Detecting Pneumonia in X-rays

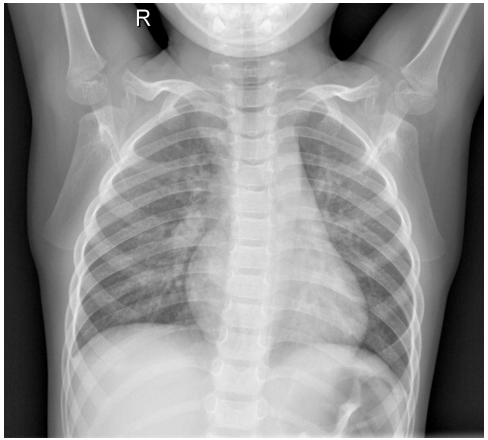
Authors: Tony Samaniego,
Malhar Pandya, Yoosuf Batliwala
-
CECS456
Dr. Wenlu Zhang
CSULB College of Engineering - Dept. of Computer Science

Introduction

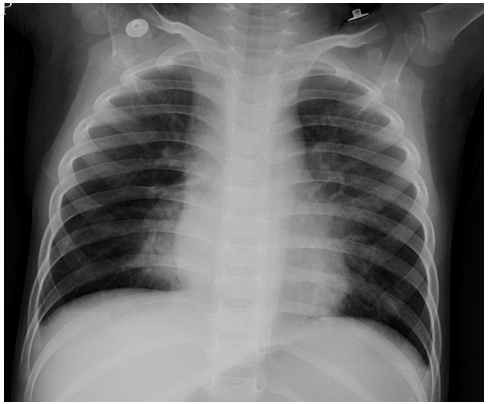
- Dataset: Kaggle’s Chest X-Ray Images(Pneumonia)
  - Given training, testing and validation folder
- Task: Determine if an X-Ray belongs to ‘normal’ or ‘pneumonia’ class
  - No other classes, binary classification
- Data is from Mendeley Data with images being 256 X 256 X 3
- Networks modelled after: alexnet2012, zfnet2013, vggnet2014

Chest X-Ray Dataset

- The dataset is classified into 2 different categories: “NORMAL” and “PNEUMONIA”
- The dataset was verified by two expert physicians before being classified for AI training.
- The dataset was partitioned with 5,216 images for training, 16 images for validation, and 624 images for testing (5,856 images in total).
- The training dataset consisted of 1,341 “NORMAL” images and 3,875 “PNEUMONIA” images resulting in a data imbalance.



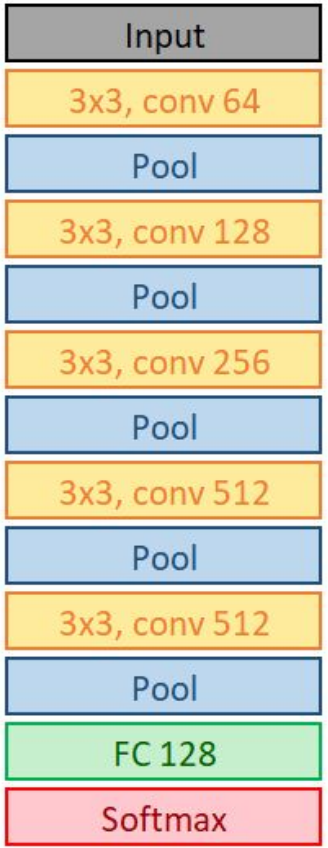
NORMAL



PNEUMONIA

VGGNET2014

- Model initially based off of the VGG16NET.
- Images resized to 224x224x3
- Has 5 convolution layers and 5 max pool layers
- 7,122,434 total and trainable parameters
- As a solution to the data imbalance:
  - Equal number of true pos and true neg samples were used in training set.
  - The model was trained for 40 epochs, tested for accuracy against the testing dataset, augmented the training dataset, and repeated for a total of 120 epochs.



Methodology

The initial training set yielded very low accuracy with the VGG16 clone. Through trial and error, I found that the more simple I made the model, the better the results I would achieve. Once I settled on my model, I began training the same model with 3 sets of augmented data. Each training session consisted of 40 epochs for a total of 120 total epochs. With each consecutive training session, the testing accuracy and F1 score increased.

		Predicted	
		1	0
Actual	1	373	17
	0	48	186

Training Session	Preprocessing
1	• Image Size: (224, 224) • Rescaling: 1./255
2	• Image Size: (224, 224) • Rescaling: 1./255 • Horizontal Flip: True • Vertical Flip: True • Random Rotation: 0.2
3	• Image Size: (224, 224) • Rescaling: 1./255 • Horizontal Flip: True • Vertical Flip: True • Random Rotation: 0.2 • Random Brightness: 0.3 • Random Contrast: 0.3

