the image belongs to the normal, CAP, and covid-19 categories. Then we input all the images of subject-level into the trained first-level classifier and output their probability vectors (pseudo-labeling). The probability vectors of all images (including slice-level and subject-level) are used as inputs for the second-level classifier, and the probability vectors of cases belonging to the normal, CAP, and covid-19 categories are output.

3.3.3 Focal loss

We use focal loss to cope with the problem of unbalanced sample categories. Focal loss is defined as follows.

$$FL(P_t) = -(1 - P_t)^{\gamma} \log(P_t) \tag{4}$$

where Pt denotes the expected probability of the model output corresponding to the correct type. γ is called the focusing parameter, $\gamma \geq 0$.

3.3.4 Attention mechanism

We used the CBAM model and the ECA model. Convolutional Block Attention Module (CBAM) represents the attention mechanism module of the convolutional module. It is an attention mechanism module that combines space and channel. It can achieve better results than senet's attention mechanism which only focuses on channels. The structure of CBAM model is given as Figure 9.

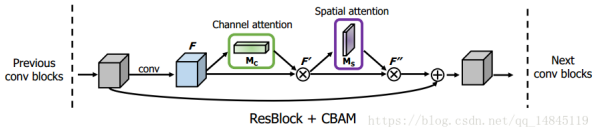


Figure 9. CBAM

ECA-Net replaces the two descending and then ascending convolutions in SENet with a more efficient connection to improve the accuracy and reduce the number of parameters at the same time[10]. The structure of ECA-Net is given as Figure 10.

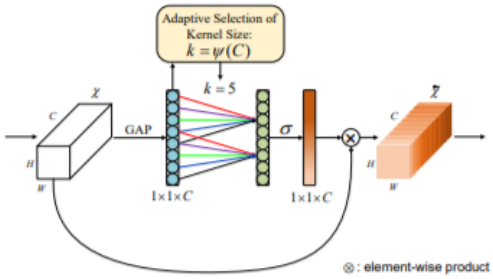


Figure 10. ECA-Net

3.3.5 Data distribution exploration

We collected and organized the creation time, access time, and editing time of the dataset images, and plotted the corresponding histograms. The creation time and access time are shown in Figure 11. The editing time is shown in Figure 12.

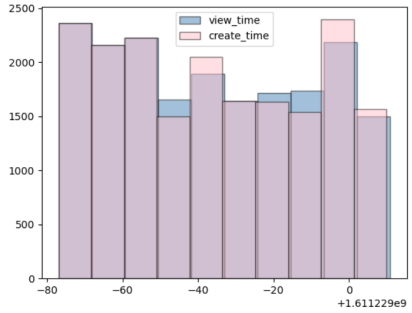


Figure 11. Histogram of creation time, view time of the dataset images

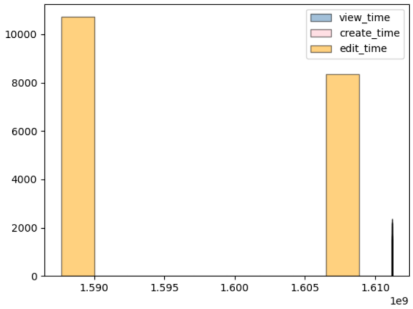


Figure 12. Histogram of editing time for the dataset images

From the figure, it can be seen that the creation time and access time of the dataset images are very similar, but the editing time of the dataset images is clearly divided into two categories, which are edited about eight months apart.

The dataset images are clearly divided into high-contrast images and low-contrast images, which is shown in Figure 13.

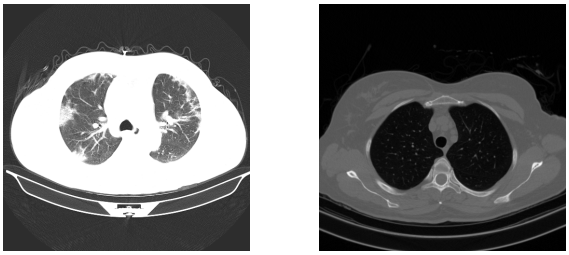


Figure 13. high-contrast image and low-contrast image

3.3.6 Gate

In the end-to-end model, the difference between the effect of high-contrast images and low-contrast images is extremely large, with the low-contrast effect being extremely poor.

