Machine Learning Engineer Nano degree
Capstone Proposal

George Samuel

April 21, 2018

Random Sample of NIH Chest X-ray Dataset

Domain Background

One of my primary areas of focus, and in fact the reason I decided to do the Machine Learning nanodegree originally, is Image processing medical image scan. The last part for this course was Deep-Learning and CNN –this part is important part for me. I interesting of medical image on the last a few years the detection of diseases by CNN or NN became more important to make diagnoses easy and safe more time.
CNN use X-ray Dataset of machine learning, computer vision and various other.

As part of the first term of the separate Chest X-ray nanodegree program, I was tasked with using many different computer vision techniques that require a decent amount of manual input and selection to arrive at the end result .With the knowledge gained from the Machine Learning nanodegree, and some of the Deep Learning course on Udacity's website, I wondered if there might be a better approach to this problem - one directly involving deep learning. Deep learning involves utilizing multiple-layered neural networks,

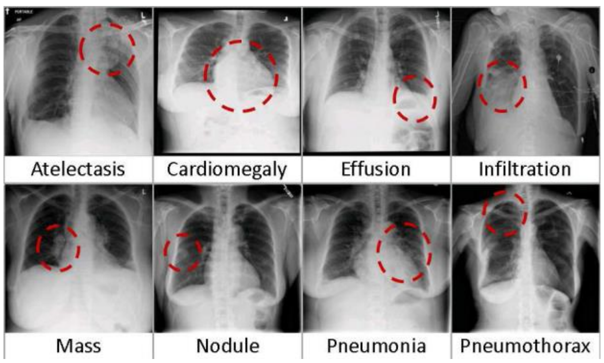
Why this matters

The chest X-ray is one of the most commonly accessible radiological examinations for screening and diagnosis of many lung diseases. A tremendous number of X-ray imaging studies accompanied by radiological reports are accumulated and stored in many modern hospitals' Picture Archiving and Communication Systems (PACS). On the other side, it is still an open question how this type of hospital-size knowledge database containing invaluable imaging informatics (i.e., loosely labeled) can be used to facilitate the data-hungry deep learning paradigms in building truly large-scale high precision computer-aided diagnosis
we present a new chest X-ray database, namely "ChestX-ray8", which comprises 108,948 frontal view X-ray images of 32,717 unique patients with the text mined
Eight disease image labels (where each image can have multi-labels), from the associated radiological reports using natural language processing. Importantly, we demonstrate that these commonly occurring thoracic diseases can be detected and even spatially-located via a unified

weakly supervised multi-label image classification and disease localization framework, which is validated using our proposed dataset. Although the initial quantitative results are promising as reported, deep convolutional neural network based "reading chest X-rays" (i.e., recognizing and locating the common disease patterns trained with only image-level labels) remains a strenuous task for fully-automated

Problem Statement

Reading and diagnosing Chest X-ray images may be an entry-level task for radiologists but, in fact it is a complex reasoning problem which often requires careful observation and good knowledge of anatomical principles, physiology and pathology. Such factors increase the difficulty of developing a consistent and automated technique for reading chest X-ray images while simultaneously considering all common thoracic diseases.



The Main Problem for this issue the diagnoses for the chest issues because there are a lot of Diseases in the same location and Diagnoses for different way Like Mass and Nodule the patient may be has the same Diseases for the the same diagnoses Also Effusion and Infiltration

Many Doctors Tell me that because if I solve for the this model of classification that help the Doctors of radiologist to make easy diagnoses

Datasets and Inputs

There are 15 classes (14 diseases, and one for "No findings") in the full dataset, but since this is drastically reduced version of the full dataset, some of the classes are sparse with the labeled as "No findings"

- Hernia - 13 images
- Pneumonia - 62 images
- Fibrosis - 84 images
- Edema - 118 images
- Emphysema - 127 images

- Cardiomegaly - 141 images
- Pleural_Thickening - 176 images
- Consolidation - 226 images
- Pneumothorax - 271 images
- Mass - 284 images
- Nodule - 313 images
- Atelectasis - 508 images
- Effusion - 644 images
- Infiltration - 967 images
- No Finding - 3044 images

File contents - This is a random sample (5%) of the full dataset:

- sample.zip: Contains 5,606 images with size 1024 x 1024
- sample_labels.csv: Class labels and patient data for the entire dataset
  - Image Index: File name
  - Finding Labels: Disease type (Class label)
  - Follow-up #
  - Patient ID
  - Patient Age
  - Patient Gender
  - View Position: X-ray orientation
  - OriginalImageWidth
  - OriginalImageHeight
  - OriginalImagePixelSpacing_x
  - OriginalImagePixelSpacing_y

https://www.kaggle.com/nih-chest-xrays/sample

Solution Statement

I plan to use deep learning toward the final solution. The reason for this is that a deep learning model may be more effective at determining the important features in a given image than a human being using manual gradient and color thresholds in typical computer vision techniques, as well as other manual programming needed within that process. Specifically, I plan to use a convolutional neural network (CNN), which are very effective at finding patterns within images by using filters to find specific pixel groupings that are important. My aim is for the end result to be both more effective at detecting the lines as well as faster at doing so than common computer vision techniques
And this the way to help and doctor to primary diagnoses

Benchmark Model

I plan to compare the results of the CNN versus the output of the computer vision-based model I used in my Chest X-ray Nano degree project. I will compare the accuracy/mean squared error of the lines to each other to see which is more effective, as well as compare the speed of the two techniques (after training, in the case of the CNN).

Evaluation Metrics

I attempt to build a "machine-human annotated" comprehensive chest X-ray database that presents the realistic clinical and methodological challenges of handling at least tens of thousands of patients (somewhat similar to "ImageNet" in natural images). We also conduct extensive quantitative performance benchmarking on eight common thoracic pathology classification and weakly-supervised localization using ChestX-ray8 database

The main goal is to initiate future efforts by promoting public datasets in this important domain.

Building

Truly large-scale, fully-automated high precision medical diagnosis systems remains a strenuous task. ChestX-ray8 can enable the data-hungry deep neural network paradigms to create clinically meaningful applications, including disease pattern mining, disease correlation analysis, automated

Common radiological report generation, etc. For future work, ChestX-ray8 will be extended to cover more disease classes and integrated with other clinical information, e.g., follow-up studies across time and patient history.

Project Design

- **Programming language** : Python 3.6+
- **Libraries :** Keras, Tensorflow, Scikit-learn, Opencv
- **Workflow :**

1- Read the Dataset and random image of diseases xray.png
2- visualize the dataset according of partitioning of the diseases
3- Convert annotated .png images into labeled numpy arrays. Discard all images with more than one pathology.
4- Make Training And Test of Data
5- Training a small convolutional neural network from scratch for further comparison with transfer learning models.
6- Describe distribution of class labels
7- use a CNN to predict each ailment based off of the X-Ray image.
8- Functions Learning Curves and Confusion Matrix
9- Apply VGG16Network(VGG is a convolutional neural network model proposed by K. Simonyan and A. Zisserman from the University of Oxford in the paper "Very Deep Convolutional Networks for Large-Scale Image Recognition" . The model achieves 92.7% top-5 test accuracy in ImageNet , which is a dataset of over 14 million images belonging to 1000 classes.)
10-Extracting features from the images with the pre-trained network and running a small fully connected network and CNN

References :https://www.nih.gov/news-events/news-releases/nih-clinical-center-provides-one-largest-publicly-available-chest-x-ray-datasets-scientific-community