ST4090 Written Report

Christopher Walter Brennan

115396296

Supervisor - Supratik Roy

Pneumonia Detection in Chest Radiographs using Convolution Neural Networks

School of Mathematical Sciences
University College Cork
Ireland
May 2019

Abstract

The use of Machine Learning and Artificial Intelligence within daily life is becoming more pervasive. The motivation of this work is to investigate how these new technologies can become even more pervasive by being able to detect diseases within humans. The example looked at in this project is using convolution neural networks to detect pneumonia on Chest Radiographs (CXR). Pneumonia appears as 'opacities' or cloudy areas with no well-defined boundaries on a scan of lungs and a convolution neural network will be trained on a previously annotated set of CXRs in order to detect pneumonia. In order to enhance the process, transfer learning was utilised in the form of using a model previously trained on the COCO (Common Objects in Context) data set. This Mask R-CNN (Region Convolutional Neural Network) is one of the best performing algorithms on the COCO data set and is easy to generalise and therefore is a good base for a model for pneumonia detection. While the ultimate goal may not be to replace a human, this technology certainly has a place within medicine to overcome fatigue-based human error and provide assistance where radiologists are in short supply.

Contents

1 Introduction

1.1 Introduction and Motivation

Science fiction has long touted the rise of the computer to the level in which it could replace a human being. Whether or not such a rise would create a dystopia or not remains to be seen. But with each passing day, this idea is becoming more and more of a reality. However, alongside potentially the software developers and engineers that create such hyper-intelligent robots or computers, one industry stands out for its resilience to the automated revolution. That industry is the medical industry. This may be partly down to an inherent mistrust of technology in such an important situation, often a life or death situation. But it is also largely down to the fact that the technology to make decisive medical decisions to a consistently accurate level is lacking. However, advances in machine learning are happening at such a massive rate that it won't be long before it is generally accepted that computers are better at tasks such as disease detection and patient prognosis than humans.

The goal of this project is to use machine learning and deep learning to help detect pneumonia using chest radiographs (CXR). Pneumonia accounts for over 15% of all deaths of children under 5 years old internationally. In 2015, 920,000 children under the age of 5 died from the disease. In the United States, pneumonia accounts for over 500,000 visits to emergency departments, and over 50,000 deaths in 2015, keeping the ailment on the list of top 10 causes of death in the country. There is, therefore, a clear motivation for a project that aims to assist in the detection of such a disease. The diagnosis of pneumonia requires a variety of tests, including a radiographer examining the X-ray for regions called 'lung opacities'. However, lung opacities are not the same as pneumonia. They are instead vague, fuzzy clouds of white in the darkness of the lungs. The diagnosis of pneumonia on CXR is complicated due to the vast number of other conditions in the lungs such as pulmonary edema, bleeding, volume loss, lung cancer, or post-radiation or surgical changes. Therefore, the intention is to take a publically available dataset of chest radiographs, train an algorithm on a training set from the dataset, and then test the accuracy of the algorithm on a test set within the dataset (completely distinct from the training set). The algorithm in question is chosen to be a Convolution Neural Network (CNN).

Examples of Machine Learning (ML) within this industry are beginning to become more commonplace, with perhaps the power of deep learning becoming most evident in ophthalmology. Deep learning has been applied to a clinically heterogeneous set of 3-d optical coherence tomography (CT) scans in the U.K. The model exhibited performance in making a referral recommendation that reached, or even exceeded that of experts on a variety of sight-threatening retinal diseases. This performance was achieved after training on only 14,884 scans [1].

Health care automation companies are even beginning to appear with one company IDx developing deep-learning based software for health providers to use when treating patients with diabetes to scan images for signs of diabetic retinopathy [2]. The company has even gone as far as receiving regulatory approval by the U.S Food and Drug Administration for the product [2].

There has even been an example of a similar project to this, albeit on a much bigger scale both in terms of complexity and data set size. The CheXNeXt algorithm is a deep learning algorithm that can be utilised to detect 14 clinically important pathologies including pleural effusion, pulmonary masses and nodules, and pneumonia (the focus of this project) in frontal-view chest radiographs [3]. The algorithm was tuned and validated on subsets of the 'National Institutes of Health ChestX-ray8' data set which includes over 100,000 chest radiographs from approximately 31,000 patients. When performance on a held-out partition consisting of images hand-annotated by a panel of cardiothoracic specialist radiologists was compared to that of 9 radiologists (6 board-certified, 3 residents), performance levels were found to be similar [3]. Most importantly though, was that at comparable accuracies, it took CheXNeXt 1.5 minutes to interpret 420 images in the validation set, whereas it took the radiologists 240 minutes [3]. Despite being in early development, the speed coupled with the precision could assist in reducing fatigue-based diagnostic error, and help with the lack of diagnostic expertise in areas of the world where radiologists are in short supply.

1.2 Pneumonia, Lungs, and Lung Opacities

An important starting point for the project was to understand what a normal pair of lungs looks like on an X-ray, what a lung opacity is, and what it looks like on an X-ray. This was done by inspecting elements of the data set. A normal chest radiograph (CXR) with good technical quality can be seen in Fig. 2.

Figure 1: Basic Anatomy of Lungs

An X-ray works by passing X-rays through the body, which reach a detector on the other side of the body. Tissues with sparse material, such as lungs filled with air, do not absorb the X-rays and

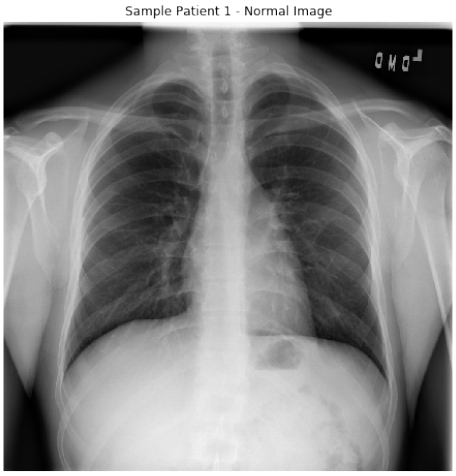

Figure 2: CXR with Good Technical Quality

appear black. Dense tissue such as bones absorb the X-rays and appear white. Grey corresponds to tissue or fluid.

Pneumonia is a lung infection that can be caused by bacteria, viruses, or fungi. Because of the infection and the body's immune response, the sacks in the lungs (termed alveoli) are filled with fluids instead of air (see Fig. 3).

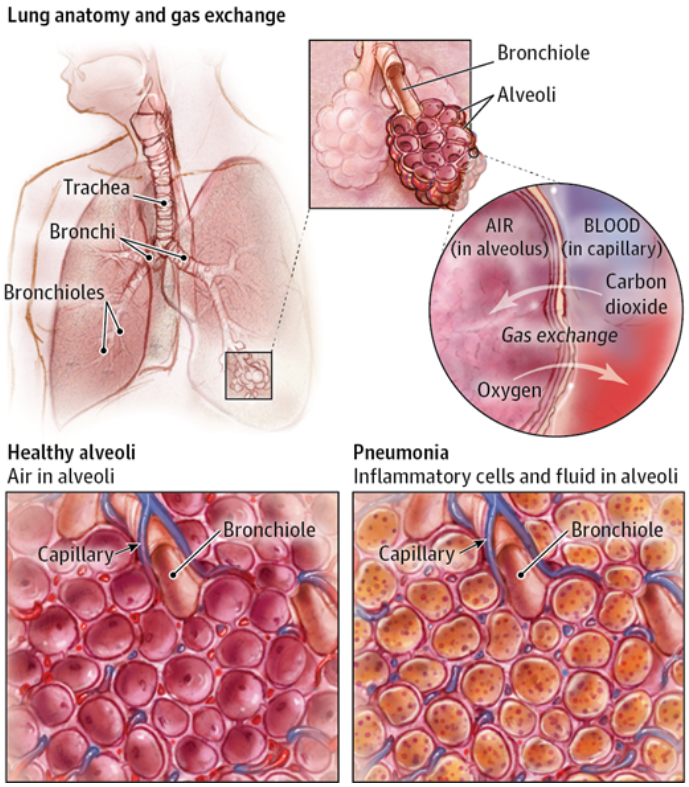

Figure 3: Lung Anatomy and Pneumonia

This causes coughing with phlegm or pus, fever, chills, and difficulty breathing. Pneumonia can range in seriousness from mild to life-threatening with it being most serious for infants, elderly people and those with health issues or weakened immune systems with symptoms and severity depending on the type of germ causing the infection, age, and overall health. A full outline of the potential symptoms can be seen in Fig. 4.

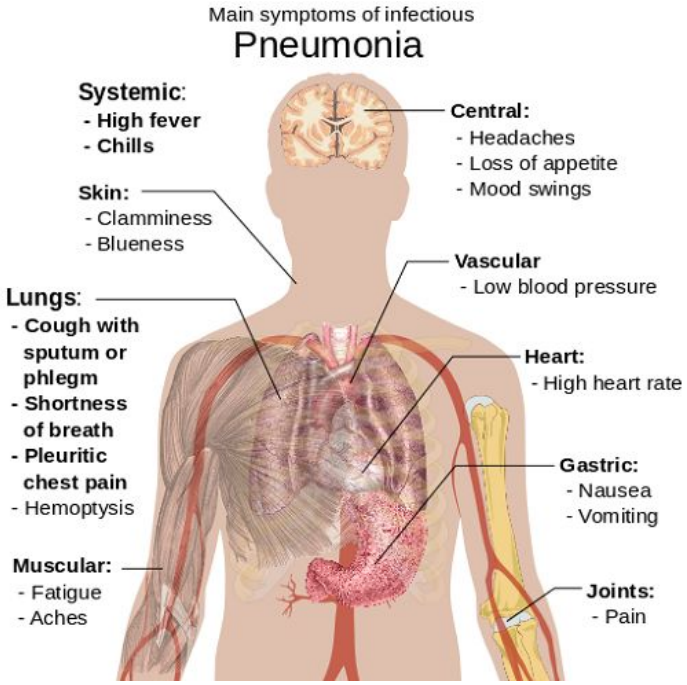

Figure 4: Symptoms Associated with Pneumonia

It is clear from the number of symptoms that a number of tests need to be carried out in order to diagnose pneumonia. One of these tests is a chest radiograph. When trying to detect pneumonia on a chest X-ray, a radiologist will focus on finding lung opacities. A lung opacity is defined to be "any area that preferentially attenuates the X-ray beam and therefore appears more opaque than the surrounding area. It is a nonspecific term that does not indicate the size or pathological nature of the abnormality" [4]. In layman's terms, this means a lung opacity is any area of the chest radiograph that is more white than it should be. Lung opacities are not homogeneous, they do not have a clear centre or clear boundaries. This leads to the issue of the fact there is a known variability in the interpretation of chest radiographs. Studies have shown that there is only a moderate level of agreement between radiologists about the presence of infiltrates, which are opacities by definition [5]. The reason that pneumonia-associated lung opacities look diffuse on the chest radiograph is that the infection and fluid that accumulate spread within the normal tree of airways in the lung. There is no clear border where the infection stops. That is different from diseases like tumours, which present entirely different from the normal lung, and do not maintain the normal structure of the

airways inside the lung. These differences are illustrated by samples from the dataset.