Therefore, we use Gate to divide the data into two categories according to high contrast and low contrast, and input them to the two models for training. The two models

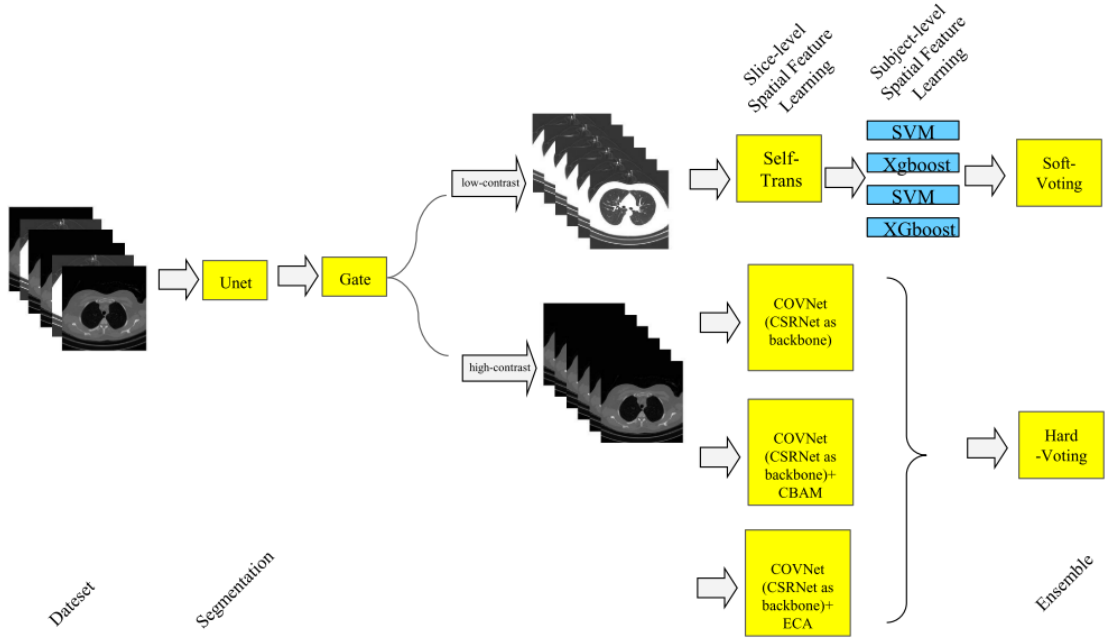


Figure 14. Model Architecture

are trained separately, and the prediction is divided into two models according to the contrast. The network structure for training is shown in Figure 14. The classification report of low-contrast image is shown in Table 1 and the classification report of high-contrast image is shown in Table 2.

Table 1. low-contrast image model evaluation

	precision	recall	f1-score	support
Normal	0.92	1.00	0.96	12
CAP	1.00	0.92	0.96	13
COVID-19	1.00	1.00	1.00	20
accuracy			0.98	45
macro avg	0.97	0.97	0.97	45
weighted avg	0.98	0.98	0.98	45

Table 2. high-contrast image model evaluation

	precision	recall	f1-score	support
Normal	0.94	1.00	0.97	16
CAP	0.94	0.94	0.94	16
COVID-19	0.90	0.82	0.86	11
accuracy			0.93	43
macro avg	0.93	0.92	0.92	43
weighted avg	0.93	0.93	0.93	43

3.4. Explainability of neural networks

3.4.1 Intermediate layer eigenvalues

We try to perform the visualization of the middle layer eigenvalues of the model with three different weights, which is shown in Figure 15.

