Replies to critical reviews

Critical review from team 10:

The presentation mentioned that for ResNet50 and Inception_V3 architecture, the models fail to achieve high classification accuracy. You mentioned the possible reason might be the dataset is not large enough to support such comprehensive.

Our response: We used some methods such as flip to do the data augmentation. But since the chest x-ray images are generated under a specific requirement of position and the patient's situation, we think that only by general data augmentation methods such as rotating the images etc. might not be so helpful and meaning for this specific input data. We would try to realize the complicated models when we get more data from COVID-19 soon since it is still updating these days.

Critical review from team 26:

It is interesting to note that the validation accuracy of the ResNet18 model is coming higher than the training accuracy throughout the sample

Our response: If the distribution of the training set and the test set is not uniform, the model might correctly capture the internal distribution pattern of the data, which may cause the internal variance of the training set to be greater than the validation set, and cause a greater error in the training set. We'll try to better divide our dataset when we get more data from COVID-19.

Critical review from team 32:

It's kind of hard to get the big picture of the work for someone without a biology background. I could understand why they wanted to use ML and how they used ML, but not the subject they are talking about. But I guess that this is not the point of the course?

Our response: Our project basically is to classify patients' chest x-ray images and tell whether they have pneumonia or COVID-19. We'll try to explain more clearly about the project subject next time.

COVID-19 DIAGNOSIS WITH CHEST X-RAY IMAGES USING DEEP LEARNING GROUP 21

Xiangwei Shao[⋆†], *Siyi Wang*[⋆†], *and Fei Xue*[⋆†]

[⋆]University of California San Diego, La Jolla, CA 92093,
[†]Department of Electrical and Computer Engineering

ABSTRACT

Due to the outbreak of Covid-19 and the possible shortage of test kits, an alternative method for rapid and accurate diagnosis is needed. Chest X-ray is a direct and useful method for diagnosing pneumonia and other lung diseases. In this project, we try to use deep learning method and chest x-ray dataset to train a classifier to help diagnose Covid-19 and compare the performance of different models. The result shows with proper model, using deep learning method can achieve good accuracy.

Index Terms— deep learning, convolutional neural network, Covid-19, computer aided diagnosis

1. INTRODUCTION

The coronavirus has affected more than 6.3 million people globally, with the U.S. experiencing the most Covid-19 cases at over 1.8 million so far. The rapid spread of Covid-19 poses serious threats to the public health and economies of nations. The mainly used method to diagnose Covid-19 is to use RT-PCR test and Antibody test. But as is known many countries have faced shortage of test kit when it outbreaked. So we believe it is important to have an alternative method for rapid and accurate diagnosis of patients with Covid-19.

Chest X-ray is a direct and useful method for diagnosing pneumonia and other lung diseases. In January, a Chinese team published a paper reporting the epidemiological, clinical, laboratory, and radiological characteristics and treatment and clinical outcomes of these patients. [1]. The typical findings of chest CT images of ICU patients on admission were bilateral multiple lobular and subsegmental areas of consolidation.The representative chest CT findings of non-ICU patients showed bilateral groundglass opacity and subsegmental areas of consolidation.Later chest CT images showed bilateral ground-glass opacity, whereas the consolidation had been resolved. All the patients present abnormalities in chest CT images [1].

But traditionally, even for an experienced doctor on radiology, it take time and effort to analyze the images, which makes it not efficient enough given the rapid spread of Covid-19 and limited hospital resources.

Using machine learning method to help detect abnormalities has been investigated. So in this project we try to use deep learning method and chest x-ray dataset to train a classifier to help diagnose Covid-19.

1.1. Convolutional neural network

Due to the development of deep learning (DL), many fields are trying to use DL methods to solve tasks [2]. Convolutional Neural Networks (CNNs) is a widely used deep learning algorithms and the most prominent category of neural networks[3]. It has a multi-layer neural network architecture. Bascially it contains convolutional layer(s) pursued by fully connected layer(s) and subsampling layers may or may not exist. Using the convolution layer and filters, CNN can obtain the features of input data and meanwhile reducing the number of parameters. This makes it perform good espicially on high dimensional data, like images and videos[3].