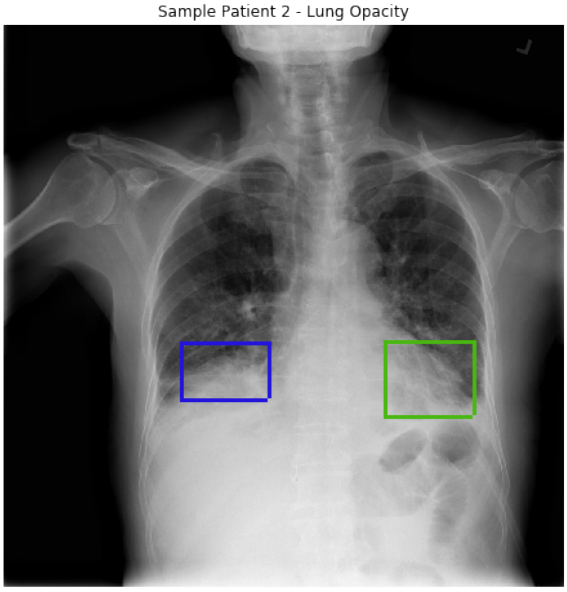

Figure 5: Lung Containing Lung Opacities

Fig. 5 shows a pair of lungs with lung opacities (marked by the boxes). The idea of lung opacities being regions that are lighter than they should be is especially clear when compared to the normal pair of lungs with good technical quality shown in Fig. 2. The regions aren't clearly defined and are essentially 'clouds'. This is in comparison to Fig. 6 which is a pair of lungs of nodules or masses. These areas are more well-defined, with clear edges. The structure of the lungs is clearly changed by these masses and it is therefore clearly not pneumonia.

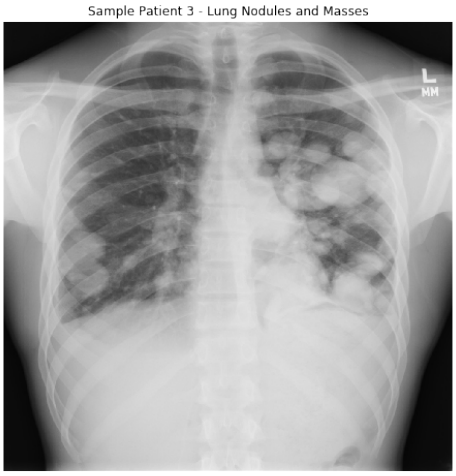

Figure 6: Lungs with Nodules or Masses

These aren't the only abnormalities present within CXRs, which serves to only highlight the difficulty of pneumonia detection; there are such a large variety of possible prognoses based on the abnormality. Fig. 7 shows a normal pair of lungs compared to lungs with pleural effusion. Pleural effusion is the build-up of excess fluid between the layers of the pleura outside the lungs. The pleura are thin membranes that line the lungs and the inside of the chest cavity and act to lubricate and facilitate breathing. On the right, clear fluid can be seen within the left lung, especially when compared to the normal pair of lungs.

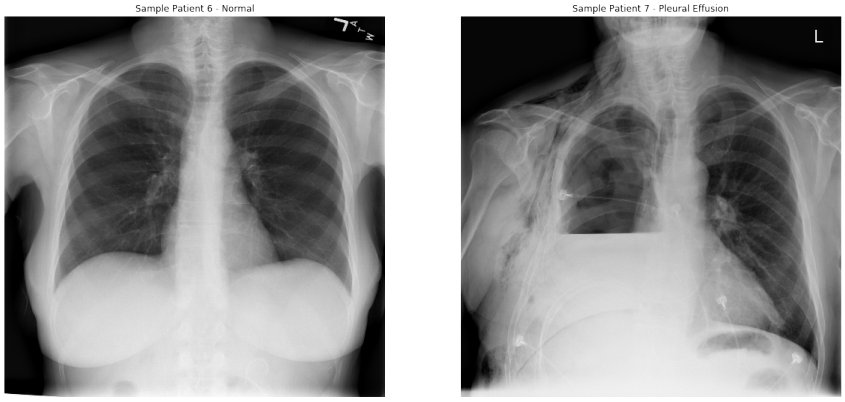

Figure 7: Normal Pair of Lungs (Left) Compared to Lungs with Pleural Effusion (Right)

Possibly the most anomalous of all abnormalities is that seen in Fig. 8. It may seem that these lung opacities are associated with a severe case of pneumonia, given the pervasiveness of the opacity. However, this phenomenon is called 'White Lung' and has a number of different explanations:

- Pneumonectomy - Removal of the lung
- Pneumonia-related lung opacity
- Severe pleural effusion

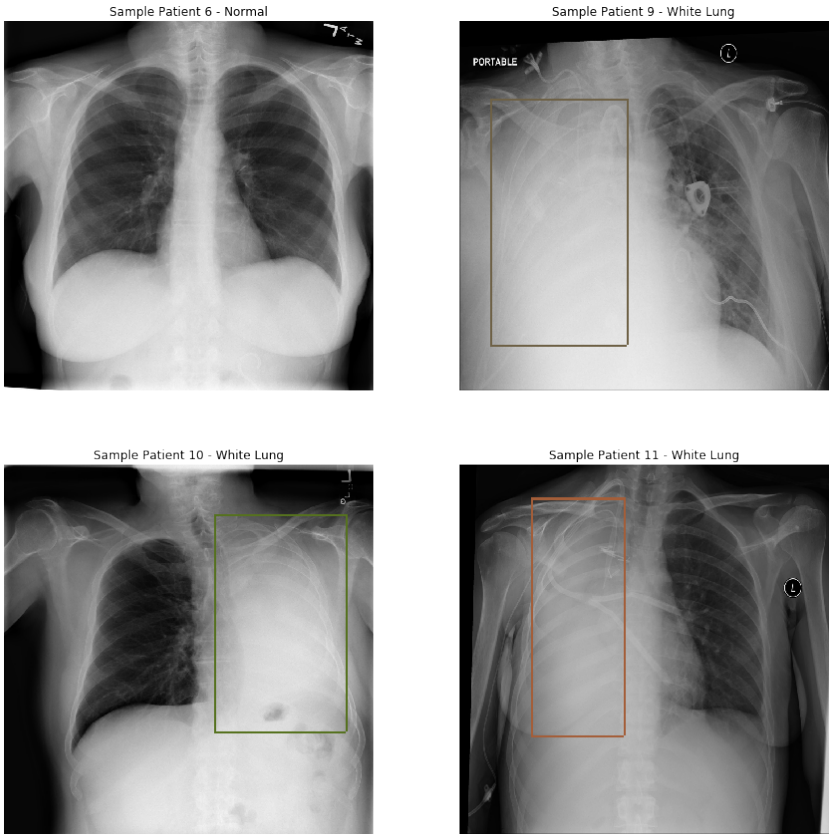

Figure 8: Normal Pair of Lungs Compared to Cases of Removed Lungs

These examples serve not only to clarify the goal of the project and what the deep learning algorithm will be inspecting and attempting to detect but also highlights the difficulty of the problem. The vast number of possible diagnoses, with often very little between deciding factors (often information not provided by the CXR such as other symptoms) creates a very difficult task for the algorithm to learn. However, the data utilised and the variety within it certainly helps to overcome this problem.

2 Data

The data for the project comes from a Radiology Society of North America Kaggle competition for the detection of pneumonia (`https://www.kaggle.com/c/rsna-pneumonia-detection-challenge`). The data was provided in two separate stages, with each stage being in the same format. The files received for Stage 1 were as follows:

- Stage 1 Images
  - stage_1_train_images.zip (training images)
  - stage_1_test_images.zip (test images)
- Stage 1 Labels - stage_1_train_labels.csv
- Stage 1 Sample Submission - stage_1_sample_submission.csv
- Stage 1 Detailed Info - stage_1_detailed_class_info.csv

All images were provided as Digital Imaging and Communications in Medicine (DICOM) files. The relevant data fields for the project are

- **patientId_** - A patientId corresponding to a unique image.
- **x_** - the upper-left x coordinate of the bounding box
- **y_** - the upper-left y coordinate of the bounding box
- **width_** - the width of the bounding box
- **height_** - the height of the bounding box
- **Target_** - Binary target variable indicating whether the sample has evidence of pneumonia

The bounding box data referred to in data fields relates to the box placed around the lung opacity by a radiologist in the training set. The information can be used to enhance machine learning, allowing the computer to know where the anomolies in the CXR are. A second stage of data was released a number of weeks after the first set that was intended to allow participants to improve their model, and submission score relevant to the competition. For use in this project, only the first stage of images was used due to both time and computing constraints. In total stage 1 had 25,684 training images and 1,000 test images. The only Stage 2 file utilised was the Stage 2 Labels and this was due to it containing the 'Target' values for the Stage 1 Test Set. Before running the Mask R-CNN model Exploratory Data Analysis was performed on the data set.