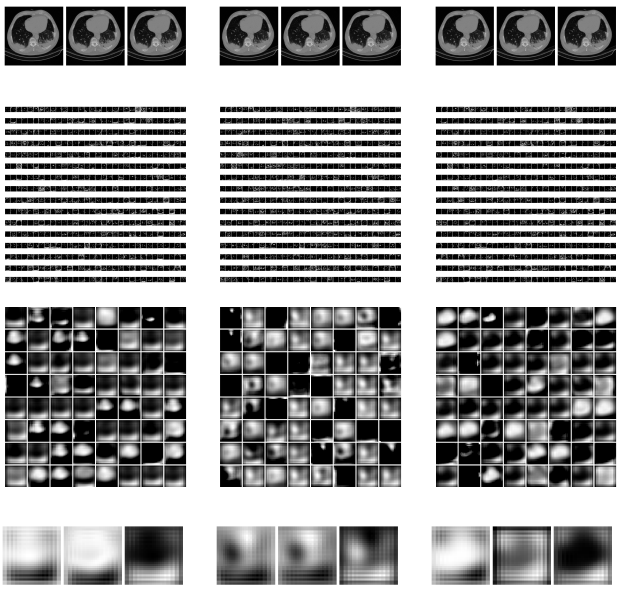


Figure 15. visualization of four intermediate layer feature values

4. Conclusion

In this project, we propose a gate-form network architecture to predict three types of cases, uninfected, community-acquired pneumonia(CAP) and covid-19, based on CT image datasets. We compared a variety of network architectures and use gate to differentiate between high and low contrast data and process them separately. We compared multiple attention mechanisms and forms of loss, and selected the best results for integrated learning, and we won first place among all participating teams.

References

[1] D. Alvarez-Melis and N. Fusi. Geometric dataset distances via optimal transport. *arXiv preprint arXiv:2002.02923*, 2020.

[2] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei. Imagenet: A large-scale hierarchical image database. In *2009 IEEE conference on computer vision and pattern recognition*, pages 248–255. Ieee, 2009.

[3] K. He, H. Fan, Y. Wu, S. Xie, and R. Girshick. Momentum contrast for unsupervised visual representation learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 9729–9738, 2020.

[4] K. He, X. Zhang, S. Ren, and J. Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016.

[5] X. He, X. Yang, S. Zhang, J. Zhao, Y. Zhang, E. Xing, and P. Xie. Sample-efficient deep learning for covid-19 diagnosis based on ct scans. *MedRxiv*, 2020.

[6] J. Hu, L. Shen, and G. Sun. Squeeze-and-excitation networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 7132–7141, 2018.

[7] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger. Densely connected convolutional networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 4700–4708, 2017.

[8] D.-H. Lee et al. Pseudo-label: The simple and efficient semi-supervised learning method for deep neural networks. In *Workshop on challenges in representation learning, ICML*, volume 3, 2013.

[9] L. Li, L. Qin, Z. Xu, Y. Yin, X. Wang, B. Kong, J. Bai, Y. Lu, Z. Fang, Q. Song, et al. Artificial intelligence distinguishes covid-19 from community acquired pneumonia on chest ct. *Radiology*, 2020.

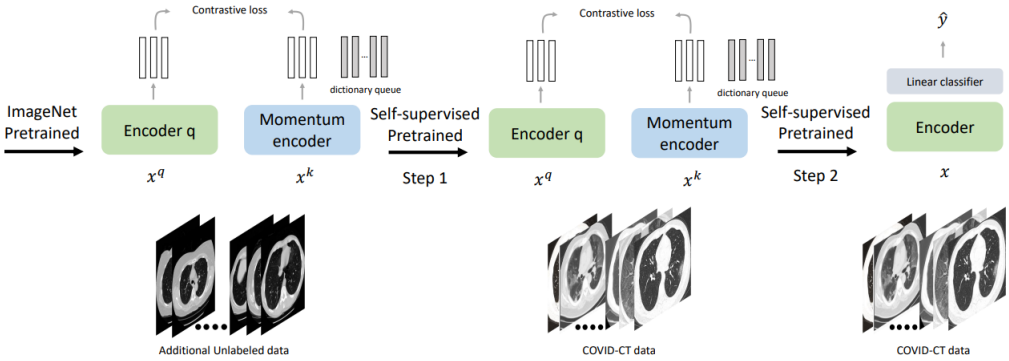
[10] P. Z. P. L. W. Z. Qilong Wang, Banggu Wu and Q. Hu. Eca-net: Efficient channel attention for deep convolutional neural networks. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2020.

[11] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, et al. Imagenet large scale visual recognition challenge. *International journal of computer vision*, 115(3):211–252, 2015.

[12] S. Woo, J. Park, J.-Y. Lee, and I. S. Kweon. Cbam: Convolutional block attention module. In *Proceedings of the European conference on computer vision (ECCV)*, pages 3–19, 2018.

A Appendix

A.1 Architecture of Self-Trans



A.2 Architecture of COVNet(CSRNet as backbone)