Training Results

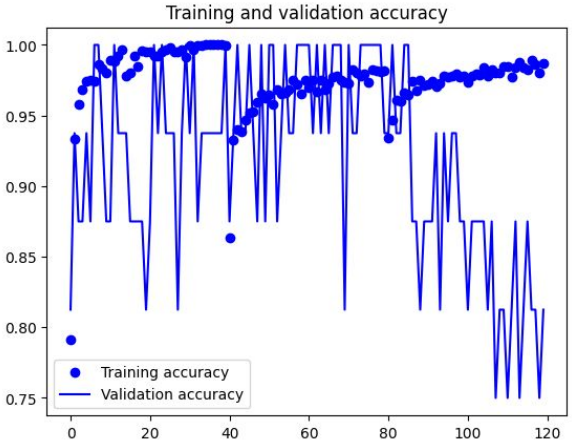
Training 1) loss: 4.505, accuracy 0.8013

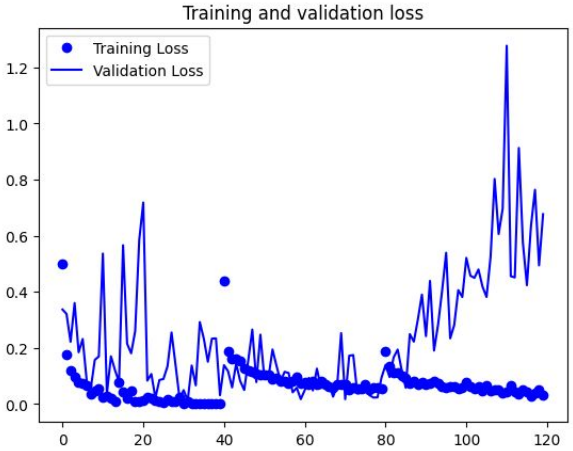
Training 2) loss: 0.856, accuracy 0.8365

Training 3) loss: 0.473, accuracy 0.8958

- Training session 2 had an higher average training and validation accuracy with the lowest training and validation loss.
- Training session 3 yielded the highest testing accuracy and lowest testing loss.

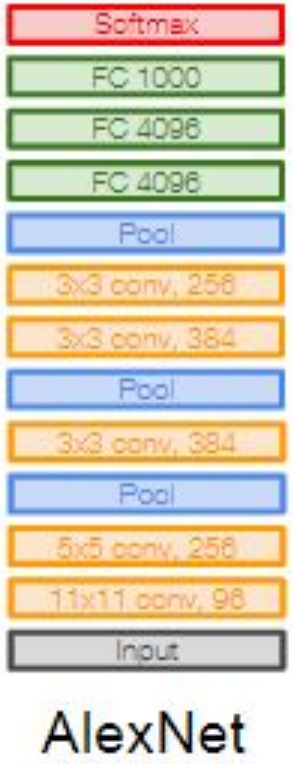
Training 1	Training 2	Training 3
• Accuracy: 80.1% • Precision: 76.0% • Recall : 99.7% • F1: 86.3%	• Accuracy: 83.7% • Precision: 80.3% • Recall : 97.9% • F1: 88.2%	• Accuracy: 89.6% • Precision: 88.6% • Recall : 95.6% • F1: 92.0%





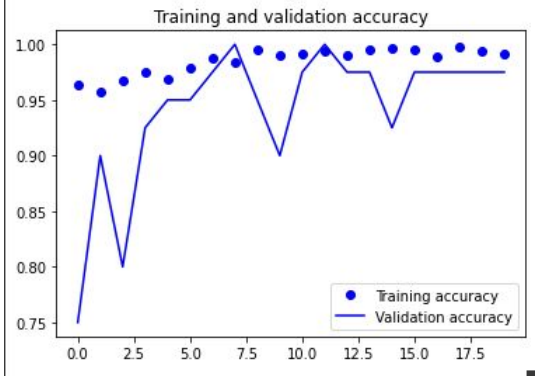
Alexnet Methodology

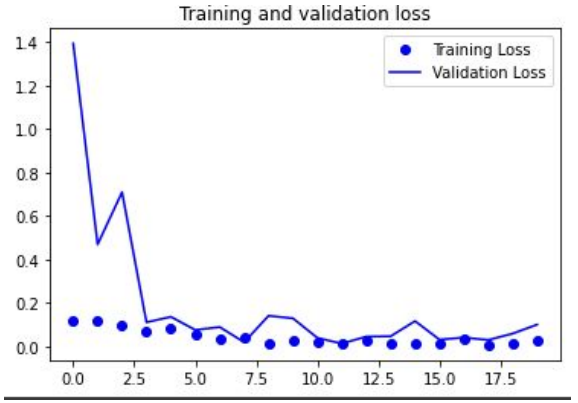
- Alexnet Paper was designed around the Imagenet LSVRC contest
- One problem with classifying objects is the variety objects offer
- Alexnet was designed to overcome this difficulty
- The model I made had around 58 million parameters
- Preprocessing:
  - Alexnet preprocessed their data by centering images on their RGB values
  - They also used horizontal flip for data augmentation and dropout on last 2 FC to combat overfitting



Experimental setup for Alexnet

Training 1	Training 2	Training 3
• Accuracy: 73.4% • Precision: 71% • Recall : 96.6% • F1: 81.8%	• Accuracy: 80.297% • Precision: 77% • Recall : 97.1% • F1: 85.8%	• Accuracy:86.86 % • Precision: 85.6% • Recall 94.8 % • F1:89.9 %
Preprocessing: • None	Preprocessing: • Rescale values[0,1] • 2 Dropouts(0.3)	Preprocessing • Rescale[0,1] • Horizontal and vertical flip • Random rotation(0.2) • Dropout first FC(0.5)

