

# CheXNet: Radiologist-Level Pneumonia Detection on Chest X-Rays with Deep Learning

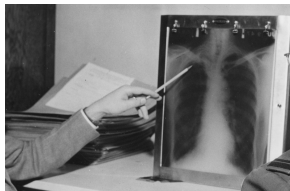
Yanan Yang

CS732 Paper Presentation, 2019  
New Jersey Institute of Technology  
Newark, NJ, USA



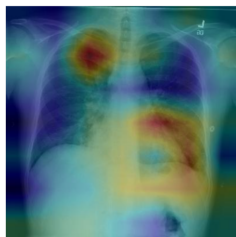
## AI Beats Radiologists at Pneumonia Detection?

- Chest X-rays are currently the best available method for diagnosing pneumonia.
- Detecting pneumonia in chest X-rays is a challenging task that relies on the availability of expert radiologists.
- ChexNet, an algorithm can detect pneumonia from chest X-rays at a level exceeding practicing radiologists.
- This would be the first example of superhuman AI performance in medicine, if so.

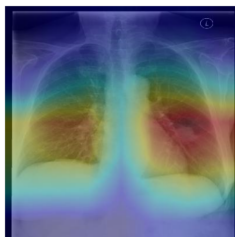


# 1. Difficulties on Pneumonia Detection

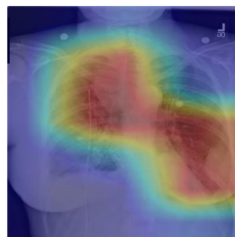
- The appearance of pneumonia in X-ray images is often vague  
can overlap with other diagnoses  
can mimic many other benign abnormalities.



(a) Patient with multifocal community-acquired pneumonia.



(b) Patient with a left lung nodule.

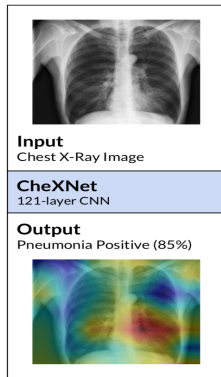


(c) Patient with primary lung malignancy and two large masses

Fig2

## 2. CNNs go deeper

- The path for information from the input layer until the output layer (and for the gradient in the opposite direction) becomes so big, that they can get vanished before reaching the other side.
- CheXNet: is a type of image analysing AI called a DenseNet (a variant of a ConvNet, similar to a ResNet) that was trained to detect abnormalities on chest x-rays, using the ChestXray14 dataset.



## ChexNet Network

- 121 layers CNN  
input: chest X-ray image  
output: the probability of pneumonia along with a heatmap localizing the areas
- Based on Dense Connection(Huang wt al. 2016), and Batch normalization(Ioff zegedy, 2015)
- Modified the loss function to optimize the sum of unweighted binary cross entropy losses

$$L(X, y) = \sum_{c=1}^{14} [-y_c \log p(Y_c = 1|X) - (1 - y_c) \log p(Y_c = 0|X)]$$



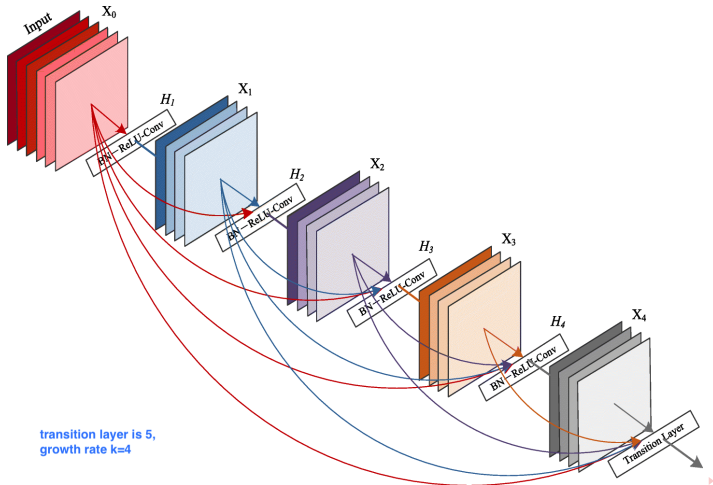
## Dense Connection Struction(Huang wt al. 2016)

