An Evaluation of Deep Learning Techniques for Chest X-Ray Abnormality Detection and Classification

Emmanuel Akama (S2227958)
(MSc.) Big-Data Technologies

Introduction

Among the various medical imaging modalities, Chest X-Ray (CXR) imaging holds a critical place due to its widespread use in diagnosing a multitude of thoracic diseases such as pneumonia, tuberculosis, and lung cancer.

Traditionally, the interpretation of CXR has been the domain of skilled radiologists who meticulously analyse these images to identify and classify abnormalities. However, this process is time-consuming, subjective, and prone to human error. The integration of deep learning techniques offers a promising solution to these challenges by providing automated, accurate, and efficient image analysis tools.

Background

The advancement of deep learning for computer vision tasks has been marked by several key developments and breakthroughs, beginning with LeCun's 1998 pioneering work on MNIST handwriting recognition, which demonstrated the potential of Convolutional Neural Networks (CNNs) for image processing tasks.

This laid the foundation for the remarkable achievements seen in the 2012 ILSVRC ImageNet competition, where AlexNet significantly outperformed traditional methods, ushering in a new era of deep learning. Subsequent advancements, such as the inception modules introduced in GoogLeNet in 2015, further pushed the boundaries of what deep learning could achieve in computer vision and visual recognition tasks.

Problem Statement

Despite these developments, the potential of these models has not been fully realized for detecting and classifying medical imaging modalities due to the lack of sufficient CXR data.

With the recent availability of public CXR datasets such as NIH Chest X-Ray 8/14, CheXpert, and RSNA Pneumonia Detection dataset, the question arises: *Can we harness the capabilities of advanced deep learning models to achieve, and possibly exceed, human-level performance in detecting and classifying pneumonia characterized as lung opacities in digital CXR images?*

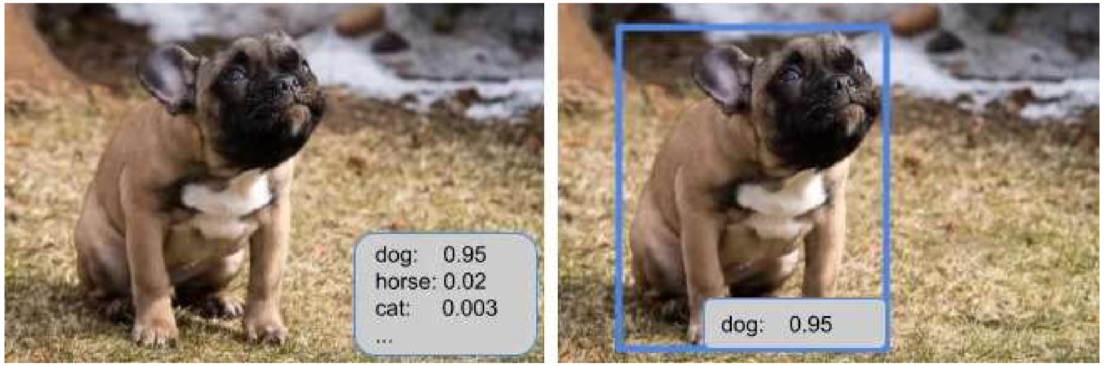
Aim & Objectives

According to the World Health Organization (WHO), pneumonia is the single largest infectious cause of death in children worldwide, accounting for approximately 15% of all deaths of children under 5 years. This burden is exacerbated in low- and middle-income countries due to factors such as limited access to healthcare, malnutrition, and a higher prevalence of underlying health.

The primary aim of this research is to evaluate the application of deep learning techniques for abnormality detection and classification of CXR images with the digital CXR image and annotations presented for the 2018 RSNA Pneumonia Detection Challenge. To support this objective, we conducted training experiments using different configurations on a Mask-RCNN model with a ResNet50/101 backbone by applying optimization techniques to enhance model performance using quantitative and qualitative assessment metrics.

Methodology

Object detection with deep neural networks is a two-fold task: (1) *Object Recognition* identifies "what" object classes are contained in the image. This is achieved using an algorithm that outputs the probability distribution across a set of pre-defined class labels in the image. (2) *Object Localization* determine "where" the object classes are located. This typically involves predicting the position of a bounding box that determines the location of the object class.