1.2. Chest X-Ray Projection

The chest x-ray is anecdotally thought to be the most frequently-performed radiological investigation globally. There are several different projection methods, including: a) PA Views, which is the standard frontal chest projection and also the best general radiographic technique to examine the lungs. The x-ray beam traverses the patient from posterior to anterior and patients must be able to stand during the projection, b)AP Views, which is an alternative frontal projection to the PA projection with the beam traversing the patient from anterior to posterior. It is more convenient for sick patients who will not be able to stand since it can be performed with the patient sitting up on the bed, but has worse visualization. And there's also supine AP views which is used for patients who are not able to sit up.[4]

2. RELATED WORK

Although Covid-19 is new to us. There are already some works on computer aided diagnosis using machine learning.

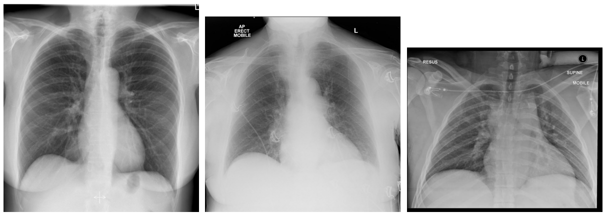

Fig. 1. Different Projection Methods for Chest X-Ray Images(a) PA Views (b) AP Views (c) Supine AP Views.

Many works using deep learning on medical image tasks has proved to achieve good performance, including skin canver classification [5], diabetic retinopathy detection [6] and so on.

For using machine learning on chest X-ray images, [7]applied the general image feature extraction method using quasi-Gabor filter on the radiographs then use a k-NN classifier to detect abnormal texture in chest X-rays. [8] segments image and uses Local binary pattern (LBP) features to classify normal and pathology on chest X-rays. [9] employed a cascade of CNN and RNN to classify abnormalities. CheXNet in [10] builds a 121-layer CNN to classify among 14 diseases using chest xrays. It is trained on ChestX-ray14, chest Xray dataset containing over 100,000 frontal view X-ray images with 14 diseases.

3. DATASETS AND FEATURES

3.1. Datasets

In our project, we aim at classifying chest X-ray images from different kinds of cases, which are normal, pheumonia and COVID-19. Thus we collected data from two different existed datasets and merged them.

The first dataset we choose is the chest x-ray dataset from Kaggle, which contains thousands of images for both normal and pheumonia cases. And the second one we use is a new dataset about covid19 chest x-ray images, which contains 182 images of confirmed covid19 cases of PA Views. In this project, since all of the chest x-ray images from Kaggle are of PA views, so in the COVID-19 dataset, we filtered out all of the images of other projection methods to ensure that our input data of the training process is of the same type and thus better classification precision.

Our task is to classify these images into the right classes. We firstly do the binary classification to classify if it's a normal case or not(pheumonia and COVID-19). In this part, since we have large amount of data from Kaggle, we sampled 1500 images from both classes and merged images of pheomonia with all 182 images from COVID-19. Thus, our dataset is of 3182 images for binary classification. Then to further classify images from pheomonia and COVID19, we did the multiple classification too. Since the size of COVID-19 dataset is currently relatively small, in order to keep the balance of data for these three classes, we reduce the images for both normal and pheomonia classes to 500 images, so our dataset is of 1182 images for multiple classification.

3.2. Features

Since all of our data are images. To further be used as input of CNNs, we resized them to size of 224*224. And we further randomly flip the images to realize general data augmentation and normalized them to [0.485, 0.456, 0.406] as the input features for CNNs.

4. METHODS

We have used several famous and effective models which achieve very good performance in other situation. More specifically, we used VGG16, ResNet18, ResNet50 and Inception_V3 as our network to be trained to diagnosis with chest X-ray.

