

COVID-19 DIAGNOSIS BASED ON DEEP LEARNING ANALYSIS OF CT IMAGES

Xuehai He[†], Fan Wang[†], and Xiaoli Yu[†]

[†]University of California San Diego, La Jolla, CA 92093-0238,

ABSTRACT

The COVID-19 disease caused by the SARS-CoV-2 virus has swept the world with unprecedented pace and severity. With the continued growth of the COVID-19 pandemic worldwide, there is growing interest in using medical imaging for the diagnosis of COVID-19 cases. In this paper, we evaluate the performance of different network backbones for diagnosing patients with COVID-19 CT images. We use transfer learning and mixup technique to improve the performance.

Index Terms— COVID, CT, transfer learning

1. INTRODUCTION

Coronavirus disease 2019 or COVID-19 is an infectious disease that has caused over 2 million infection in the USA, as of June 11rd, 2020. One major hurdle in controlling the spreading of this disease is the shortage of tests. Currently, the tests are mostly based on reverse transcription polymerase chain reaction (RT-PCR), which are always in great shortage. To address this, hospitals have been utilizing alternative diagnosis methods. Because the majority of COVID-19 cases have similar CT images features ground-glass opacities in the early stage and pulmonary consolidation in the late stage, it enables radiologists and trained AI tools to distinguish COVID-19 from other viral pneumonias and do the correct prediction. And many hospitals began to use computed tomography (CT) scans for COVID-19 diagnosis and screening, which has been demonstrated to be effective.

There are already numerous paper published on utilizing deep learning methods in COVID-19 diagnoses. For example, [1] used a 3D deep CNN on chest scans to detect COVID-19. [2] developed a deep learning based CT diagnosis system to assist clinicians to identify patients with COVID-19. [3] developed a deep learning algorithm by modifying the inception transfer-learning model to provide clinical diagnosis ahead of the pathogenic test. In this paper we follow this step and use deep learning based method to do diagnosis based on CT images. Because the availability of datasets of CT images of COVID-19 patients is still not ample due to its novelty, though quickly increasing, our project used the COVID-CT [4] dataset as the training, validating, and testing dataset,

which is made by collecting images from COVID19-related papers.

2. DATASETS

We use the COVID-CT [4] dataset, which contains 349 COVID-19 CT images from 216 patients and 463 non-COVID-19 CTs. The number of training, validation, and testing images is 425, 118, and 203. The CT images have different sizes. These images are from 216 patient cases. For patients labeled with positive, 169 of them have age information and 137 of them have gender information. Figure 1 shows the age distribution of patients with COVID-19. Figure 2 shows the gender ratio of patients with COVID-19. Male patients are more than female patients, with a number of 86 and 51 respectively.

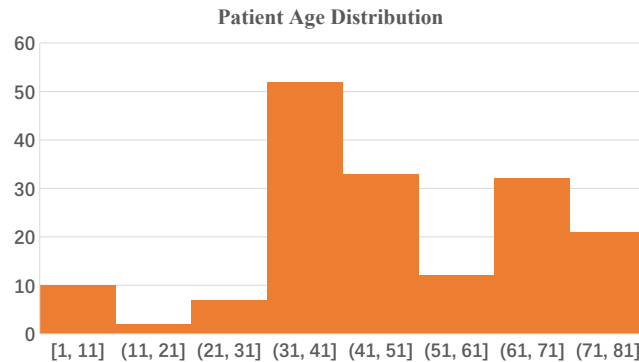


Fig. 1. Age distribution of COVID-19 patients.

3. METHODS

In this section, we will introduce our approaches to model and make predictions for COVID-2019 based on the CT image. In particular, because the size of the dataset is small, one major challenge we are confronted with is addressing issues that can prevent our model from generalizing. Therefore, we do an augmentation on the training data to avoid overfitting to the training data, and add the mixup to it to avert the memorization of noise and artifacts. We also train our models from scratch—starting from a shallow CNN network to

[†] ECE 228 Course Project Report, Group 32.