3 Exploratory Data Analysis

An important step in the machine learning process is Exploratory Data Analysis. This is not machine learning in itself, but it does provide useful information that can augment the machine learning process. Exploratory Data Analysis is termed to be a method of analysing data sets in order to summarise their principal characteristics, most often with visual methods. While statistical models can be utilised, EDA is mainly for seeing what the data can show us outside of formal modelling or hypothesis testing. The breakdown of classes for the training set is as follows:

Class	Target	Patient Count
Lung Opacity	1	8964
No Lung Opacity/ Not Normal	0	11500
Normal	0	8525

Table 1: Classes of Training Set

Through looking at the metadata it was discovered that two different view positions existed for the CXR; PA and AP. This was done by looking for the number of unique entries for BodyPartExamined, ViewPosition, and PatientSex, and seeing which was also most frequent:

	BodyPartExamined	ViewPosition	PatientSex
count	25684	25684	25684
unique	1	2	2
top	CHEST	PA	M
freq	25684	13979	14593

Table 2: Uniqueness

It was also noticed, maybe somewhat unsurprisingly, that some images had multiple bounding boxes. The full breakdown was as follows:

Boxes	Patients
1	22506
2	3062
3	105
4	11

Table 3: Uniqueness

It must be noted at this point that evidently, by this result, even the images that didn't have the Target variable set to 1, indicating the presence of pneumonia, had a bounding box in the data, the x, y, width and height variables were simply set to 0. It was obviously important at this stage to look at a sample of images with the bounding boxes imposed upon them to get a general feel for the data. A sample of different images, some with lung opacities, some not normal (but no lung opacities), and some normal images can be seen below in Fig. 9

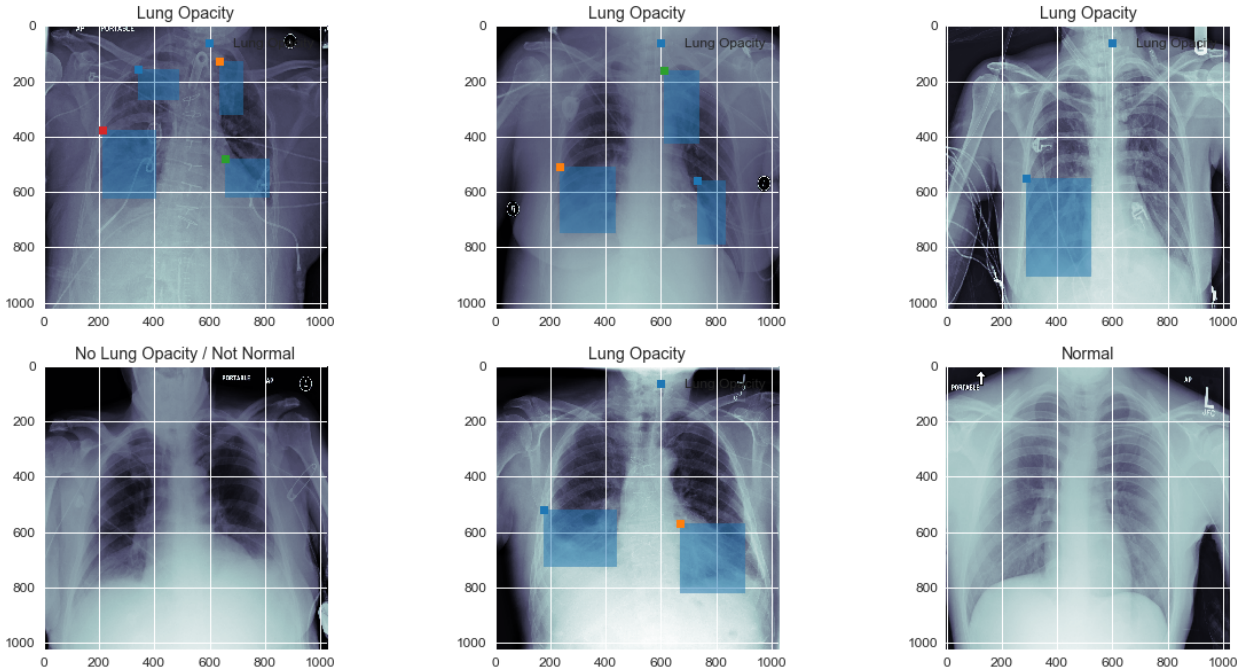

Figure 9: Sample of Images with Bounding Boxes

It was obvious at this stage that it would be worthwhile to see, in general, where the opacities occur in the image. This was done by finding centres of the bounding boxes and plotting these using scatter plots and bar charts as illustrated in Fig. 10. This is in no means a scientific plot, considering the centre of the bounding box doesn't, in general, have any importance to the lung opacity and is

simply imposed by the radiologist that viewed the image. Nevertheless, it is interesting to see the dispersion of the points on the scatterplots to form to ellipse shapes, almost like lungs, and for the centres to most frequently occur in the centre of the two ellipses with relatively little skew. These shapes are only further emphasised by heat maps illustrated in Fig. 11, again showing the two ellipse shapes. And for curiosity, these heat maps were imposed over random images to again get a general feel for the data (see Fig. 12).