4.1. VGG16

VGG is a convolutional neural network model proposed by Simonyan and Zisserman in the document "Very Deep Convolutional Networks for Large Scale Image Recognition", whose name comes from the abbreviation of the Visual Geometry Group of Oxford University where the author is located. The model participated in the 2014 ImageNet image classification and positioning challenge and achieved excellent results: it ranked second in the classification task and ranked first in the positioning task. The fig.2 shows the six structural configurations of VGG: In the figure above, each column corresponds to a structural configuration. The green part in the figure indicates the structure adopted by VGG16. We conducted a specific analysis of VGG16 and found that VGG16 contains a total of 13 convolutional layers, 3 fully connected layers and 5 pool layers. VGG has the following features: (1) All convolutional layers use 3*3 convolution kernels. (2) The model is composed of a stack of convolutional layers and pooling layers, and it is relatively easy to form a deep network structure. (3) The convolutional layers and pooling layers form a block structure. Each block contains several convolutional layers and a pooling layer. VGG16 has a very simple network structure but will consume more computing resources.

4.2. ResNet18 & ResNet50

Within a certain range, increasing the depth of the network will improve the accuracy of the network. However, when the network depth is too large, there will be a degradation problem, that is, as the depth increases, the performance of the network will become worse and worse, directly reflected in the accuracy of the training set will decline. The ResNet network solves this problem, and after the problem is solved, the depth of the network has increased by several orders of

ConvNet Configuration						
A	A-LRN	B	C	D	E	
11 weight layers	11 weight layers	13 weight layers	16 weight layers	16 weight layers	19 weight layers	
input (224 × 224 RGB image)						
conv3-64	conv3-64 **LRN**	conv3-64 **conv3-64**	conv3-64 conv3-64	conv3-64 conv3-64	conv3-64 conv3-64	block 1
maxpool						
conv3-128	conv3-128	conv3-128 **conv3-128**	conv3-128 conv3-128	conv3-128 conv3-128	conv3-128 conv3-128	block 2
maxpool						
conv3-256 conv3-256	conv3-256 conv3-256	conv3-256 conv3-256	conv3-256 conv3-256 **conv1-256**	conv3-256 conv3-256 **conv3-256**	conv3-256 conv3-256 conv3-256 **conv3-256**	block3
maxpool						
conv3-512 conv3-512	conv3-512 conv3-512	conv3-512 conv3-512	conv3-512 conv3-512 **conv1-512**	conv3-512 conv3-512 **conv3-512**	conv3-512 conv3-512 conv3-512 **conv3-512**	block4
maxpool						
conv3-512 conv3-512	conv3-512 conv3-512	conv3-512 conv3-512	conv3-512 conv3-512 **conv1-512**	conv3-512 conv3-512 **conv3-512**	conv3-512 conv3-512 conv3-512 **conv3-512**	block5
maxpool						
FC-4096						
FC-4096						
FC-1000						
soft-max						

Fig. 2. Structure of VGG Network[11]

magnitude.The core of ResNet is the residual learning module. The structure of this module is shown in the Fig.3. The residual module can make the network in the forward process, when the output of the shallow layer has been optimized, so that the layers behind the deep network can achieve the role of identity mapping. In addition, the redidual module will significantly reduce the value of the parameters in the module, so that the parameters in the network have a more sensitive response to the value of the reverse conduction loss. Although the problem of return loss is not fundamentally solved, the parameter Decreasing, relatively speaking, increases the effect of return loss, and also produces a certain regularization effect. The configuration pf ResNet network is shown in the Fig.4.

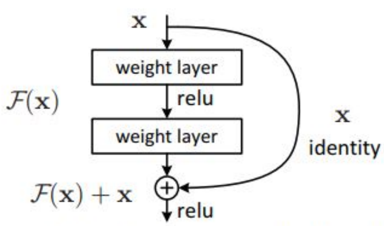

Figure 2. Residual learning: a building block.