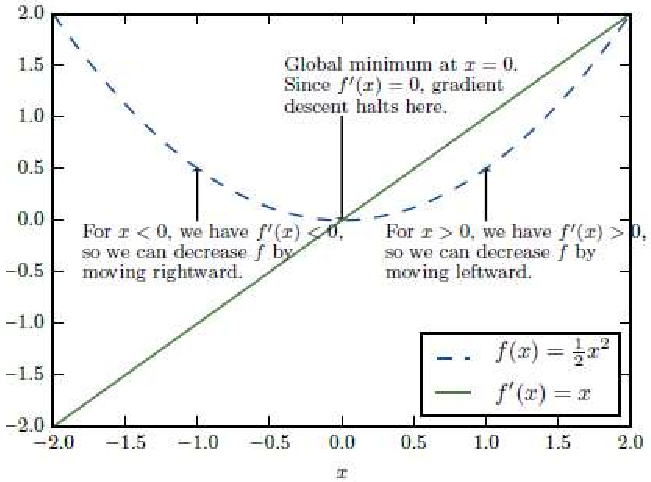
Approaches to Deep Neural Network-based Object Detection & Classification

There are two major approaches to deep neural network-based object detection. This includes the single-shot and region-based detector approaches.

Single-shot detectors like YOLO perform object detection as well as classification in a single forward pass of the network. This enhances their speed and accuracy, making them more suitable for real-time applications. Region proposal-based detectors, on the other hand, use a multi-stage approach to extract features from each region proposal for classification and simultaneously train a regression model to predict the correct bounding boxes from learned offsets. Although, this process is computationally expensive, the Mask R-CNN is specially designed to minimize these effects.

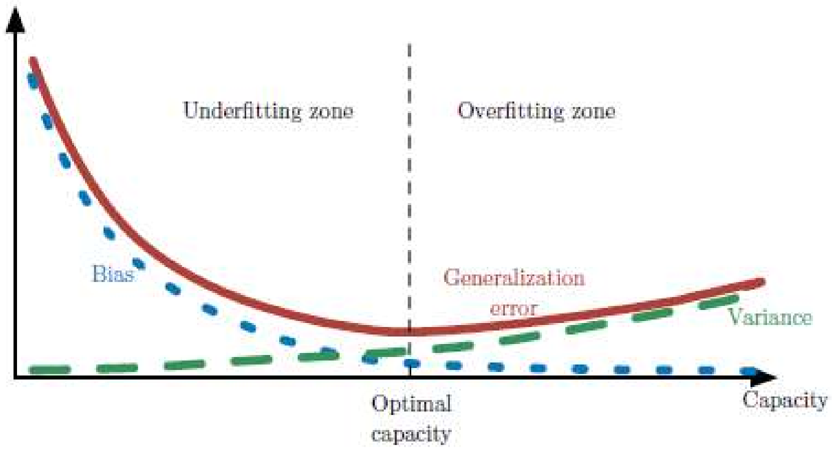
Model Training & Optimization Techniques

The performance of a deep neural network hinges significantly on the training and optimization techniques employed to optimize model parameters (weights and biases) using a cost function $f(L,\theta)$ to minimize the loss L, with a set of model parameters θ. The loss L indicate the difference between the predicted outputs of a model $\hat{x}_i$ and the true outputs x_i.



This normally involve the use of regularization techniques that reduces the generalization error by adding penalty terms using techniques like dropout, early stopping, and data augmentation that aid in capturing essential patterns in the data while avoiding overfitting.

A model that generalizes well will have similar performance on both the training data and unseen test data. The most common way to evaluate generalization is through an appraisal of model's performance metrics (such as, accuracy, precision, recall, F1-score) on validation and test sets.