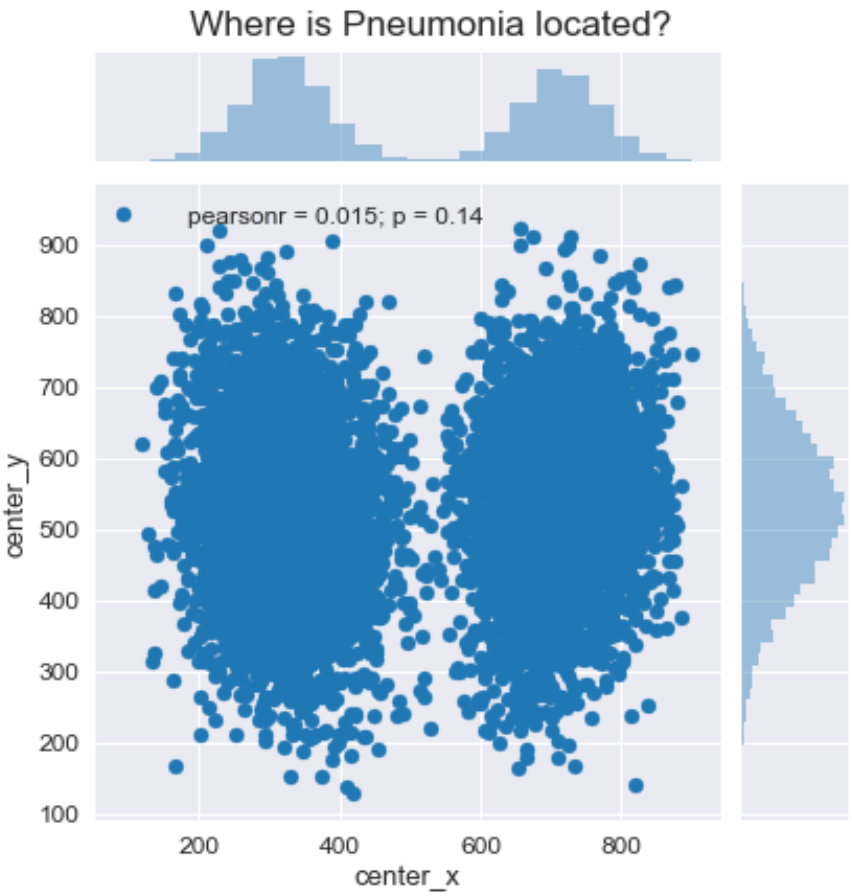

Figure 10: Scatter Plot

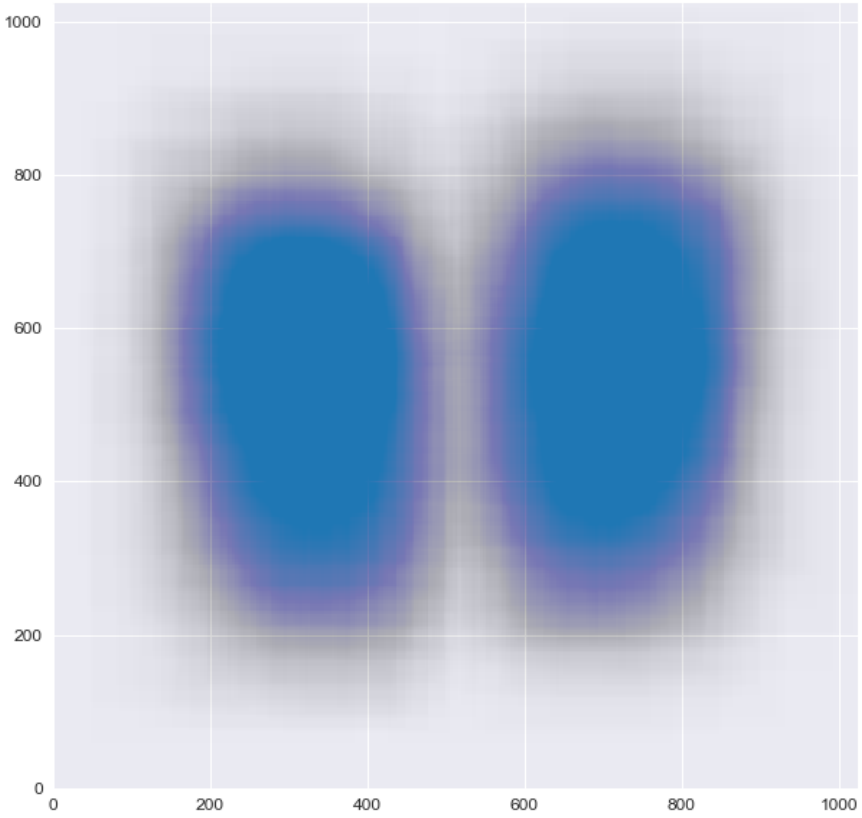

Figure 11: Heatmap of Centres of Bounding Boxes

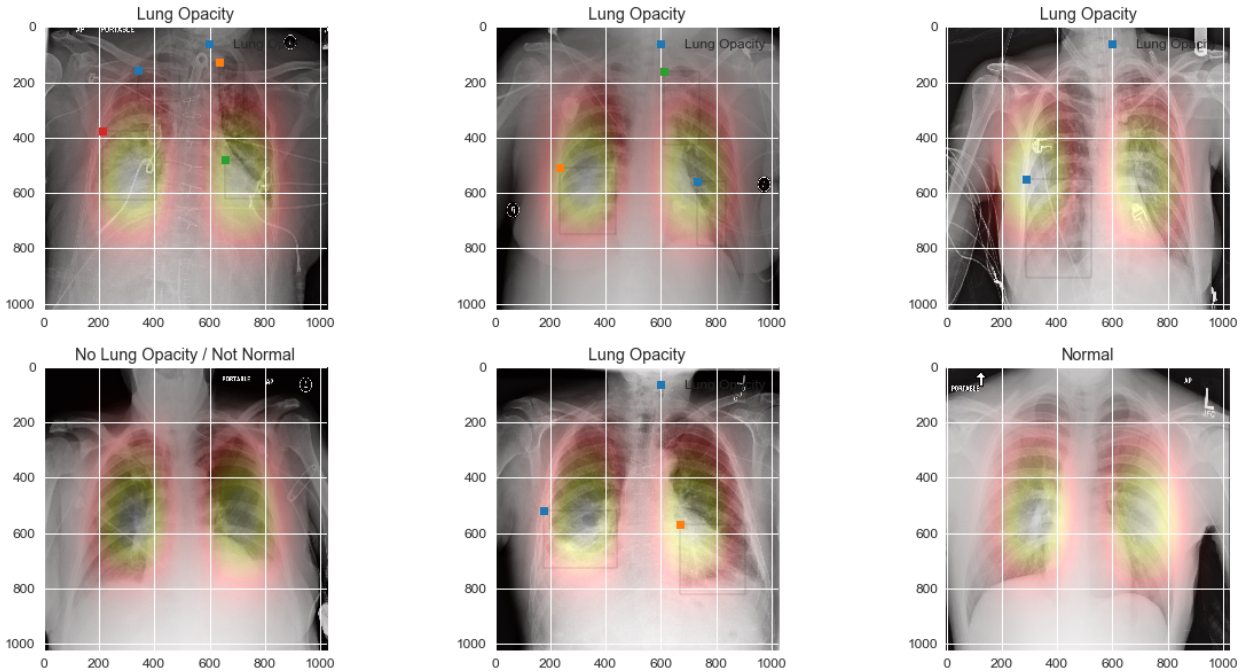

Figure 12: Heatmap Imposed over Sample X-rays

4 Methods

4.1 Mask - Region Convolution Neural Network

Before discussing the method used in this project some background will be provided in order to make the algorithm easier to understand by placing it in context.

4.1.1 Neural Networks

Neural networks are computing systems vaguely inspired by biological neural networks that animal brains are made up of and are therefore more correctly termed as artificial neural networks [6]. Instead of being an algorithm itself, a neural network can be considered as a framework for many different machine learning algorithms, allowing them to work in tandem to process complex data [7]. Neural networks are said to 'learn' to perform certain tasks by being shown examples, and are generally not programmed with any task-specific rules. A neural network consists of nodes (artificial neurons) and connections (synapses in a biological brain) that transmit a signal from one node to another. When a signal is passed to a node it can process it and send a new signal to the artificial neurons it is connected to. In common neural network implementations, the signal is a real number and the output from each node is calculated using some non-linear function of the sum of the inputs. The nodes and connections, or edges, have weights associated with them that change as the network 'learns' from the examples provided to it. The artificial neurons can sometimes have a threshold related to them that only allows the signal to be passed to it if the sum of the signals is greater than the threshold. In order to create some order between the artificial neurons, they are usually split into different layers with different layers often having different transformations on their input. So in reality the signal will travel from the input layer to an output layer, usually after passing through a number of hidden layers, which is illustrated in Fig. 13. Usually, the more complex the task, the more hidden layers are required.

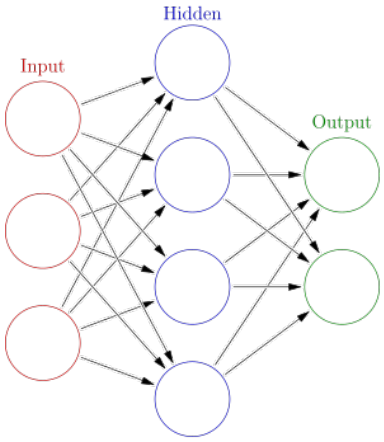

Figure 13: Basic Neural Network (*Image by Glosser.ca on Wikipedia.org*)

4.1.2 Convolution Neural Networks

Convolution Neural Networks (CNN) are a type of deep neural networks that are renowned for their ability to analyse visual imagery. Just like conventional artificial neural networks (ANN), convolution neural networks are based off of biological processes [8] [9]. In this case, the connections between neurons resemble an animal's visual cortex. In an animal, individual cortical neurons will only respond to stimuli in a restricted region called a receptive field, with all receptive fields partially overlapping to cover the entire visual field. Again a CNN will consist of an input layer, hidden layers, and an output layer. The difference from a basic neural network is that a CNN's hidden layers consist of convolution layers, RELU layers, pooling layers, fully connected layers, and normalisation layers. This all leads to a number of differences from a basic neural network. Firstly, unlike a basic neural network, not all nodes between layers are connected in order to match that process in a visual cortex. Next, a CNN is organised in 3-d, rather than 2-d, with a width, height and depth parameter. Finally, the output layer is a single vector of probabilities. Fig. 14 illustrates these differences clearly.

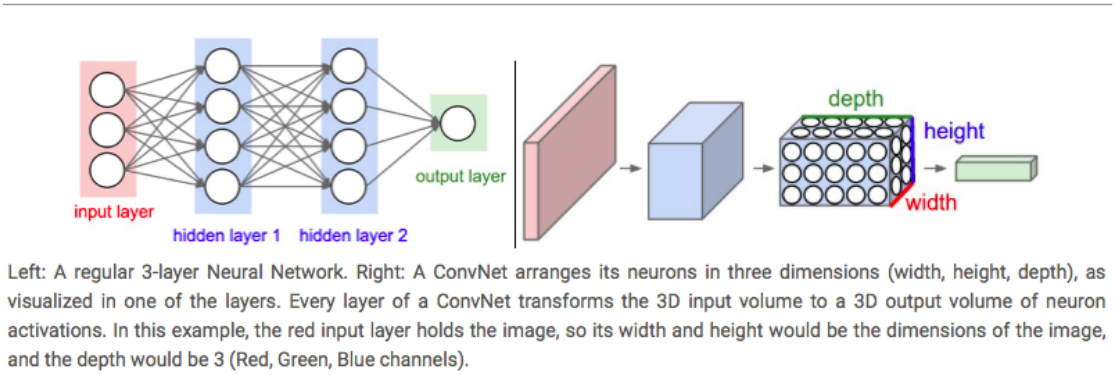

Figure 14: Comparison between ANN and CNN

As the name would suggest, convolution is the principal building block of a CNN. Convolution refers to the mathematical combination of two functions to produce a third. Or in other words, combining two sets of data. For a CNN convolution is performed on the data in the form of a filter or kernel that creates a feature map. Convolution occurs by moving this filter over the data, with matrix multiplication occurring at each point, and the result summed for the feature map.

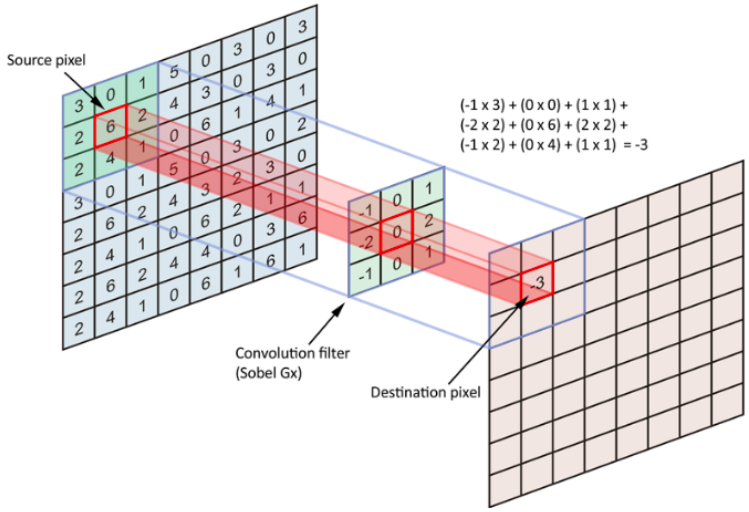

Figure 15: Example of Convolution

Fig. 15 illustrates how this works, with the depth coming from the RGB channels of the image. Different convolutions are performed on the input, each with different filters, and the results of the different feature maps are combined at the last step to produce the output of the convolution layer.

Just like any other neural network, the output of convolution layers must be passed through an activation function in order to make the output nonlinear. The most commonly used activation functions are the ReLu function or rectified linear function, given by the equation:

$$f(x) = \max(0, x) \tag{1}$$

or the Softmax function which squashes the output vector into a categorical probability distribution in the following manner:

$$\sigma(z)_j = \frac{e^{Z_j}}{\sum_{k=1}^{K} e^{Z_k}} \tag{2}$$

where $j = 1, 2, ..., K$.

Convolution layers are often proceeded by pooling layers. Pooling is utilised to reduce the dimensionality in order to lower the number of parameters and computation time of the network. This will shorten training time and help with overfitting which is when the network loses it generalisability by fitting the data it is trained on too well. Max pooling is the most commonly used type of pooling

and takes the max value in each window. This allows the network to keep the significant information but reduce the feature map size. Fig. 16 illustrates an example of max pooling using a 2x2 filter and a stride length of 2 (stride refers to how much we move along the data before performing the process again).

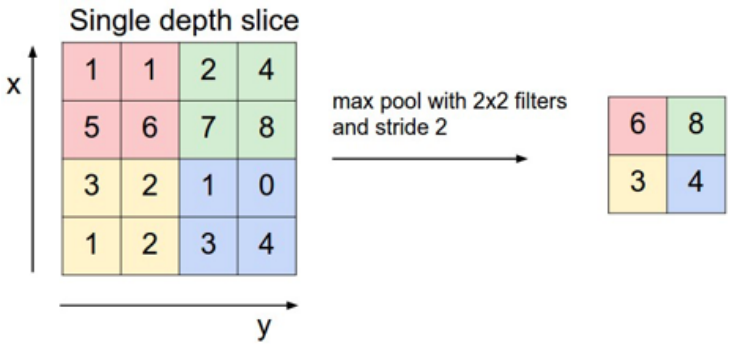

Figure 16: Example of Max Pooling

The final few layers of a CNN are in principle similar to a basic neural network for the purposes of classification. However, these fully connected layers only accepted 1-d data and therefore the 3-d data must be flattened in order to classify.