Fig. 3. Structure of Residual Module[12]

layer name	output size	18-layer	34-layer	50-layer	101-layer	152-layer
conv1	112×112	7×7, 64, stride 2				
conv2_x	56×56	3×3 max pool, stride 2				
		$\begin{bmatrix} 3\times3, 64 \\ 3\times3, 64 \end{bmatrix}\times2$	$\begin{bmatrix} 3\times3, 64 \\ 3\times3, 64 \end{bmatrix}\times3$	$\begin{bmatrix} 1\times1, 64 \\ 3\times3, 64 \\ 1\times1, 256 \end{bmatrix}\times3$	$\begin{bmatrix} 1\times1, 64 \\ 3\times3, 64 \\ 1\times1, 256 \end{bmatrix}\times3$	$\begin{bmatrix} 1\times1, 64 \\ 3\times3, 64 \\ 1\times1, 256 \end{bmatrix}\times3$
conv3_x	28×28	$\begin{bmatrix} 3\times3, 128 \\ 3\times3, 128 \end{bmatrix}\times2$	$\begin{bmatrix} 3\times3, 128 \\ 3\times3, 128 \end{bmatrix}\times4$	$\begin{bmatrix} 1\times1, 128 \\ 3\times3, 128 \\ 1\times1, 512 \end{bmatrix}\times4$	$\begin{bmatrix} 1\times1, 128 \\ 3\times3, 128 \\ 1\times1, 512 \end{bmatrix}\times4$	$\begin{bmatrix} 1\times1, 128 \\ 3\times3, 128 \\ 1\times1, 512 \end{bmatrix}\times8$
conv4_x	14×14	$\begin{bmatrix} 3\times3, 256 \\ 3\times3, 256 \end{bmatrix}\times2$	$\begin{bmatrix} 3\times3, 256 \\ 3\times3, 256 \end{bmatrix}\times6$	$\begin{bmatrix} 1\times1, 256 \\ 3\times3, 256 \\ 1\times1, 1024 \end{bmatrix}\times6$	$\begin{bmatrix} 1\times1, 256 \\ 3\times3, 256 \\ 1\times1, 1024 \end{bmatrix}\times23$	$\begin{bmatrix} 1\times1, 256 \\ 3\times3, 256 \\ 1\times1, 1024 \end{bmatrix}\times36$
conv5_x	7×7	$\begin{bmatrix} 3\times3, 512 \\ 3\times3, 512 \end{bmatrix}\times2$	$\begin{bmatrix} 3\times3, 512 \\ 3\times3, 512 \end{bmatrix}\times3$	$\begin{bmatrix} 1\times1, 512 \\ 3\times3, 512 \\ 1\times1, 2048 \end{bmatrix}\times3$	$\begin{bmatrix} 1\times1, 512 \\ 3\times3, 512 \\ 1\times1, 2048 \end{bmatrix}\times3$	$\begin{bmatrix} 1\times1, 512 \\ 3\times3, 512 \\ 1\times1, 2048 \end{bmatrix}\times3$
	1×1	average pool, 1000-d fc, softmax				
FLOPs		1.8×10^9	3.6×10^9	3.8×10^9	7.6×10^9	11.3×10^9

Fig. 4. Structure of ResNet Network[12]

4.3. Inception_V3

After 2014, the increase in model size became the main research method to improve model performance (VGG, GoogLeNet). Although VGG has achieved good performance, it is too computationally intensive; GoogLeNet's Inception architecture can achieve very good performance under strict constraints on memory and computational budget. In addition, GoogLeNet has only 5 million parameters (1/12 AlexNet parameter amount). The parameter amount of VGG is 3 times that of AlexNet. Inception is less computationally intensive than VGG, but has higher performance. This makes it possible to use Inception in the context of big data or mobile environments.The design criteria of inception is (1) to avoid representation bottlenecks, especially the layers in front of the network. The representation is the activations, and the size of the activations should gradually decrease. (2) Higher-dimensional representations are easier to handle and easier to train (converge). (3) Spatial aggregation on the low-dimensional embedding space (Spatial aggregation) hardly affects the representation ability. The explanation is that there is a strong correlation between connected neurons, and the information is redundant. (4) Balance of network width and depth. The balance of the two can bring better performance.