Interpreting Pneumonia from Chest Radiographs

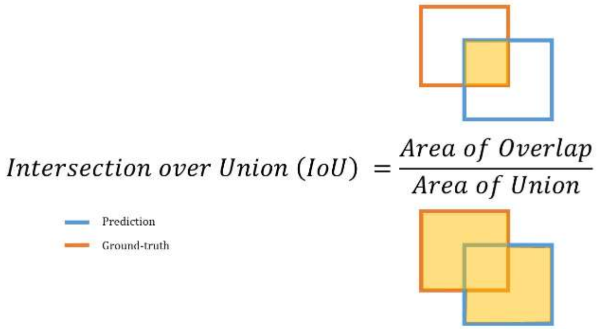
Traditional methods of detecting pneumonia typically begins with a visual examination of the patient's CXR radiograph by a trained radiologist or medical professional to assess the lungs for areas of increased opacity, which may indicate fluid, pus, or other materials in the alveoli that is characteristic of a patient with pneumonia.

With recent advances in artificial intelligence and deep learning, automated systems can be developed to assist in interpreting digital CXR radiographs. This can improve diagnostic accuracy by providing a second opinion, highlight areas of concern.

Model Evaluation Method

Intersection over Union (IoU) is a common metric used to evaluate the accuracy of object detection models. The IoU is a measure of the overlap between the areas of the predicted and ground truth bounding boxes.

The IoU values typically range from 0 (in cases where there's no overlap between the predicted and ground truth boxes) to 1 (indicating a perfect overlap between the predicted and ground truth boxes). A higher IoU indicates a better match between the predicted and actual bounding boxes, with a threshold (e.g., 0.5) often used to determine whether a prediction is considered a true positive.



Model Evaluation Metrics

Precision: Measures the proportion of true positive detections among all positive detections

Recall: Measures the proportion of true positive detections among all ground truth instances.

Both metrics depend on the number of true positive (TP), false positive (FP) and false negative (FN) predictions.

- ***True Positives (TP)***: This is an indication of the correctly detected objects where the predicted bounding box overlaps sufficiently with the ground truth bounding box. This is defined by the IoU threshold.
- ***False Positives (FP)***: This is an indication of the predicted bounding boxes that do not overlap sufficiently with the ground truth bounding box.
- ***False Negatives (FN)***: This is an indication of the ground truth objects that were not detected by the model.

Model Results

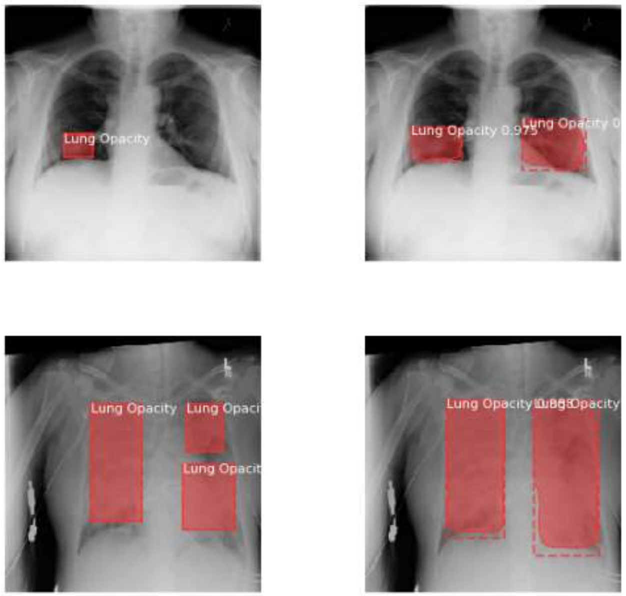
The comparative performance of the pre-trained (head with MS-COCO weights) and complete training schedule of our Mask R-CNN model on the validation dataset containing 1,500 samples is shown in the table below.

