Pneumonia Detection using Convolutional Neural Networks

Dhawal Chaturvedi, Dhivya Swaminathan, Hasika Mahtta

Abstract

In the United States, pneumonia accounts for over 500,000 visits to emergency departments [1] and over 50,000 deaths in 2015 [2], keeping the ailment on the list of top 10 causes of death in the country. While common, accurately diagnosing pneumonia is a tall order. It requires review of a chest radiograph (CXR) by highly trained specialists and confirmation through clinical history, vital signs and laboratory exams. We propose a machine learning algorithm that will automatically detect pneumonia and locate lung opacities on chest radiographs. 26684 chest radiographs were obtained from the Radiological Society of North America. A Convolutional Neural Network was modelled and we report an accuracy of 0.8006.

Keywords

pneumonia, radiographs, lung opacities, machine learning algorithms, convolutional neural network, epochs

Contents

Introduction

In the United States, pneumonia accounts for over 500,000 visits to emergency departments [1] and over 50,000 deaths in 2015 [2], keeping the ailment on the list of top 10 causes of death in the country. Detecting pneumonia requires review of Chest Radiograph (CXR) by trained professionals, supported by clinical history. A number of factors including patient positioning can alter a CXR, thus making it difficult to interpret. Fig.1 shows the deaths due to Pneumonia in the Unites States.

In this study, a machine learning methodology is proposed to detect pneumonia and identify bounding boxes for lung opacities using a convolutional neural network (CNN). CNN is a class of feed-forward artificial neural networks, commonly used in analyzing images. CNNs use an implementation of multilayer perceptrons that require very minimal preprocessing [3]. They are also known as shift invariant or space invariant artificial neural networks (SIANN), based on their shared-weights architecture and translation invariance characteristics [4] [5].

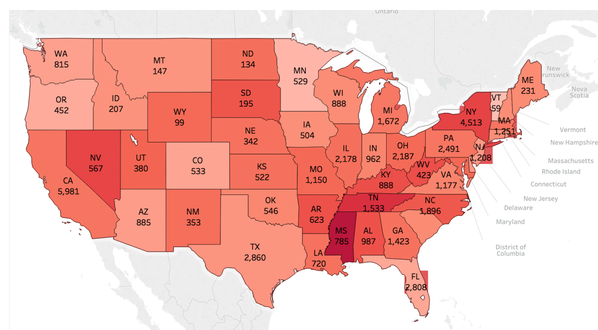

Figure 1. Pneumonia Statistics across the United States

1. Background

Diagnosing pneumonia requires review of a chest radiograph (CXR). Pneumonia manifests as an area or areas of increased opacity on CXR. CXRs are the most commonly performed diagnostic imaging study.This is basically to explore the potential for machine learning algorithms to automate initial detection (imaging screening) of potential pneumonia cases in order to prioritize and expedite their review. We hope that this technology can improve healthcare delivery and increase access to medical imaging expertise in parts of the world where access to skilled radiologists is limited. Rajpurkar et al. [6] focused on detecting pneumonia from chest X-rays at a level exceeding practicing radiologists. They found that CheXNet exceeds average radiologist performance on the F1 metric. They have extended CheXNet to detect all 14 diseases in Chest X-ray and achieve state of the art results on all 14 diseases.

Schalekamp et al. [7] have assessed the effect of bone suppression imaging on observer performance in detecting lung nodules in chest radiographs. Their work concluded that bone suppressed images improve radiologists' detection performance for pulmonary nodules, especially for those of moderate and subtle conspicuity. Zhou et al. [8] described a fully automated segmentation and recognition scheme, which is designed to recognize lung anatomical structures in the human chest by segmenting the different chest internal organ and tissue regions sequentially from high-resolution chest CT images. Sedai, Suman and Mahapatra [10] proposed a novel weekly supervised method to localize chest pathologies using class aware deep multiscale feature learning. CNN features learn to classify pathology responses from the intermediate feature maps along with the class specific layer relevance weights for coarser and deeper layers. The network switches between 2 modes during training and test phases; 1) Classification CNN (C-CNN) and 2) Attention CNN (A-CNN), respectively. In prediction phase, A-CNN combines the convolutional feature maps from individual layers using the learned layer relevance weights to obtain the multiscale attention map. The proposed multiscale attention map is robust against pathology size as it encapsulates the feature maps from both coarse and fine layers. Kalinovsky [11] presents the resuts of the first, exploratory stage of research and developments on segmentation of lungs in X-Ray chest images (Chest Radiographs) using Deep Learning methods and Encoder - Decoder Convolutional Neural Networks(ED-CNN). They concluded that the results obtained with this study may be considered as a promising tool for automatic lung segmentation in chest X-Ray images.