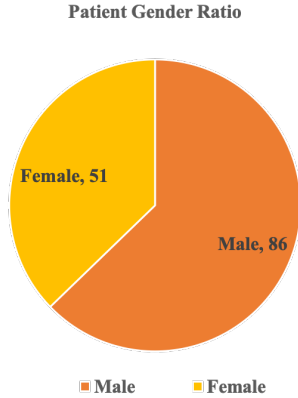


Fig. 2. The gender ratio of COVID-19 patients. The ratio of male:female is 86:51.

observe whether we have already gone into the field of overfitting. And then several different network are considered: one is the relative traditional VGGNet pre-trained on the ImageNet [5] with dropout added so that it is less prone to overfitting. Another one is the DenseNet pre-trained on other CT datasets, with batch normalization layers in each block of its original skeleton and skip-connections introduced in its structure that contribute to reducing overfitting, to promote the generalization of the model. Similarly, We have also tested models such as ResNet [6], EfficientNet [7]. And we do the transfer learning, the result turns out that DenseNet-169 tends to have the best performance, so we add the mixup [8] technique to DenseNet-169 and compare the performance of it under different parameters settings in mixup.

3.1. Data Augmentation

The data augmentation is implemented to improve generalization. We first apply different random affine transformation on our dataset, to avoid the potential overfitting problem since the number of training CT volumes was really limited. We implement random affine transformation, random crop, and flip. The random affine transformation consists of translation and $(15^\circ \pm 10^\circ)$ rotation.

3.2. Mixup

Considering the limited labels of the dataset, we adopt the mixup [8] method to improve the generalization of models and reduce the overfitting.

The mixup extends the training distribution by constructing virtual training examples from the existing examples

$$\begin{aligned}\tilde{x} &= \lambda x_i + (1 - \lambda)x_j, \\ \tilde{y} &= \lambda y_i + (1 - \lambda)y_j\end{aligned}$$

where x_i, x_j are raw input CT reshaped image vectors and y_i, y_j are binary labels, and $\lambda \sim \text{Beta}(\alpha, \alpha)$.

3.3. Shallow CNN network

We first construct a shallow convolutional neural network. The architecture of the simple CNN consists of a combination of a 2d-convolution with an ReLU activation as the basic block and the block repeats four times. A global average pooling layer is used after the blocks for the dimensional reduction of data to prevent overfitting, and a fully connected layer concatenated at the last.

3.4. VGGNet network

We also implement the VGGNet [9] trained from scratch and pre-trained on the ImageNet [5], which is a more powerful neural network architecture than simple CNN to do the prediction. We replace the last fully-connected layer of it so that it can output the confidence score for binary labels.

3.5. DenseNet network

DenseNets [10] have been shown to be the best architecture for X-rays predictive models in [11, 12]. So we test it for CT-based predictive models. There are many medical-used DenseNet pretrained models: for example one pre-trained on the ChestX-ray 14 dataset [11], and the one one given in [12] is pre-trained using NIH aka Chest X-ray14 [13], PC aka PadChest [14], CheX aka CheXpert [15], MIMIC-CXR [16], OpenI [17], Google [18], and Kaggle aka the RSNA Pneumonia Detection Challenge¹. To be fair to comparisons, we still use the version pretrained on the ImageNet dataset. And we also test the DenseNet trained from scratch.

3.6. ResNet network

ResNet [6] is a residual learning framework to ease the training of networks that are substantially deeper than those used previously. Actually DenseNet is improved based on the ResNet.

3.7. EfficientNet network

EfficientNet is the newest one among the models we evaluate and it is quite novel. The parameters in EfficientNet [7] are obtained by doing neural architecture search to design a new baseline network and scale it up to obtain this family of models.

3.8. Evaluation Metrics

We evaluate our approaches using three metrics: (1) accuracy, which measures the percentage of diagnostic prediction

¹<https://www.kaggle.com/c/rsna-pneumonia-detection-challenge>

that matches exactly with the ground-truth; (2) F1-score [19], which is the average of precision and recall. It is used to evaluate the accuracy of our binary models by measuring the average overlap between the predicted diagnosis and the ground-truth; (3) AUC [20], which is also an indicator to measure the performance of the classifier that can be used to determine the performance per task as it accounts for imbalance in labels.