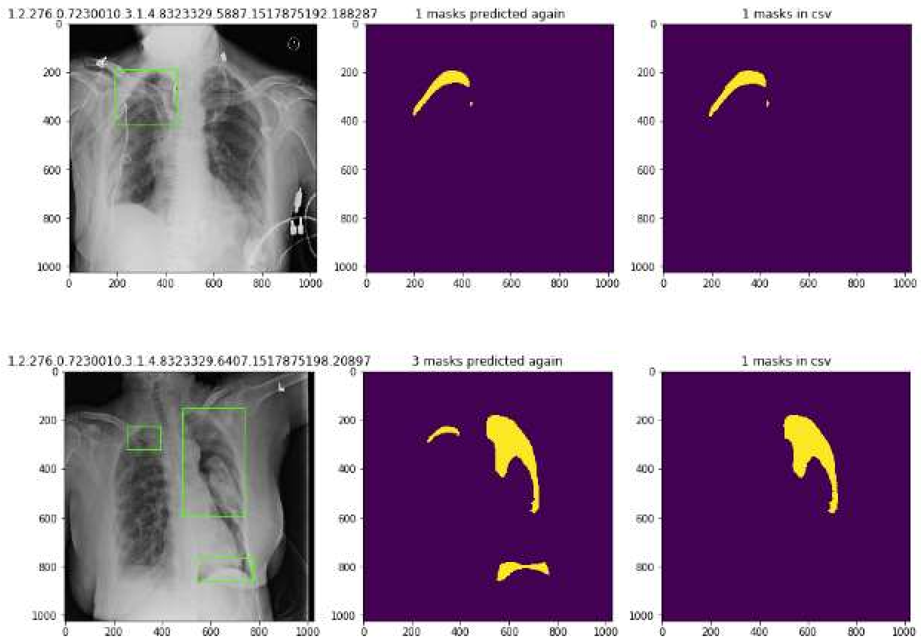
Network Backbone	Trained Layers	Is the Data Augmented?	IoU	mAP	mAR	F1
ResNet50	Head	No	***0.25***	***0.9308***	***0.9530***	***0.9418***
			0.50	***0.6058***	***0.6655***	***0.6342***
			0.75	***0.1285***	***0.1825***	***0.1508***
	Complete	Yes	0.25	0.8346	0.8402	0.8374
			0.50	0.4988	0.4794	0.4889
			0.75	0.0750	0.1032	0.0869
ResNet101	Head	No	0.25	0.8777	0.9062	0.8917
			0.50	0.5769	0.5868	0.5818
			0.75	0.0982	0.1557	0.1204
	Complete	Yes	0.25	0.8521	0.8762	0.8640
			0.50	0.5769	0.5868	0.5818
			0.75	0.0897	0.1482	0.1118

The comparative performance analysis showed varying mAP, mAR, and F1 scores across different IoU thresholds on the ResNet 50 and 101 variants, with results from ResNet101 lagging behind across all thresholds. As expected, the models trained with pre-trained MS-COCO weights out-performed the ones with complete training on all layers (even with data augmentation). This is possible due to (1) number of training samples in the MS-COCO dataset, and (2) training is optimized for the network head rather than the complete network. See the highlighted portion of the report.

Although, both ResNet50 and ResNet101 variants are designed to mitigate the vanishing/exploding gradient problem through their residual connections, the increased depth of ResNet101 can still make it more susceptible to this issue, particularly in situations where the network parameters are not carefully tuned.

In the figures below, we present graphical illustrations of the model's performance with samples from the validation and training datasets, respectively. The images presented in the first column represent the ground truth of abnormality locations in the validation and testing datasets. The second and third columns represent the prediction of corresponding class and mask locations.





LIVE DEMO

Conclusion

Among the various medical imaging modalities, Chest X-Ray (CXR) imaging holds a critical place due to its widespread use in diagnosing a multitude of thoracic diseases such as pneumonia, tuberculosis, and lung cancer.

Traditionally, the interpretation of CXR has been the domain of skilled radiologists who meticulously analyse these images to identify and classify abnormalities. However, this process is time-consuming, subjective, and prone to human error. The integration of deep learning techniques offers a promising solution to these challenges by providing automated, accurate, and efficient image analysis tools.

Limitations & Future Work

Although this research provides valuable insights into the detection of pneumonia using deep learning techniques, the following limitations are acknowledged and recommended for future work.

1. The limited scope of the RSNA pneumonia detection dataset may not be adaptable to for real-world scenarios. More robust datasets like the NIH Chest X-Ray 8/14 or CheXNet can be used instead.

2. While the *Waleed Abdulla's* implementation of the Mask R-CNN model might be well-optimized for general object detection and classification tasks, it may not be fully tailored for the specific nuances present in digital CXR images. Hence, specific in-code adaptations might be required to enhance performance of the model in production environments

THANK YOU FOR YOUR TIME

Questions are welcome