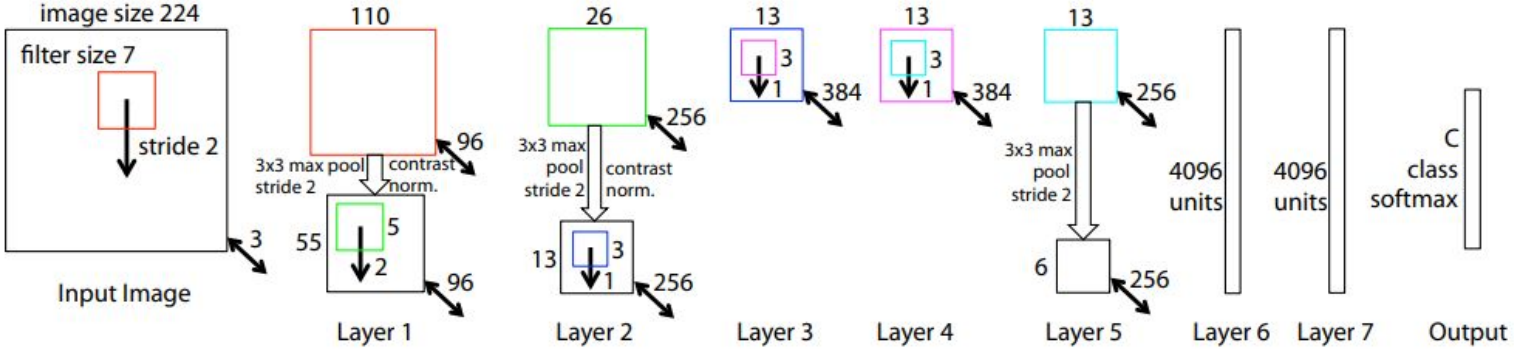




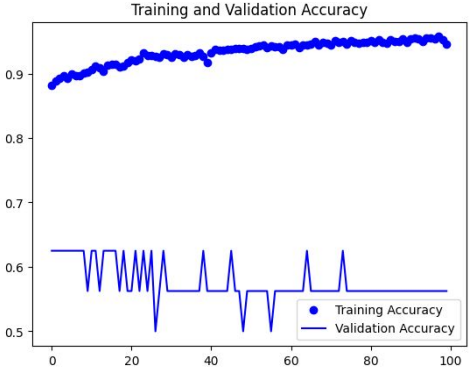
		Predicted	
		1	0
Actual	1	370	20
	0	62	172

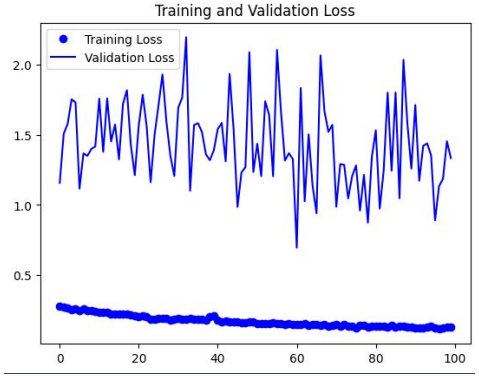
ZFnet Methodology

- Based on the Alexnet model, with improvements such as reduced filter sizes, and local contrast operations.
- Uses a deconvolutional network—essentially the CNN in reverse—to give a visualization of the types of features that it detects in an image.
- Results were hovering around 62.5% with no preprocessing of the images.
- ZFnet paper specified 256x246 crop as well as cropping to center and random flips.
- Paper trained for 70 epochs, I trained for 100.



ZFnet Structure & Results





Training Session with Above Preprocessing (100 epochs)

- Accuracy on test data: 90.71%
- Accuracy on training data: 96.22%
- Precision: 88%
- Recall: 99%
- F1: 93%

		Predicted	
		1	0
Actual	1	387	3
	0	55	179

	Preprocessing
100 epochs	• Resizing: 256,256 • Rescaling: 1.0, 127.5 • Random flip: Horizontal and Vertical • Random Rotation: 0.2 • Random Contrast: 0.5 • Random Crop: 127,255

Analysis of Results

- Three different models, with different pre-processing and methodologies, so it's hard to compare.
- Our preprocessing (resizing, rescaling, augmentation) improved all of our results tremendously.
- Overall, ZFnet achieved the highest accuracy out of the three, but all three were fairly close.

VGGNet Clone	AlexNet	ZFnet
Accuracy: 89.6%	Accuracy: 86.86%	Accuracy: 90.71%

Conclusion

- The reason we picked the X-Ray Pneumonia Dataset is because it was a dataset with highly practical application.
- We analysed our models by comparing the validation accuracy with our test accuracies.
- In conclusion, the network you pick matters, for example VGG outperformed Alexnet
- The way you preprocess the data also has a huge impact