We also evaluate the precision and the recall separately apart from the F1-score. Precision is the number of true positives divided by the total number of elements labeled as belonging to the positive class, and recall is the number of true positives divided by the total number of elements that actually belong to the positive class, where the number of “true positive” is defined as the number of correctly predicted COVID-19 pneumonia cases/images and “positive” class is defined as the COVID-19 pneumonia patients.

4. TRANSFER LEARNING

Given the target task where the amount of training data is limited, training deep neural networks on this small-sized dataset has high risk of overfitting. To alleviate overfitting, we consider the use of transfer learning [21], which is a commonly used technique in deep learning to learn more effective representations. The “transfer” is a method which usually done by pretraining the deep neural network on the large dataset, that aims to make use of the generalized, learned features in large-scale, labelled datasets, and apply those learned features to some different but related dataset to compensate the lack of data in the related dataset. It has long been used to improve performance on computer vision tasks, as well as tasks in the medical domain, such as medical image classification and recognition tasks [22]. One can pretrain the feature extraction layers in the network on large-sized external data from some source tasks, then finetune these layers on the target data. The abundance of source data enables the network to learn expressive and generalizable representations that are robust to overfitting, which can be leveraged to assist in the learning of the target task with more resilience to overfitting.

In our project, we plan to implement the transfer learning methodology in the way that we pretrain a common neural architecture, for instance ResNet, on the available large-scale dataset, for instance ImageNet. Then, modify it to make it applicable to the COVID-19 CT image dataset to alleviate the situation of lacking enough images. We aim at mending these gaps through our project and come up with a model that could be satisfactory.

Of course, there are things need to be addressed when trying to apply the transfer learning technique. On one hand, the neural network architecture could impact the viability of transfer learning, as certain architectures may respond better in transfer learning than other architectures. On the other hand, the training dataset of our selection, ImageNet, though being frequently cited, has a significant gap with the COVID-

19 CT images that we aim at. The ImageNet dataset features objects that are often seen in our daily lives, while the CT images, well, are simply not. We want to evaluate the effectiveness of transfer learning for CT-based COVID-19 diagnosis relies for the following three factors by experiments:

- **Task:** The complexity and difficulty of the target task determines the requirement of network capacity and the amount of source data. The task of medical imaging diagnosis is usually different from natural image classification. For example, the standard task for ImageNet classification has 1000 classes, while the CT-based COVID-19 diagnosis is essentially a binary classification problem. Such difference brings challenge to the direct transfer learning from ImageNet classification to our task. And we evaluate this by comparing it with the models trained from scratch.
- **Data:** The training data for medical imaging tasks is usually difficult to obtain, especially for the CT-based COVID-19 screening task. Existing COVID-19 CT data is scarce and usually expensive to access, proposing higher requirements for the transfer learning strategy to yield better diagnosis performance. Generally, as the number of labeled data involved in the training process increases, the performance of the model will be improved. That is why we introduce the ImageNet dataset into our pretraining process.
- **Model:** The performance of the CT-based diagnosis also depends on the neural network architectures. We study the effects of different neural network architectures by evaluating the transfer learning with different backbone networks, including VGG-16 [9], ResNet-18 [6], ResNet-50 [6], DenseNet-121 [10], DenseNet-169 [10], EfficientNet-b0 [7], EfficientNet-b1 [7], and even shallow CNN network to see the performance.

5. EXPERIMENTATION

In this section, we will perform experiments with different settings that includes the randomly initialized networks and the transferred networks, to demonstrate that the transfer learning methodology is superior and could actually solve the problem of the data shortage. The results and the corresponding analysis will be presented in the following section.