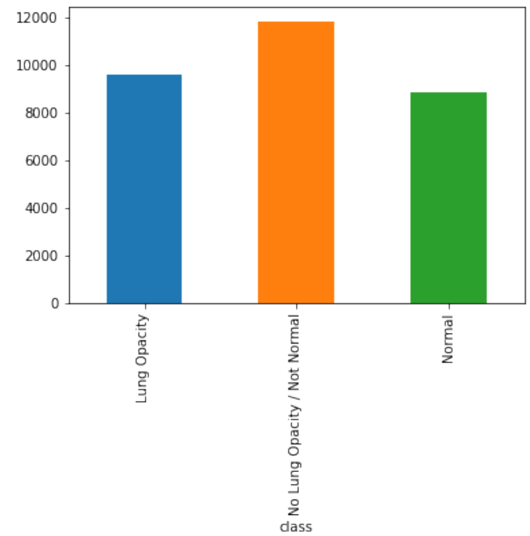

Figure 2. Dataset Skewness

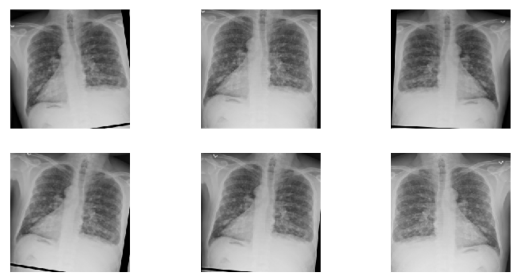

Figure 3. Augmented Images

2. Models and Methodology

2.1 Data Preprocessing

2.1.1 Removing NaN values

The Nan values present in the training set should not be removed as they give information to model about whether the sample has lung opacity or not. The samples with Nan values in the bounding box dimensions are those with no lung opacity.

2.1.2 Skewness of Dataset

Out of the 26,664 samples,about 12000 were labelled as Lung Opacity, 9000 had no lung opacity but were not normal either and about 6000 were Normal. The skewness graph is shown is Fig 1.

2.1.3 Image Augmentation

Due to the relatively small size of our dataset, there was a need to resample the images and augment them to increase the training set and also reduce the chances of over-fitting. This was successfully done by mirroring images side-ways and zooming the images. Some of the examples of augmented images are shown in Fig 2.

2.2 CNNs in Image Segmentation: From R-CNN to Mask R-CNN [9]

2.2.1 R-CNN

R-CNN takes in an image, and creates the bounding boxes and the labels for each object in the image selective search process. It takes in the sub-regions of the image corresponding to objects and outputs the new bounding box coordinates for the object in the sub-region.

It overall generates a set of proposals for bounding boxes. It runs the images in the bounding boxes through a pre-trained AlexNet and finally an SVM to see what object the image in the box is. It runs this box through a linear regression model to output tighter coordinates for the box once the object has been classified.

2.2.2 Fast R-CNN

R-CNN is quite slow due to the reason that it requires forward pass of the CNN (AlexNet) for every single region proposal for every single image. And it has to train three different models separately. The CNN to generate image features, the classifier that predicts the class, and the regression model to tighten the bounding boxes. This makes the pipeline extremely hard to train. This makes the pipeline really hard to train.

Fast R-CNN solves these both problems by using the technique known as RoIPool (Region of Interest Pooling).It shares the forward pass of a CNN for an image across its subregions. the CNN features for each region are obtained by selecting a corresponding region from the CNN's feature map. Then, the features in each region are pooled (usually using max pooling). It uses a single network to compute all the three models to extract image features (CNN) , classify (SVM), and tighten bounding boxes (regressor). Here, SVM classifier is replaced with a softmax layer on top of the CNN to output a classification and adds a linear regression layer parallel to the softmax layer to output the bounding box coordinates.

To summarize, Fast R-CNN takes images with region proposals and outputs the object classifications of each region along with tighter bounding boxes.

2.2.3 Faster R-CNN

In Fast R-CNN, the region proposals are created using Selective Search which is a slow process and hence is a drawback of the overall process. To overcome this shortcoming, Faster R-CNN is was introduced. In faster R-CNN, the region proposals are dependent upon the features of the image that were already calculated wit the forward pass of the CNN. A single CNN is used to both implement region proposals and the classification. This , only one CNN is required to be trained and generates region proposals cost - free.

Faster R-CNN uses a CNN feature extractor to extract image features. Then it uses a CNN region proposal network to create region of interests (RoIs). We apply RoI pooling to warp them into fixed dimension. It is then feed into fully connected layers to make classification and boundary box prediction.

