Predicting Thorax Diseases with NIH Chest X-Rays

CS229 Machine Learning: Final Project

Joy Hsu - joycj@stanford.edu, Kush Khosla - kkhosla@stanford.edu, Peter Lu - peterlu6@stanford.edu
Department of Computer Science, Stanford University

Background

The use of diagnostic imaging has increased dramatically in recent years. A substantial number are chest x-rays used to diagnose a plethora of conditions. These diagnoses are still primarily done by radiologists manually poring over each scan, with no automated triaging or assistance. We aim to use deep learning to predict thorax disease categories using chest x-rays and their metadata with greater than first-pass specialist accuracy.

Problem Statement

Our problem can be cast as a multiclass image classification problem with 15 different labels. We aim to provide a proof of concept of an automated chest x-ray diagnosis system by utilizing the NIH dataset. Deep learning is used to improve the multiclass classification accuracy of thorax disease classification, measured against a baseline of softmax regression.

Datasets

This dataset contains over 110,000 gray scale identically-sized images of size 1024 x 1024 from over 30,000 unique patients. The labels given corresponds to 14 common thorax disease types as well as a "no finding" category of healthy x-rays. Other data given include age, gender, number of visits to hospital, etc. We utilized and processed the above data to feed into our neural net.

Methods

Our custom model included two initial convolution layers with drop out, concatenated with a 50 layer residual net (ResNet50), with additional layers to add in categorical features.

- The initial convolution layers with 2 x 2 strides aim to downsample the 1024 x 1024 image to 256 x 256, which is closer to the 224 x 224 dimensions for the canonical ResNet structure.
- The residual layers learn the subtraction of features from input with shortcut connections, as the net directly connects the input of the (n) th layer to that of the (n + x) th layer.
- The last merged layer consists of the original outputs to the residual net with a concatenated array of other additional data (age, gender, # of visits to hospital)
- Relu activation and the adam optimizer were used for the model as well as various hyperparameter tuning strategies.

We compare our sequential model to softmax regression of all features, which also outputs a prediction of probabilities. We use a custom top k class metric to compute accuracy, after making a prediction based on the output.

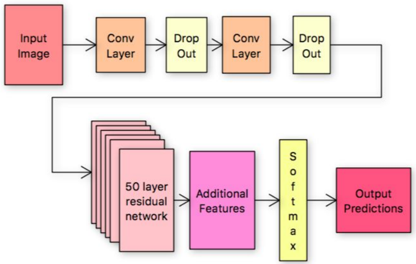

Figure: neural net architecture combining residual network with additional features to output softmax prediction.

Experimental Evaluation and Findings

0.58 accuracy for neural net   0.08 accuracy for softmax regression   0.04 accuracy for random weight initialization softmax regression

Atelactasis	Cardiomegaly	Consolidation	Edema	Effusion	Emphysema	Fibrosis	Hernia	Infiltration	Mass	No Finding	Nodule	Pleural Thickening	Pneumonia	Pneumothorax	
-2.22E-02	-6.12E-01	-4.51E-02	-4.44E-02	-1.21E-02	-2.83E-02	-1.89E-02	-5.13E-01	-3.16E-02	-7.96E-01	-3.54E-02	-3.36E-02	-5.21E-02	-3.40E-02	-4.86E-02	Age
4.98E-01	-4.40E-01	1.48E-01	-1.45E+00	-2.93E-01	-2.46E-01	-3.61E-01	-4.04E-01	-8.78E-03	1.02E-02	-1.55E-01	3.60E-01	5.16E-01	-1.10E-01	9.04E-01	Gender
2.00E-01	2.36E-01	2.00E-01	2.36E-01	2.16E-01	1.76E-01	1.36E-01	-1.20E-01	1.96E-01	-4.70E-01	1.67E-01	1.60E-01	2.13E-01	1.88E-01	2.10E-01	# visits

Conclusions and Future Directions

- The residual net demonstrated greater ability to model complex features than softmax regression, reaching higher accuracy than first-pass specialist accuracy in current diagnostic imaging.
- Future work:
  - Utilizing more storage to increase batch size and filter amount
  - Saliency maps for model interpretability on pixels
- Weight analysis from last layer of neural net:
  - Females are more likely to have edema
  - Males are more likely to have pneumothorax
  - Increased number of hospital visits is positively correlated with many diseases

References

[1] NIH, "NIH clinical center provides one of the largest publicly available chest x-ray datasets to scientific community." 2017.
[2] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," *CoRR*, vol. abs/1512.03385, 2015.