5.1. Settings

We implement the network using PyTorch [23] and train it on GTX 1080Ti GPUs from UCSD datahub. Input images are resized to 224-by-224. We perform data augmentation on the training set. Each training image is augmented with random cropping with a scale of 0.5, horizontal flip, random contrast, and random brightness. We optimize them using Adam [24]

with parameter settings for learning rate of 0.0001 and a mini-batch size of 4. All networks are trained in an end-to-end manner, only CT volumes are fed in as input and the final output alone was supervised without any manual intervention. The commonly used binary cross-entropy loss function on classification task is used to calculate the loss between model predictions and ground-truth labels. To demonstrate the efficacy of the transfer learning methodology, we do two sets of experiments, pretraining on ImageNet and training from scratch respectively, as shown in Table 1.

6. RESULTS

Table 1. Performance comparison between randomly initialized networks (Rand.), ImageNet pretrained networks (Trans.)

	Accuracy		F1		AUC	
	Rand.	Trans.	Rand.	Trans.	Rand.	Trans.
Shallow CNN	0.57	0.63	0.52	0.61	0.64	0.72
VGG-16	0.66	0.75	0.58	0.74	0.74	0.81
ResNet-18	0.66	0.74	0.66	0.73	0.76	0.82
ResNet-50	0.68	0.80	0.72	0.81	0.76	0.88
DenseNet-121	0.77	0.76	0.78	0.78	0.81	0.85
DenseNet-169	0.80	0.83	0.79	0.81	0.86	0.86
EfficientNet-b0	0.71	0.77	0.71	0.78	0.76	0.87
EfficientNet-b1	0.69	0.79	0.62	0.79	0.76	0.81

Table 2. Performance of models on the test set with mixup

	Accuracy	Precision	Recall	F1	AUC
DenseNet ($\alpha=0.1$)	90.5%	0.94	0.95	0.94	0.81
DenseNet ($\alpha=0.2$)	89.0%	0.99	0.88	0.93	0.92
DenseNet ($\alpha=0.4$)	91.5%	0.95	0.97	0.96	0.88
DenseNet ($\alpha=0.6$)	88.0%	0.91	0.94	0.93	0.79
DenseNet ($\alpha=1$)	95.5%	0.98	0.97	0.97	0.93
DenseNet ($\alpha=8$)	96.5%	0.96	0.97	0.94	0.88
DenseNet ($\alpha=32$)	85.5%	0.96	0.86	0.91	0.83
DenseNet($\alpha=0.6$)	84.0%	0.88	0.93	0.90	0.70

As can be seen from the result of each evaluation metric, the deep learning-based COVID-19 diagnostic algorithm used in our study is effective. In our study, only 98 CT-images are used for training, but the best F1 result we have obtained is 0.81 from DenseNet-169, which suggests that the proposed deep learning algorithm is very powerful and the task of AI assisted COVID-19 detection is achievable. We use Grad-CAM [25] to generate heatmaps on input images and the heatmaps indicate which regions are more important to the models for judging whether the images are COVID19-positive or negative. Figure 3 show the visualizations for ImageNet pretrained DenseNet-169 network on one CT-Image examples. The visualizations are based on the NORM5 layer in DenseNet-169.

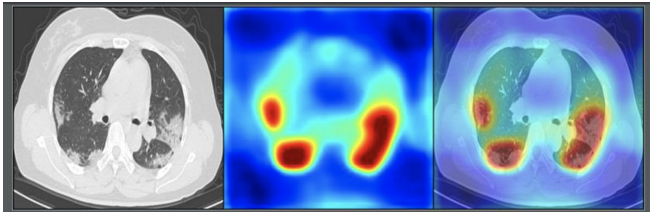


Fig. 3. Grad-CAM visualizations for ImageNet pretrained DenseNet-169 network. Warmer color denotes more attention paid by the model.

We also use t-SNE [26] embeddings of the dataset in Figure 6 on the ImageNet pretrained DenseNet-169 network used for extracting visual features of CTs. We can see the positive and negative examples are somewhat separable, which demonstrates that the model can learn representations of images.