5. EXPERIMENTS & DISCUSSION

We have conducted two set of experiments with VGG16, ResNet18, ResNet50 and Inception_V3 network mentioned above. The experiments are binary classification task and mutiple classification task. For binary classification task, We try to tell whether the person had pneumonia from the chest X-ray (COVID-19 and other pneumonia). For multiple classification task, We try to tell whether the person had pneumonia from the chest X-ray and the specific type of pneumonia(COVID-19 or other pneumonia). Our experiment environment is UCSD datahub with python3, 1 GPU, 8 CPU and 16G RAM. The results of experiments are as follows:

5.1. VGG16

From the results in Fig.5, we an see that the performance of VGG16 is pretty well. The classification correctness can be more than 90%. And the loss is also very low.

5.2. ResNet18

From the results in Fig.6 , we can see that the performance of ResNet18 is even better than VGG16. The classification correctness can reach 95%, which is a great success.

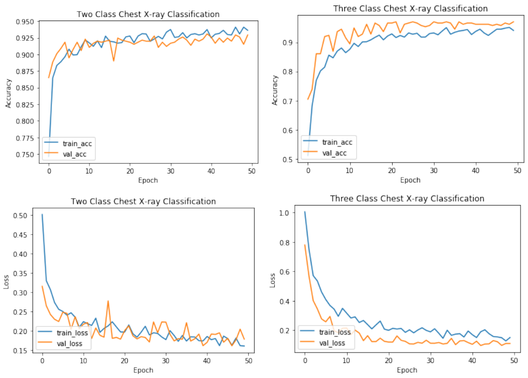

Fig. 5. Results of VGG16 experiments

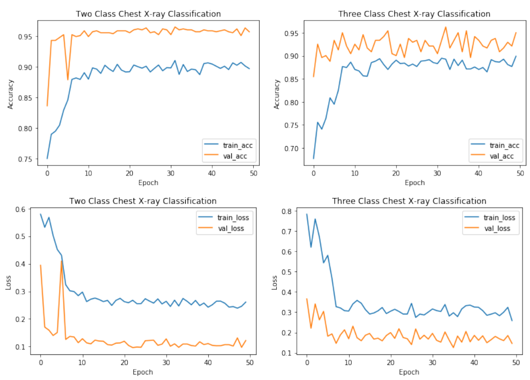

Fig. 6. Results of ResNet18 experiments

5.3. ResNet50

But for ResNet50, as Fig.7 showing, things are going bad. We can see the validation_accuracy is very low and validation loss is large while training result is opposite.

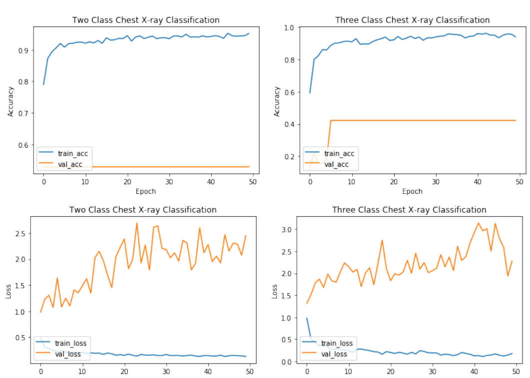

Fig. 7. Results of ResNet50 experiments

5.4. Inception_V3

For Inception_V3, result in Fig.8, it is not very good and the plots fluctuate very much. Clearly, overfitting occurs when we use ResNet50 and Inception version3.

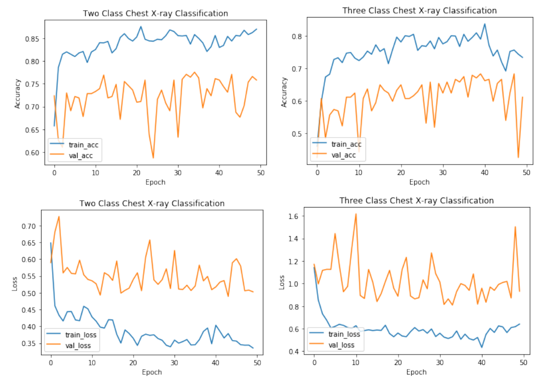

Fig. 8. Results of Inception_V3 experiments

5.5. Discussion

From the results above, we can see that when using VGG16 and ResNet18, the classification accuracy can be more than 90%, which verifies the feasibility of distinguishing COVID-19 and common pneumonia with deep learning. But when using ResNet50 and Inception_V3, the results are bad. This is because ResNet50 has 50 layers and Inception_V3 has 47 layers while our dataset is not large enough to support such complicated models. So overfitting happens, then our classification tasks failed.