Training is also similar to that of a basic neural network, utilising backpropagation/ gradient descent in order to find the correct weights at each step to make accurate predictions. This step is obviously a bit more complex mathematically considering the convolution operations.

‘

4.1.3 R-CNN and Mask R-CNN

Convolution neural networks encounter issues when utilised as a detection algorithm, that is when attempting to draw a bounding box around an object of interest. This problem gets complicated when attempting to draw more than one bounding box. Say, for example, a CNN is set up to detect cats and draw a bounding box around them. If images of cats are supplied to the CNN, some images may have more than one cat in them and this is where problems start to arise since the CNN can’t know there are multiple objects to detect beforehand. It is impossible in this case to build a standard CNN followed by a fully connected layer as the length of the output layer is variable (detecting different numbers of objects). It is possible to solve this problem by taking different regions of the image and use a CNN to classify within that region. However, the objects of interest may have different spatial locations and aspect ratios and therefore the number of regions that need to be searched would be huge and computationally this would ’blow up’. This is where R-CNN comes in. R-CNNs take a specified number of regions (≈ 2000), called region proposals, and just work with

them [10]. These region proposals are generated using a selective search algorithm [10], which works by initially generating many candidate regions, using a greedy algorithm to combine similar regions, and then uses the generated regions to create region proposals. After selecting the region proposals, they are warped into squares that are fed into a CNN. The CNN is essentially a feature extractor with an output of a 4096-d dense layer feature vector that is fed into a Support Vector Machine in order to classify the presence of the object of interest within that particular region [10]. Obviously, this method has its downfalls, still taking a long time to train considering it must train each image on 2000 regions, and also the selective search algorithm is fixed and therefore region proposals may be poor.

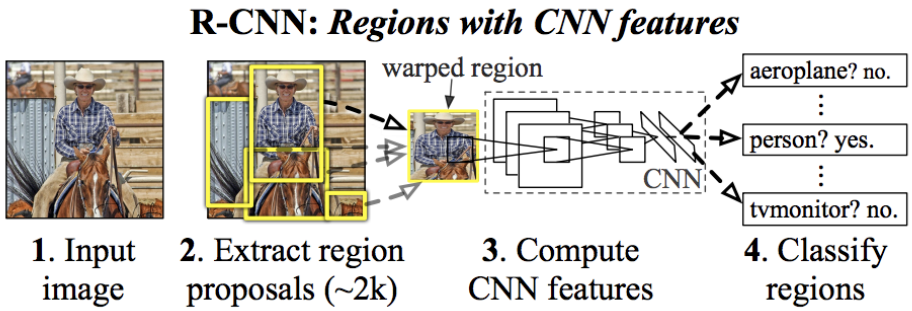

Figure 17: RCNN

Some of the drawbacks were eliminated in future work by the same authors with an algorithm called Fast-RCNN [11]. However, the true improvements came from Faster-RCNN, which is the basis for the model used in this project. For Faster-RCNN the image is provided as an input for a convolutional network that produces a feature map. A separate network is then used to predict region proposals [12]. The region proposals are reshaped and then used to classify the image within the proposed region.

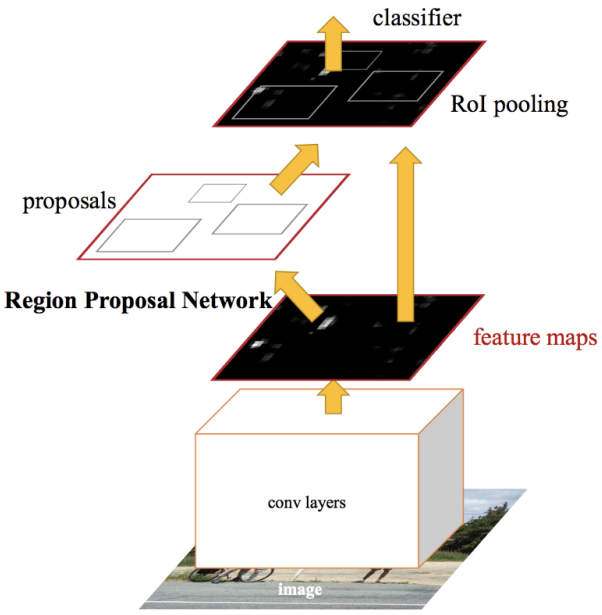

Figure 18: Faster-RCNN

Mask R-CNN is very similar to Faster-RCNN, possessing the same framework with an additional process that creates a more generalisable network [13]. This extra branch for predicting an object mask along with the original branch for bounding box recognition [13]. This essentially means that along with the bounding box, Mask R-CNN attempts to cut out the exact region the object occupies. For lung opacities with no clearly defined boundary, it clearly makes sense to use Mask R-CNN.

However, creating a brand new neural network that follows the Mask R-CNN framework would be an arduous task, which would probably lead to relatively poor predictions due to the training set size (approx. 25000). While this is quite large, given the complexity of detecting pneumonia, it would be hard to create a model from scratch that is acceptable. That is why transfer learning is used. This is essentially re-training a previously trained model. In their simplest form, CNNs are edge or pattern detectors, especially at the beginning of the hidden layers. Therefore utilising a neural network that has class-leading accuracy on the COCO dataset would clearly improve the prediction and classification of pneumonia [13]. Therefore a pre-trained Mask R-CNN model (Matterplot Mask R-CNN Implementation) was re-trained on the RSNA Pneumonia dataset in order to predict pneumonia.

5 Results

A Mask R-CNN on a ResNet50 backbone, pretrained on the COCO dataset was trained on the RSNA dataset. The training utilised a Colabatory Google Python notebook, enhanced by a GPU, with the

training taking place on one GPU, with 8 images per GPU and 200 steps per epoch (stage) for a total of 16 stages ($8 \times 200 \times 16 = 26,000 > 25,600 =$ dataset size). So as to prevent overfitting, the learning rate was adjusted at different stages of the training. The learning rates for each epoch can be seen in Table 4.

Epochs	Learning Rate
1-2	0.012
3-6	0.006
7-16	0.0012

Table 4: Learning Rates for Each Epoch

Total training time for the Mask R-CNN was approximately 6.5 hours. For each epoch, a number of different loss functions, as per the Mask R-CNN model used, were calculated. These different metrics and their respective values for each epoch can be seen in Fig. 19.

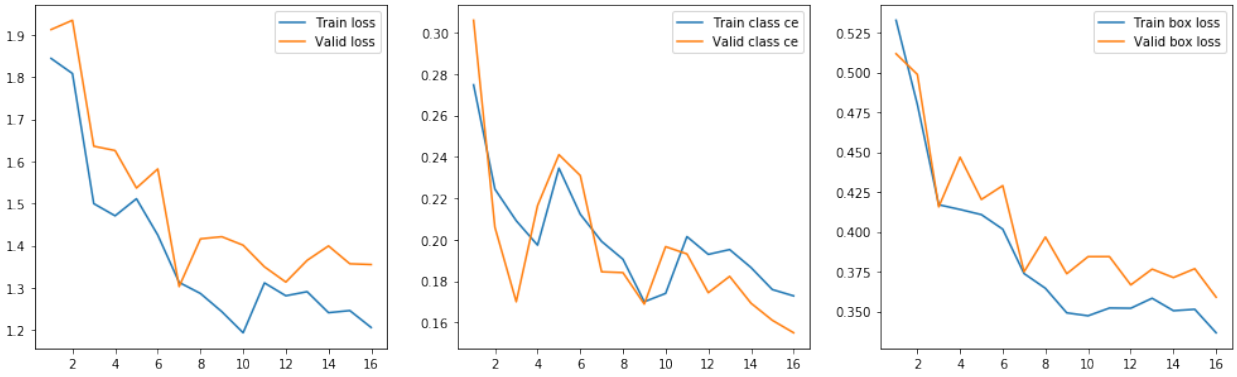

Figure 19: Losses

With the training set being split into training and validation set, the goal was to minimise the overall validation loss, which occured during epoch 7. The weights from epoch 7 were therefore utilised for prediction on the test set of images. This test set contained 1,000 images in total. Due to the nature of the data set and how the information regarding test set bounding boxes was only released in stage 2, comparisons between predictions and ground truths come from the validation set.

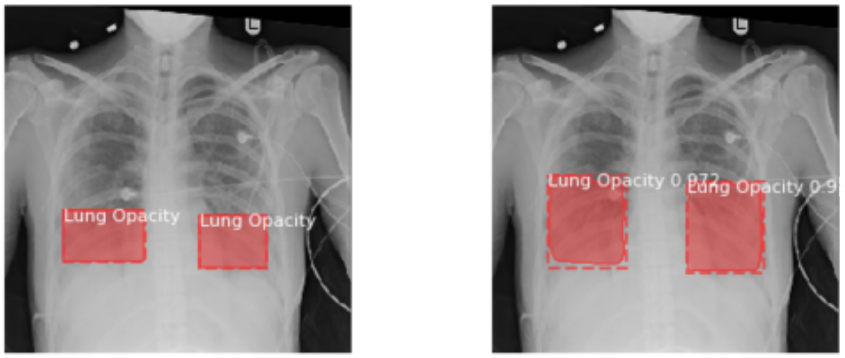

Figure 20: Accurate Prediction

Fig. 20 illustrates an example of a good prediction within the validation set (Left is ground truth, right is prediction for all predictions). The Mask R-CNN has managed to correctly detect that lung opacities were present and where these lung opacities are with high certainty. The bounding boxes from the Mask R-CNN are considerably larger in area than the ground truth, but this is potentially to be expected considering how lung opacities do not have well-defined boundaries. This trend of large bounding boxes can be seen across other predictions. This is exemplified by Fig. 21.