2.2.4 Mask R-CNN

Mask R-CNN, extends Faster R-CNN by adding a branch for predicting an object mask in parallel with the existing branch for bounding box recognition. It is simple to train and adds only a small overhead to Faster R-CNN. It is easy to generalize to other tasks that helps estimating the human poses in the same framework.

In Mask R-CNN, a fully Convolutional Network (FCN) is added on top of the CNN features of Faster R-CNN to generate a binary mask that depicts whether a given pixel is the part of an object or not (Figure 4). It takes a CNN Feature Map and outputs matrix with with 1s on all locations where the pixel belongs to the object and 0s elsewhere. This is known as binary mask.

One important implementation which was done by the Mask R-CNN authors to make this methodology work was to pass the image through RoIAlign instead of RoIPool so that the regions of the feature map selected by RoIPool corresponds more precisely to the regions of the original image. This is required because pixel level segmentation requires pixel level preciseness and fine-grained alignment.In RoIAlign, we avoid rounding down the number of pixels during its calculation to avoid misalignment caused by RoIPool and instead use bi linear interpolation to get precise results. Once these masks are generated , Mask R-CNN combines

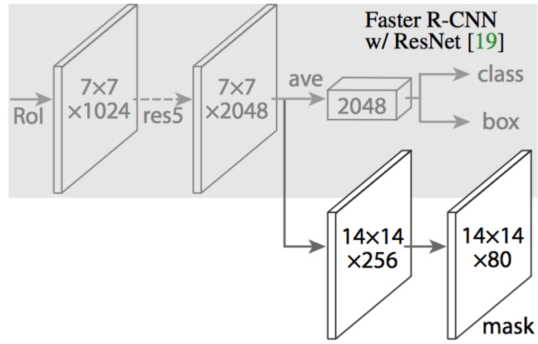

Figure 4

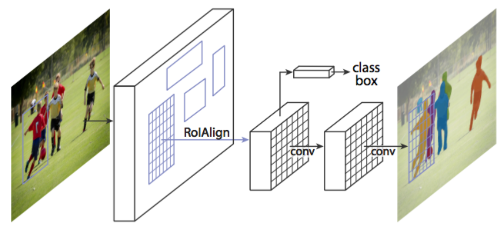

Figure 5. RoIAlign instead of RoIPool

them with the classifications and bounding boxes from Faster R-CNN to generate such wonderfully precise segmentation.

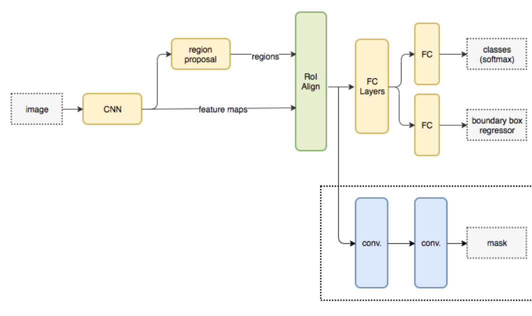

Figure 6. Mask R-CNN Architecture

3. Experiments and Results

The experiments were based on Mask-RCNN model for object-detection and instance segmentation with a backbone of residual networks - Resnet50 and Resnet101. These were the only 2 backbone models that were supported by matterports' implementation of Mask-RCNN. Initially, only the last 2 layers of the models were trained along with using pre-trained weights for the inner layers. The weights were pre-trained on COCO Dataset. After not getting any good results, entire model was trained. The model training relied heavily on image augmentation due to small training set. The augmentation was done by random horizontal and vertical flips, random rotations in the range of (-2,2) degrees and introducing random level of blur and noise. The amount of brightness/gamma augmentations was limited since it could invalidate the original labels.

Different learning rates were also used in a single training. For the initial training a higher learning rate was used as compared to training for later epochs. It was noticed that Mask-RCNN overfits the model easily without image augmentation. The validation-loss was a lot higher than training loss when no image augmentation was used even for low epochs. This wasn't the problem when image-augmentation was used and training was done for all layers.

To evaluate the performance of the model, Mean Average Precision (MAP) and Intersection over Union thresholds: An Average Precision of IoU is calculated for the thresholds $\in \{0.40, 0.45, 0.50, 0.55, 0.60, 0.65, 0.70, 0.75\}$

The IoU of a set of predicted and actual values and Average precision value are given by the formula:

$$\text{IoU(A, B)} = \frac{A \cap B}{A \cup B}$$

$$\text{Avg Precision} = \frac{TP}{TP + FP + FN}$$

The average precision of an image is given as the mean of the precision values at each IoU threshold indicated as t in $\{0.40, 0.45, 0.50, 0.55, 0.60, 0.65, 0.70, 0.75\}$ is given by:

$$\frac{1}{|thresholds|} \sum_t \frac{TP_t}{TP_t + FP_t + FN_t}$$

After submissions, an IoU value for test set of 0.063 was achieved on Mask-RCNN with Resnet50 as backbone and an IoU value for test set of 0.067 with backbone as Resnet101.