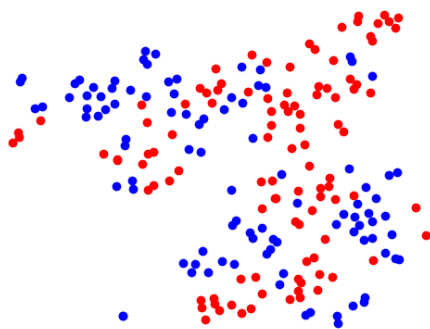


Fig. 4. t-SNE embeddings of positive and negative CTs. Blue and red points represent COVID19-positive and negative examples respectively. Please scale to view better

And we furtherly do mixup on DenseNet-169, the performance is shown in Table 2 with different mixup hyperparameter α . As can be seen, when $\alpha = 8$, the performance tends to be the best. In general, our models are able to generate accurate predictions. However, due to the limitation of the size of data, the performance on the test set is not always stable. With more CT images included in the training, this problem will be addressed, the overfitting will be reduced, and there will be higher performance on the test set.

7. CONCLUSION

In our report, we presented a dedicated deep learning model to facilitate the diagnosis of COVID-19 with given chest CT scan of a patient. Our method that mainly utilized the methodology of transfer learning aimed to alleviate the deficiency of available data as this disease is so novel that any large-scale database is still in the process of being established. We hope that this virus could be eliminated in the near future.

8. REFERENCES

- [1] Chuansheng Zheng, Xianbo Deng, Qing Fu, Qiang Zhou, Jiawei Feng, Hui Ma, Wenyu Liu, and Xinggang Wang. Deep learning-based detection for covid-19 from chest ct using weak label. *medRxiv*, 2020.
- [2] Ying Song, Shuangjia Zheng, Liang Li, Xiang Zhang, Xiaodong Zhang, Ziwang Huang, Jianwen Chen, Huiying Zhao, Yusheng Jie, Ruixuan Wang, et al. Deep learning enables accurate diagnosis of novel coronavirus (covid-19) with ct images. *medRxiv*, 2020.
- [3] Shuai Wang, Bo Kang, Jinlu Ma, Xianjun Zeng, Mingming Xiao, Jia Guo, Mengjiao Cai, Jingyi Yang, Yaodong Li, Xiangfei Meng, et al. A deep learning algorithm using ct images to screen for corona virus disease (covid-19). *medRxiv*, 2020.
- [4] Jinyu Zhao, Yichen Zhang, Xuehai He, and Pengtao Xie. Covid-ct-dataset: a ct scan dataset about covid-19. *arXiv preprint arXiv:2003.13865*, 2020.
- [5] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. Imagenet: A large-scale hierarchical image database. In *CVPR*, 2009.
- [6] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *CVPR*, 2016.
- [7] Mingxing Tan and Quoc V Le. Efficientnet: Rethinking model scaling for convolutional neural networks. *arXiv preprint arXiv:1905.11946*, 2019.
- [8] Hongyi Zhang, Moustapha Cisse, Yann N Dauphin, and David Lopez-Paz. mixup: Beyond empirical risk minimization. *arXiv preprint arXiv:1710.09412*, 2017.
- [9] Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. 2014.
- [10] Gao Huang, Zhuang Liu, Laurens Van Der Maaten, and Kilian Q Weinberger. Densely connected convolutional networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 4700–4708, 2017.
- [11] Pranav Rajpurkar, Jeremy Irvin, Kaylie Zhu, Brandon Yang, Hershel Mehta, Tony Duan, Daisy Ding, Aarti Bagul, Curtis Langlotz, Katie Shpanskaya, et al. Chexnet: Radiologist-level pneumonia detection on chest x-rays with deep learning. *arXiv preprint arXiv:1711.05225*, 2017.
- [12] Joseph Paul Cohen, Mohammad Hashir, Rupert Brooks, and Hadrien Bertrand. On the limits of cross-domain generalization in automated x-ray prediction. *arXiv preprint arXiv:2002.02497*, 2020.
- [13] Xiaosong Wang, Yifan Peng, Le Lu, Zhiyong Lu, Mohammadhadi Bagheri, and Ronald M Summers. Chestx-ray8: Hospital-scale chest x-ray database and benchmarks on weakly-supervised classification and localization of common thorax diseases. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 2097–2106, 2017.
- [14] Aurelia Bustos, Antonio Pertusa, Jose-Maria Salinas, and Maria de la Iglesia-Vayá. Padchest: A large chest x-ray image dataset with multi-label annotated reports. *arXiv preprint arXiv:1901.07441*, 2019.
- [15] Jeremy Irvin, Pranav Rajpurkar, Michael Ko, Yifan Yu, Silvana Ciurea-Ilcus, Chris Chute, Henrik Marklund, Behzad Haghgoo, Robyn Ball, Katie Shpanskaya, et al. Chexpert: A large chest radiograph dataset with uncertainty labels and expert comparison. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 33, pages 590–597, 2019.
- [16] Alistair EW Johnson, Tom J Pollard, Seth Berkowitz, Nathaniel R Greenbaum, Matthew P Lungren, Chihying Deng, Roger G Mark, and Steven Horng. Mimic-cxr: A large publicly available database of labeled chest radiographs. *arXiv preprint arXiv:1901.07042*, 1(2), 2019.
- [17] Dina Demner-Fushman, Marc D Kohli, Marc B Rosenman, Sonya E Shooshan, Laritza Rodriguez, Sameer Antani, George R Thoma, and Clement J McDonald. Preparing a collection of radiology examinations for distribution and retrieval. *Journal of the American Medical Informatics Association*, 23(2):304–310, 2016.
- [18] Anna Majkowska, Sid Mittal, David F Steiner, Joshua J Reicher, Scott Mayer McKinney, Gavin E Duggan, Krish Eswaran, Po-Hsuan Cameron Chen, Yun Liu, Sreenivasa Raju Kalidindi, et al. Chest radiograph interpretation with deep learning models: Assessment with radiologist-adjudicated reference standards and population-adjusted evaluation. *Radiology*, 294(2):421–431, 2020.
- [19] Cyril Goutte and Eric Gaussier. A probabilistic interpretation of precision, recall and f-score, with implication for evaluation. In *European Conference on Information Retrieval*, 2005.
- [20] Charles X Ling, Jin Huang, Harry Zhang, et al. Auc: a statistically consistent and more discriminating measure than accuracy. In *Ijcai*, volume 3, pages 519–524, 2003.