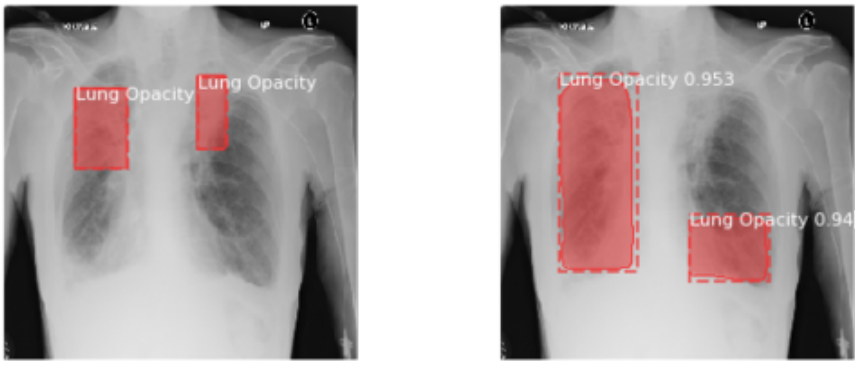

Figure 21: Semi-Accurate Prediction

The Mask R-CNN has managed to correctly predict that lung opacities exist. However, one bounding box is considerably larger than the ground truth and the other is in the incorrect location. Without any radiology knowledge, it is difficult to speak in absolution about CXRs and lung opacities, but there seems to be a trend across the dataset that the Mask R-CNN is good at detecting 'clouds' or other shapes, but is not good at distinguishing whether this is a pneumonia-related lung opacity, another sort of opacity, or simply an artefact. Take for example Fig. 22 and Fig. 23.

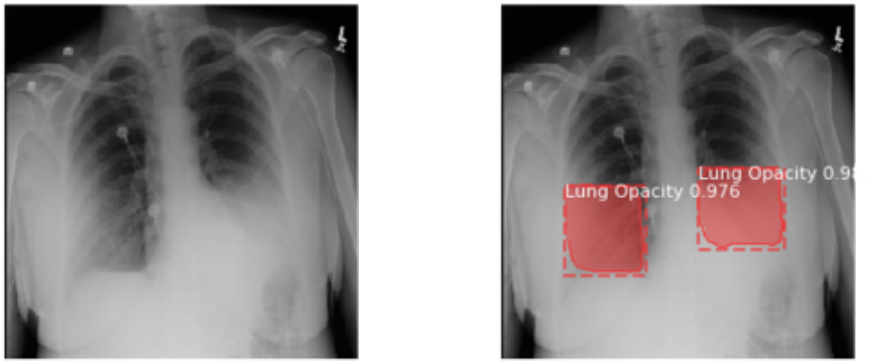

Figure 22: Bad Prediction

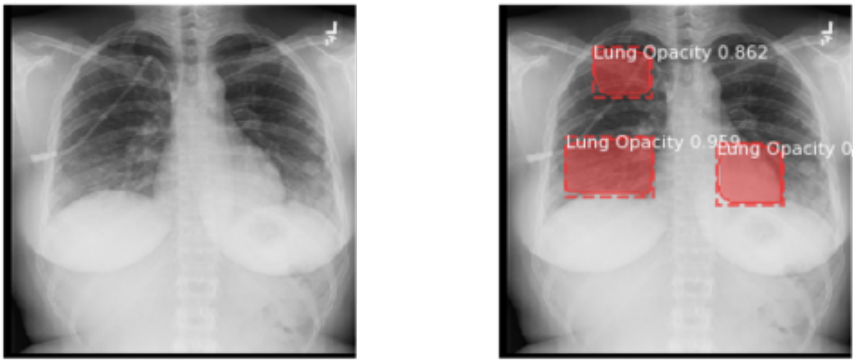

Figure 23: Bad Prediction

The Mask R-CNN identifies cloudier areas as lung opacities relating to pneumonia. In Fig. 23 even the heart is identified as a lung opacity. It is evident the Mask R-CNN is too quick to identify whiter areas as lung opactities. This is ultimately seen in the confusion matrix for the test set.

		prediction outcome		
		0	1	**total**
actual value	0 (No Lung Opacity)	357 35.7%	290 29%	647
	1 (Lung Opacity)	41 4.1%	312 31.2%	353
	total	385	615	

In total, the Mask R-CNN had an accuracy of 66.9% in the classification sense. Obviously, in reality, it may not be sufficient to just say whether a person has or hasn't got pneumonia, and may be necessary to know where it is or how pervasive it is. With the bounding box predictions, it would be possible to give some measure of how accurate the model is in terms of error in bounding boxes but that is not done in this project. The most notable property of the confusion matrix is the false positive value of 44.8%. Overall the misclassification rate is 33.1%, or 331 in 1000, but 290 of these misclassified cases are false positive cases. The Mask R-CNN is obviously too sensitive when it comes to detecting lung opacities and considers too many areas of whiteness as a lung opacity.

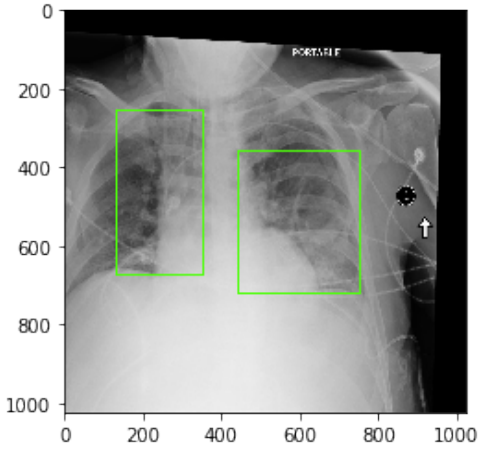

Figure 24: Sample Prediction 1

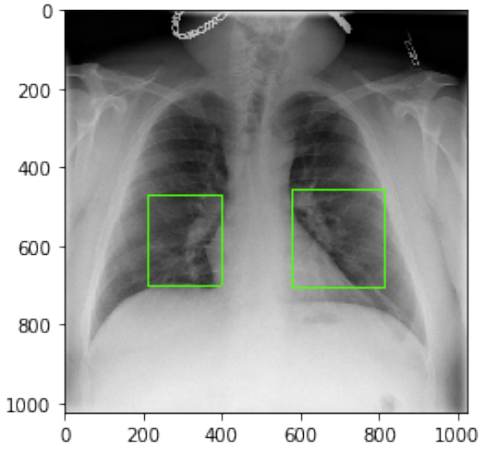

Figure 25: Sample Prediction 2

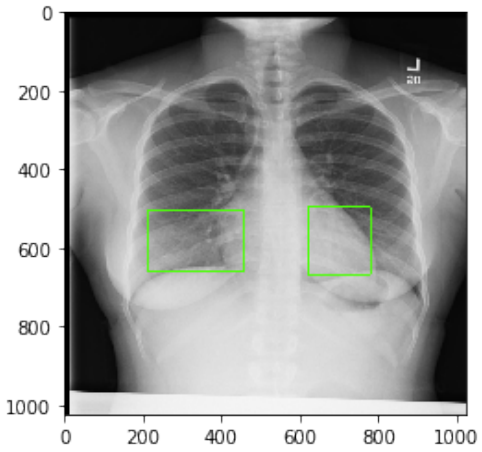

Figure 26: Sample Prediction 3

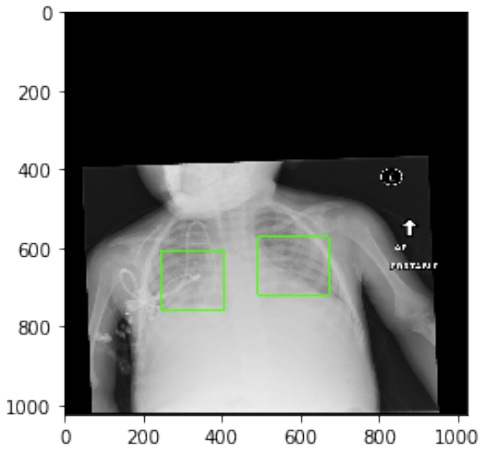

Figure 27: Sample Prediction 4

Fig. 24-Fig. 27 illustrates examples of predictions on the test set and seems to highlight the idea that the Mask R-CNN is too sensitive to relatively white areas. In particular, in all of them, the heart (right bounding box) is considered a lung opacity which obviously is incorrect. While some of these bounding boxes may be correct, the Mask R-CNN is clearly too sensitive and this obviously leads onto improvements and limitations of the Mask R-CNN used and the project itself.

6 Discussion

As a first attempt at such a technical problem such as detecting lung opacities, which radiologists study for years to perfect, an accuracy of 66.9% is acceptable, but can obviously be massively improved upon. The first thing that would improve the accuracy is using the stage 2 data available on Kaggle. Increasing the number of images trained on to 26,684 would obviously improve the detection slightly.

On top of this, it was noticed during exploratory data analysis that a large number of scans did not fill the entire image. And while augmentations are regularly used for neural networks to prevent overfitting, the extent to which these scans deviate from the norm is too much. These images had poor technical quality; either the X-ray didn't cover the full image or the X-ray is skewed within the image. The most appropriate way to detect this was deemed to be a black pixel count. The percentage of black pixels was calculated for each image in the training set and the boxplot shown in Fig. 28 shows quite a number of outliers.

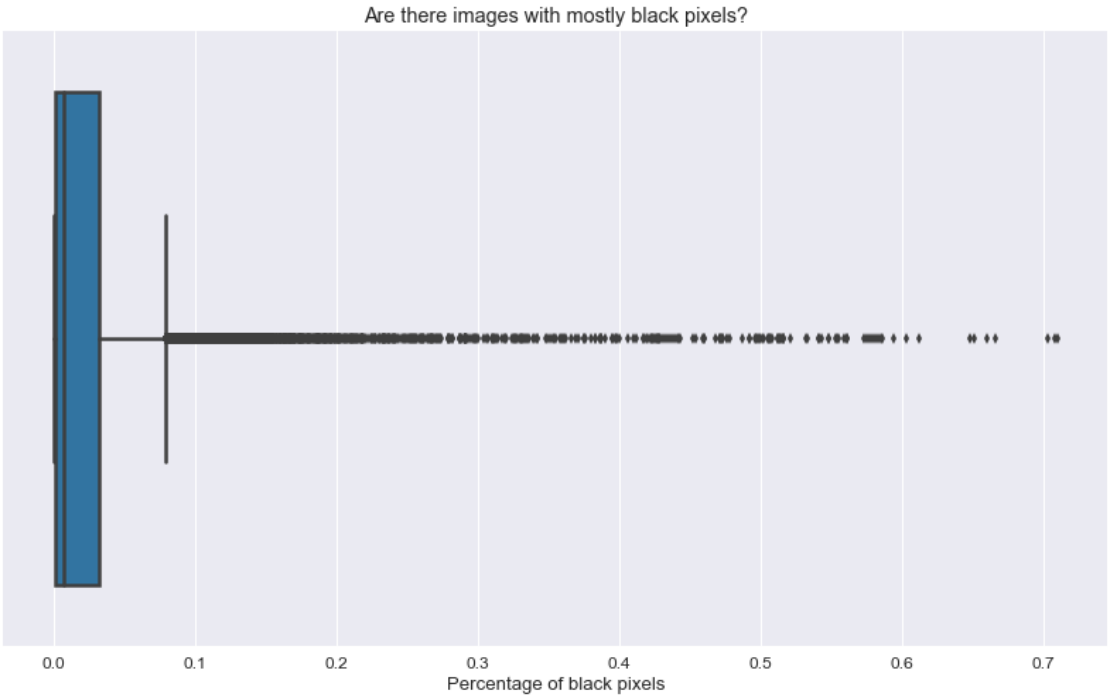

Figure 28: Boxplot of Percentage of Black Pixels

Obviously, for the most part, these outliers are normal chest radiographs. However, a lot will be highly skewed scans that may be affecting accuracy. To help this a cut-off could be somewhat arbitrarily imposed to differentiate normal chest radiographs with good technical quality and images that needed further examination. For example, taking the cut-off as 55% black pixels, scans above this threshold can be manipulated and transformed to fit the full image. An example can be seen in Fig. 29 and it is clear the technical quality of these images isn't good.