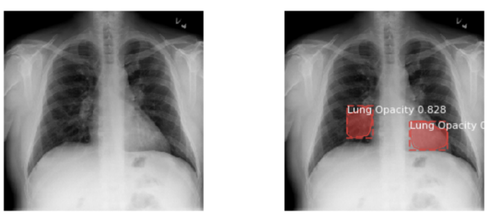

Figure 7. Chest X-Ray with bounding boxes identifying Lung opacities

The Figure 7 indicates the chest X-Ray with lung opacities. The red area shown in the figure is the actual area where the lung opacity is present and the dotted lines denote the bounding box as identified by the Masked RCNN model employed.

4. Summary and Conclusions

From the results above, we could see that while the bounding box covers more area than the actual lung opacity area, the lung opacities are rightly identified. In other words, the sensitivity metric is high. Since the kaggle competition provided test data set does not include the actual labels/areas of lung opacities, metrics like sensitivity, specificity and F1-Score could not be provided.

It is to be noted that the state of the art model which employs Retina Net has an IoU value of 0.25475. Albeit significantly higher than the Masked-RCNN model, Retina Net can be explored as part of the future work as it is deemed out of scope given the time frame constraint for the project.

References

[1] Rui P, Kang K. *National Ambulatory Medical Care Survey: 2015 Emergency Department Summary Tables. Table 27.*

[2] *Deaths: Final Data for 2015. Supplemental Tables. Tables I-21, I-22.*

[3] LeCun, Yann. *LeNet-5, convolutional neural networks*. Retrieved 16 November 2013.

[4] Zhang, Wei (1988). *Shift-invariant pattern recognition neural network and its optical architecture.* Proceedings of annual conference of the Japan Society of Applied Physics.

[5] Zhang, Wei (1990). *Parallel distributed processing model with local space-invariant interconnections and its optical architecture.* Applied Optics. 29 (32): 4790–7. Bibcode:1990ApOpt..29.4790Z. doi:10.1364/AO.29.004790. PMID 20577468

[6] Pranav Rajpurkar, Jeremy Irvin, Kaylie Zhu, Brandon Yang, Hershel Mehta, Tony Duan, Daisy Ding, Aarti Bagul, Curtis Langlotz, Katie Shpanskaya, Matthew P. Lungren, Andrew Y. Ng. *CheXNet: Radiologist-Level Pneumonia Detection on Chest X-Rays with Deep Learning* CoRR,abs/1711.05225 2017

[7] Schalekamp, Steven, van Ginneken, Bram, Meiss, Louis, Peters-Bax, Liesbeth, Quekel, Lorentz GBA, Snoeren, Miranda M, Tiehuis, Audrey M, Wittenberg, Rianne, Karssemeijer, Nico, Schaefer-Prokop, Cornelia M *Bone suppressed images improve radiologists' detection performance for pulmonary nodules in chest radiographs*, European journal of radiology, volume 82, number 12, pages 2399–2405, year 2013, publisher: Elsevier

[8] Zhou, Xiangrong and Hayashi, Tatsuro and Hara, Takeshi and Fujita, Hiroshi and Yokoyama, Ryujiro and Kiryu, Takuji and Hoshi, Hiroaki, *Automatic segmentation and recognition of anatomical lung structures from high-resolution chest CT images*, volume 30, number 5, pages 299–313, year 2006, publisher : Elsevier

[9] Kaiming He and Georgia Gkioxari and Piotr Dollár and Ross B. Girshick,*He2017MaskR, 2017 IEEE International Conference on Computer Vision (ICCV)*,year 2017 ,2980-2988,

[10] Deep multiscale convolutional feature learning for weakly supervised localization of chest pathologies in X-ray images,author=Sedai, Suman and Mahapatra, Dwarikanath and Ge, Zongyuan and Chakravorty, Rajib and Garnavi, Rahil, booktitle=International Workshop on Machine Learning in Medical Imaging,pages=267–275,year=2018, organization=Springer

[11] Lung image segmentation using deep learning methods and convolutional neural networks,author=Kalinovsky, Alexander and Kovalev,Vassili,year=2016,publisher=Minsk: Publishing Center of BSU