6. CONCLUSION & FUTURE WORK

In this project, we tried different CNN models to classify chest x-ray images. Results of VGG16 and ResNet18 show that CNN does well in classifying images. But when the models become more complicated, it will need more input data to support the training or the result will be relatively worse. Our future work is to find ways to extend our data, like collecting more covid-19 chest x-ray data as it is still updating, to further see how complicated models perform. And we want also tune our parameters more carefully to get the best performance for each model.

7. REFERENCES

[1] Chaolin Huang, Yeming Wang, Xingwang Li, Lili Ren, Jianping Zhao, Yi Hu, Li Zhang, Guohui Fan, Jiuyang Xu, Xiaoying Gu, Zhenshun Cheng, Ting Yu, Jiaan Xia, Yuan Wei, Wenjuan Wu, Xuelei Xie, Wen Yin, Hui Li, Min Liu, and Bin Cao. Clinical features of patients infected with 2019 novel coronavirus in wuhan, china. *The Lancet*, 395, 01 2020.

[2] Yann LeCun, Y. Bengio, and Geoffrey Hinton. Deep learning. *Nature*, 521:436–44, 05 2015.

[3] Arohan Ajit, Koustav Acharya, and Abhishek Samanta. A review of convolutional neural networks. pages 1–5, 02 2020.

[4] Andrew Murphy. Chest (ap erect view). [EB/OL]. https://radiopaedia.org/articles/chest-ap-erect-view-1?lang=us/ Accessed April 4, 2010.

[5] Andre Esteva, Brett Kuprel, Roberto Novoa, Justin Ko, Susan Swetter, Helen Blau, and Sebastian Thrun. Dermatologist-level classification of skin cancer with deep neural networks. *Nature*, 542, 01 2017.

[6] Varun Gulshan, Lily Peng, Marc Coram, Martin Stumpe, Derek Wu, Arunachalam Narayanaswamy, Subhashini Venugopalan, Kasumi Widner, Tom Madams, Jorge Cuadros, Ramasamy Kim, Rajiv Raman, Philip Nelson, Jessica Mega, and Dale Webster. Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs. *JAMA*, 316, 11 2016.

[7] Mira Park, Jesse S. Jin, and Laurence S. Wilson. Detection of abnormal texture in chest x-rays with reduction of ribs. In *VIP*, 2003.

[8] Juan Manuel Carrillo de Gea, Ginés García-Mateos, José Fernández-Alemán, and José Hernández. A computer-aided detection system for digital chest radiographs. *Journal of Healthcare Engineering*, 2016:1–9, 05 2016.

[9] Hoo-Chang Shin, Kirk Roberts, Le Lu, Dina Demner-Fushman, Jianhua Yao, and Ronald Summers. Learning to read chest x-rays: Recurrent neural cascade model for automated image annotation. 03 2016.

[10] Pranav Rajpurkar, Jeremy Irvin, Kaylie Zhu, Brandon Yang, Hershel Mehta, Tony Duan, Daisy Ding, Aarti Bagul, Curtis Langlotz, Katie Shpanskaya, Matthew Lungren, and Andrew Ng. Chexnet: Radiologist-level pneumonia detection on chest x-rays with deep learning. 11 2017.

[11] Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*, 2014.

[12] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016.

Individual contributions

- Siyi Wang

  She worked on data processing, model construction and report writing

- Xiangwei Shao

  He worked on model tuning, writing report and preparing slides

- Fei Xue

  He worked on model construction, model tuning and report writing