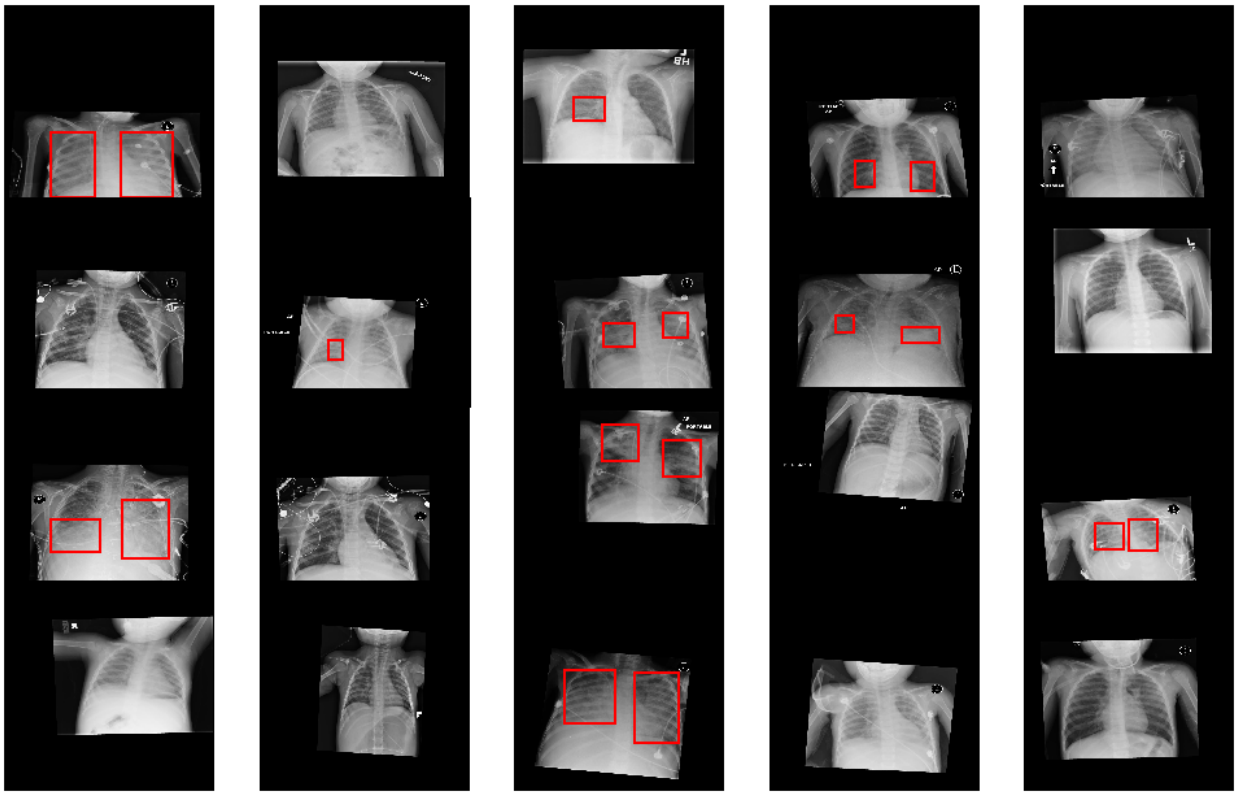

Figure 29: Sample of Outliers

A bounding box was fit around these images to isolated the scan from the black pixels (see Fig. 30) and these isolated scans were transformed to fit the full image (see Fig. 31).

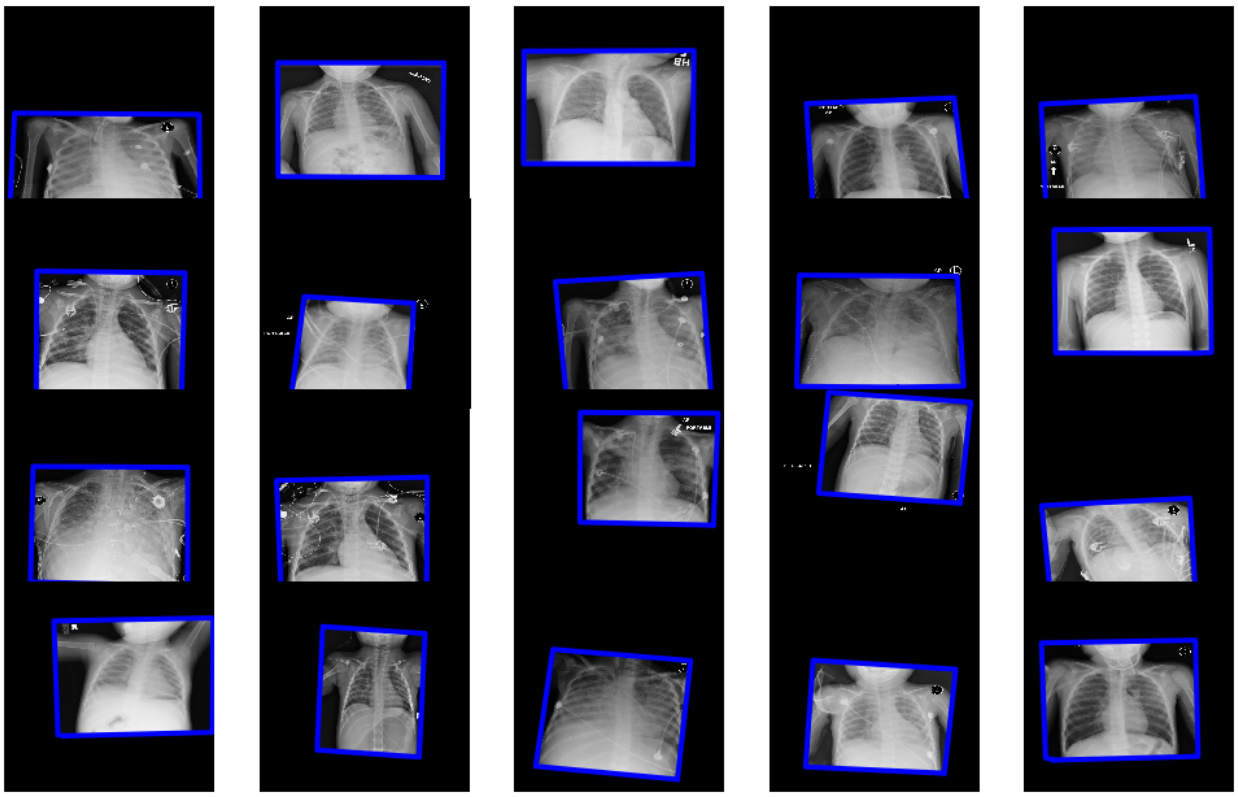

Figure 30: Sample of Outliers and the Isolated Image

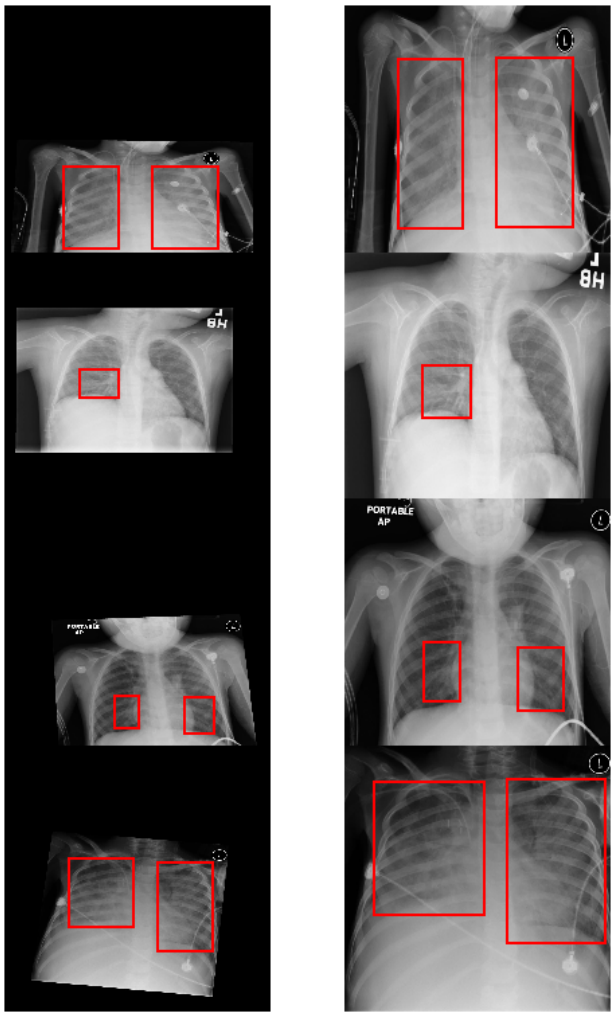

Figure 31: Transformed Outliers

It is possible that if this transformed data set was used for training with augmentations, the accuracy would increase due to the uniformity of the data.