- [21] Sinno Jialin Pan and Qiang Yang. A survey on transfer learning. *IEEE Transactions on knowledge and data engineering*, 22(10):1345–1359, 2009.
- [22] Andre Esteva, Brett Kuprel, Roberto A Novoa, Justin Ko, Susan M Swetter, Helen M Blau, and Sebastian Thrun. Dermatologist-level classification of skin cancer with deep neural networks. *Nature*, 542(7639):115–118, 2017.
- [23] Adam Paszke, Sam Gross, Francisco Massa, Adam Lerer, James Bradbury, Gregory Chanan, Trevor Killeen, Zeming Lin, Natalia Gimelshein, Luca Antiga, et al. Pytorch: An imperative style, high-performance deep learning library. In *Advances in Neural Information Processing Systems*, pages 8024–8035, 2019.
- [24] Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.
- [25] Ramprasaath R Selvaraju, Michael Cogswell, Abhishek Das, Ramakrishna Vedantam, Devi Parikh, and Dhruv Batra. Grad-cam: Visual explanations from deep networks via gradient-based localization. In *Proceedings of the IEEE international conference on computer vision*, pages 618–626, 2017.
- [26] Laurens van der Maaten and Geoffrey Hinton. Visualizing data using t-sne. *Journal of machine learning research*, 9(Nov):2579–2605, 2008.

Individual contributions

- Xuehai He

He worked with others on the selection of the project idea and producing ppt. He worked on the writing of the code and the report. He also did the last part of the final representation

- Fan Wang

He is responsible for collecting all information resources and producing ppt. He also did the introduction part of the final representation.

- Xiaoli Yu

He worked on brainstorming the ideas and verified viability. He worked on the presentation and ppt. He worked on the report. He also did the second part of the final representation.