Obviously, one major method of improving accuracy is to improve the model itself. This can come in two forms; add extra steps to the Mask R-CNN already used, or use something other than a Mask R-CNN. In both cases, the easiest way to improve the accuracy would be to reduce the false positive rate. Success has been found in using the Mask R-CNN for detection and then using another neural network such as a Densely Connected Convolutional Network for classification. While traditional convolutional networks with L layers have L connections, one between each layer and its subsequent layer, a Dense Convolutional Network (DenseNet) has $\frac{L(L+1)}{2}$ direct connections, as illustrated in Fig. 32.

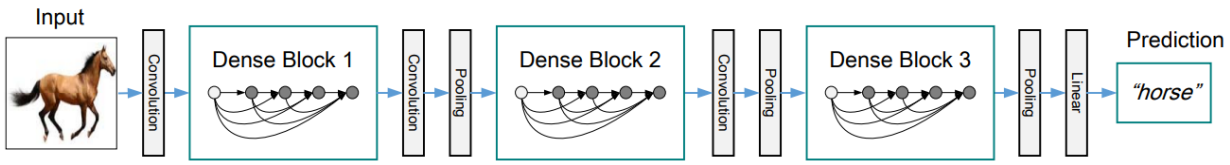

Figure 32: DenseNet Method

For each layer, the feature-maps of every preceding layer is utilised as inputs [14]. DenseNets offer a number of advantages such as alleviating the vanishing gradient problem, strengthening feature propagation, encouraging feature re-use, and substantially reducing the number of parameters [14]. A number of different DenseNet architectures are used commonly and are illustrated in Fig. 33

Layers	Output Size	DenseNet-121	DenseNet-169	DenseNet-201	DenseNet-264
Convolution	112 × 112	7 × 7 conv, stride 2			
Pooling	56 × 56	3 × 3 max pool, stride 2			
Dense Block (1)	56 × 56	$\begin{bmatrix} 1\times1 \text{ conv} \\ 3\times3 \text{ conv} \end{bmatrix} \times 6$	$\begin{bmatrix} 1\times1 \text{ conv} \\ 3\times3 \text{ conv} \end{bmatrix} \times 6$	$\begin{bmatrix} 1\times1 \text{ conv} \\ 3\times3 \text{ conv} \end{bmatrix} \times 6$	$\begin{bmatrix} 1\times1 \text{ conv} \\ 3\times3 \text{ conv} \end{bmatrix} \times 6$
Transition Layer (1)	56 × 56	1 × 1 conv			
	28 × 28	2 × 2 average pool, stride 2			
Dense Block (2)	28 × 28	$\begin{bmatrix} 1\times1 \text{ conv} \\ 3\times3 \text{ conv} \end{bmatrix} \times 12$	$\begin{bmatrix} 1\times1 \text{ conv} \\ 3\times3 \text{ conv} \end{bmatrix} \times 12$	$\begin{bmatrix} 1\times1 \text{ conv} \\ 3\times3 \text{ conv} \end{bmatrix} \times 12$	$\begin{bmatrix} 1\times1 \text{ conv} \\ 3\times3 \text{ conv} \end{bmatrix} \times 12$
Transition Layer (2)	28 × 28	1 × 1 conv			
	14 × 14	2 × 2 average pool, stride 2			
Dense Block (3)	14 × 14	$\begin{bmatrix} 1\times1 \text{ conv} \\ 3\times3 \text{ conv} \end{bmatrix} \times 24$	$\begin{bmatrix} 1\times1 \text{ conv} \\ 3\times3 \text{ conv} \end{bmatrix} \times 32$	$\begin{bmatrix} 1\times1 \text{ conv} \\ 3\times3 \text{ conv} \end{bmatrix} \times 48$	$\begin{bmatrix} 1\times1 \text{ conv} \\ 3\times3 \text{ conv} \end{bmatrix} \times 64$
Transition Layer (3)	14 × 14	1 × 1 conv			
	7 × 7	2 × 2 average pool, stride 2			
Dense Block (4)	7 × 7	$\begin{bmatrix} 1\times1 \text{ conv} \\ 3\times3 \text{ conv} \end{bmatrix} \times 16$	$\begin{bmatrix} 1\times1 \text{ conv} \\ 3\times3 \text{ conv} \end{bmatrix} \times 32$	$\begin{bmatrix} 1\times1 \text{ conv} \\ 3\times3 \text{ conv} \end{bmatrix} \times 32$	$\begin{bmatrix} 1\times1 \text{ conv} \\ 3\times3 \text{ conv} \end{bmatrix} \times 48$
Classification Layer	1 × 1	7 × 7 global average pool			
		1000D fully-connected, softmax			

Figure 33: DenseNet Architectures

DenseNets have achieved significant improvements over other state-of-the-art neural networks on a number of highly competitive object recognition benchmark tasks [14] with two-phase approaches (CNN then DenseNet) having proven success of minimising false positives in other areas [15]. In this example, 'hard negative backgrounds' were classified as pedestrians due to their human-look-alike pattern. With the Mask R-CNN trained on the RSNA data incorrectly classifying the heart as a lung opacity due to an almost cloud-like depiction on the scans, using a two-phase classification approach with a DenseNet architecture such as DenseNet-169 could greatly improve overall accuracy.

On top of an enhanced two-phase classification detection approach, many other neural network architectures could be utilised for this problem that may perform better. Facebook AI Research's RetinaNet is an example of one of the best one-stage object detectors, surpassing top-performing, two-stage methods [16]. It identifies extreme foreground-background class imbalance as the primary obstacle for one-stage detectors and proposes *focal loss* to remedy this [16]. Focal loss modifies the cross-entropy loss function to down-weight easy examples which overwhelm cross entropy - making up the majority of the loss and dominate the gradient - and focus training on hard negatives. This one stage method could also offer potentially improved performance, and with faster training it could be trained on more images in order to further improve the accuracy. This is only one of the many other options to choose from but would definitely be worth considering.

7 Conclusion

In conclusion, it is clear that the use of neural networks within the medical industry is going to greatly increase in the future both due to their accuracy and their ability to assist radiographers and help them be less error-prone. While the overall accuracy of the Mask R-CNN wasn't overly impressive, it does offer an insight into the power neural networks, and other Machine Learning algorithms, can have. As an undergraduate with no prior experience to Deep Learning, I was capable of training a model that is in some way capable of reading CXRs. Many papers written by many people have been referenced within this report and these people are some of the brightest minds, working on the most cutting-edge technology. If an undergraduate with no prior experience can achieve this, it is very clear these researchers can achieve much more. This project illustrates what a computer is capable of, but these researchers are truly on the precipe of a great revolution. Diseases will be more easily detected, correct care plans will be followed more regularly, and fewer errors will be made. It not only the first world that will benefit but deep learning will also allow the standard of healthcare to greatly increase in many third world companies to increase. It is remarkable to think that computers will soon be able to do these tasks with absolute confidence, and despite some fears, it is definitely part of the technological revolution that should be embraced.

Code

Code for the project is available on GitHub at the link: `https://github.com/cwjjb96/ST4090`

References

[1] J. De Fauw, J. R. Ledsam, B. Romera-Paredes, S. Nikolov, N. Tomasev, S. Blackwell, H. Askham, X. Glorot, B. O'Donoghue, D. Visentin, et al., Clinically applicable deep learning for diagnosis and referral in retinal disease, Nature medicine 24 (9) (2018) 1342.

[2] M. D. Abràmoff, P. T. Lavin, M. Birch, N. Shah, J. C. Folk, Pivotal trial of an autonomous ai-based diagnostic system for detection of diabetic retinopathy in primary care offices, Npj Digital Medicine 1 (1) (2018) 39.

[3] P. Rajpurkar, J. Irvin, R. L. Ball, K. Zhu, B. Yang, H. Mehta, T. Duan, D. Ding, A. Bagul, C. P. Langlotz, et al., Deep learning for chest radiograph diagnosis: A retrospective comparison of the chexnext algorithm to practicing radiologists, PLoS medicine 15 (11) (2018) e1002686.

[4] L. R. Goodman, Felson's principles of chest roentgenology, a programmed text, Elsevier Health Sciences, 2014.

[5] M. I. Neuman, E. Y. Lee, S. Bixby, S. Diperna, J. Hellinger, R. Markowitz, S. Servaes, M. C. Monuteaux, S. S. Shah, Variability in the interpretation of chest radiographs for the diagnosis of pneumonia in children, Journal of hospital medicine 7 (4) (2012) 294–298.

[6] M. van Gerven, S. Bohte, Artificial neural networks as models of neural information processing, Frontiers Media SA, 2018.

[7] Build with ai.
URL https://deepai.org/machine-learning-glossary-and-terms/neural-network

[8] K. Fukushima, Neocognitron: A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position, Biological cybernetics 36 (4) (1980) 193–202.

[9] M. Matsugu, K. Mori, Y. Mitari, Y. Kaneda, Subject independent facial expression recognition with robust face detection using a convolutional neural network, Neural Networks 16 (5-6) (2003) 555–559.

[10] R. Girshick, J. Donahue, T. Darrell, J. Malik, Rich feature hierarchies for accurate object detection and semantic segmentation, in: Proceedings of the IEEE conference on computer vision and pattern recognition, 2014, pp. 580–587.

[11] R. Girshick, Fast r-cnn, in: Proceedings of the IEEE international conference on computer vision, 2015, pp. 1440–1448.

[12] S. Ren, K. He, R. Girshick, J. Sun, Faster r-cnn: Towards real-time object detection with region proposal networks, in: Advances in neural information processing systems, 2015, pp. 91–99.

[13] K. He, G. Gkioxari, P. Dollár, R. Girshick, Mask r-cnn, in: Proceedings of the IEEE international conference on computer vision, 2017, pp. 2961–2969.

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

[15] S. K. Choudhury, R. P. Padhy, P. K. Sa, Faster r-cnn with densenet for scale aware pedestrian detection vis-à-vis hard negative suppression, in: 2017 IEEE 27th International Workshop on Machine Learning for Signal Processing (MLSP), IEEE, 2017, pp. 1–6.

[16] T.-Y. Lin, P. Goyal, R. Girshick, K. He, P. Dollár, Focal loss for dense object detection, in: Proceedings of the IEEE international conference on computer vision, 2017, pp. 2980–2988.