Replies to critical reviews

Critical review from team 10:

Your work is overall excellent and meaningful! You choose COVID-19 diagnosis, a pretty new (so also challenging) topic, for deep learning application, and this is awesome and exciting. Your chosen dataset seems to be a good choice for this topic due to its good quality and well-labeled virtue.

The algorithm you presented is also effective and applicable to this topic. You introduce two well-developed CNN network architectures – ResNet and DenseNet succinctly but clearly.

Your project also attracts us due to its great breadth. You conduct a comprehensive experiments among ResNet, DenseNet, and EfficientNet. You also present your models with result from either transfer learning or randomly initialized. You make a comparison among these results and give reasonable results. Your results show that transfer learning and DenseNet is effective for CT-based COVID-19 diagnosis.

However, there are some unclear parts in your presentation video which can be improved further. First, you show a very interesting feature visualization and classification distance figure. They are intuitive and can be self-explained, but we are hoping to see the runtime result (in your code live-run) of these figures and a more detailed explanation of how you get them.

Moreover, you show the comparison results among different networks, they are good to read, but we are interested in the explanation of why they may get these results (more concretely, why DenseNet beats the other two? Do you have other proof besides the result numbers?).

Last, although the breadth of this project is awesome, can you dive into the method of your final choice deeper (basically means you just fix to use the transfer learned DenseNet, amend it and see if you can get better results)?

You generally do a good job and develop a well applicable and effective method for COVID-19 diagnosis, thank you for your efforts!

Our response:

1. The GradCam code and TSNE code for figures will be appended in the software when submitting the report.

2. The DenseNet beats the other two because it contains shorter connections between layers close to the input and those close to the output. There are also papers introducing the superiority of DenseNet over medical problems.

3. The performance can be further improved, we will show it in our reports.

Critical review from team

21:

Advantages:

1. The project of this group has a comprehensive introduction of model. And why they use these models.
2. They have compared the COVID recognition accuracy using different models. Then has a good result and conclusion.

Some improvements/unclear:

1. They can use plot or graph to make the result much more direct to be seen. With the table they provide, the results are not intuitive.
2. The author only do a survey about transfer learning, missing part about CT classification.
3. They don't explain why different models have different performance.

Our response:

1. The table is actually quite clear about the performance actually. Even top conferences

like Neurips, CVPR, ICCV have many papers using tables to present their work, not to say this stupid course.

2. We use transfer learning to do CT classification. What does 'missing part about CT classification' means? The detailed survey will be given in the report.

3. Due to the time limitation, it is impossible to explain all about these. In one word, because of model structure and it is quite obvious.

Critical review from team

26:

- The use of GRAD-CAM was very informative. The images of the chest X-rays with the heatmap of features allowed for greater clarity. Would have liked some type of focus on the ethos behind why this is useful. Maybe some kind of linking to how this is useful from a scientific or medical diagnosis perspective beyond just the ML relevance.
- Feature extraction slide 12- the idea seemed very great, but it was unclear what this slide was trying to show visually. What do you mean by the distance between the Covid and non-covid features? What exactly are the images we are looking at. Are they layers of the net? Could do a better job relating it to the physiology of lungs maybe? It's understandable if the team doesn't have any biology experts, but just a thought.
- A lot of the details like data augmentation etc could've been explained in the presentation instead of a line by line of these subjects in the demo.
- For something like the Covid19 dataset a full confusion matrix would be very useful. Since this is a medical diagnosis a full understanding of the workings of the model on each specific case is very important beyond just accuracy or true positive rates.
- Overall, the machine learning parts of the project were very well done, however I felt that there could be a little bit more done to link those aspects of the project with the science/medicine parts of the project, to truly highlight the application.

Our response:

1. We use CT scans, but not xray-images. This is useful as it shows the features selection behind the classification.

2. Please google what is TSNE plot.

3. The data augmentation techniques will be given in detail in the report.

4. We have shown the TP, FP, TN, FN in the slides actually.

5. We will give some in the report.