- Transition layer
  1. connect each layer to every other layer with the same filter size, where the dimensions of the feature maps remains constant within a block, but the number of filters changes between them
  2. not sum the output feature maps of the layer with the incoming feature maps but concatenate them, where  $X(i) = H([X(1), X(2), X(3), , X(i - 1)])$
- Growth rate K

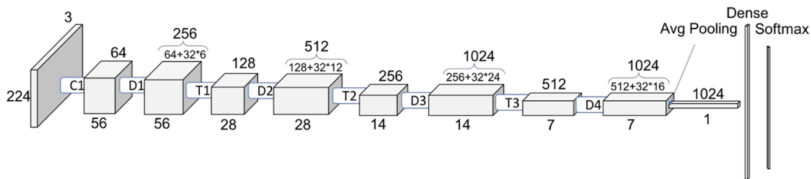
Since we are concatenating feature maps, this channel dimension is increasing at every layer. If we make  $H_l$  to produce k feature maps every time, then we can generalize for the l-th layer:  $K(l) = K(0) + K * (l - 1)$



# Dense Connection Struction(Huang wt al. 2016)



## Dense Connection Struction(Huang wt al. 2016)

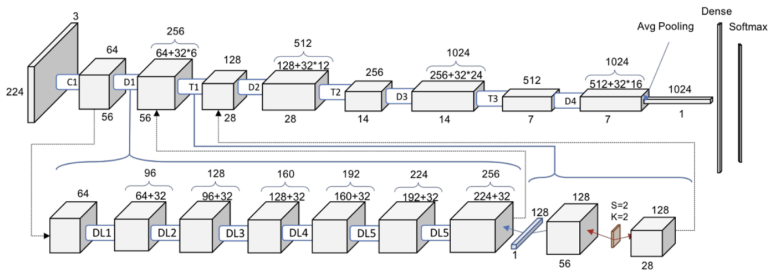


- simple scheme on the architecture of the DenseNet-121: the measures under each volume represent the sizes of the width and depth  
the numbers on top represents the feature maps dimension
- concatenating (concatenate means add dimension, but not add values!) new information to the previous volume, which is being reused





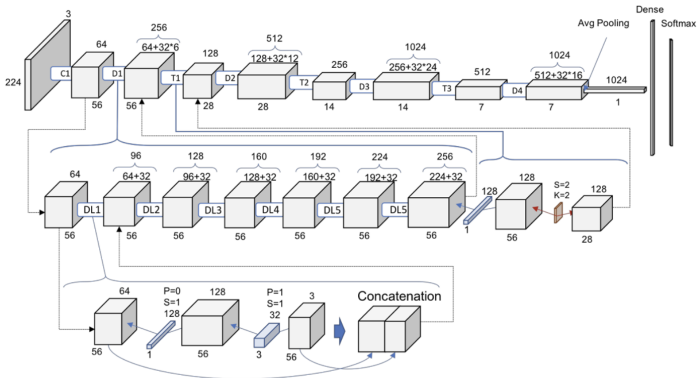
# DenseNet: Go deeper and understand what is actually happening inside every block



- From 64 to 256 after 6 layers: adding to the previous volume these 32 new feature maps; The volume within a Dense Block remains constant
- Transition Block performs as 1x1 convolution with 128 filters, followed by a 2x2 pooling with a stride of 2. The volume and the



# DenseNet: Go 1 level deeper



- 1x1 convolution with 128 filters to reduce the feature maps size, and the perform a more expensive 3x3 convolution (include the padding to ensure the dimensions remain constant)



## ChexNet: Loss Function

- For a single example in the training set, optimizing the weighted binary cross entropy loss:

$$L(X, y) = -w_+ \cdot y \log p(Y = 1|X) - w_- \cdot (1 - y) \log p(Y = 0|X)$$

- where  $p(Y = i|X)$  is the probability that the network assigns to the label  $i$ ,  
 $w_+ = |N|/(|P|+|N|)$ ,  
 $w_- = |P|/(|P|+|N|)$  with  $|P|$  and  $|N|$  the number of positive cases and negative cases of pneumonia in the training set respectively.



## ChexNet: Model Architecture

- 121-layer Dense Convolutional Network (DenseNet) (Huang et al., 2016)
- Optimized loss function
- Replaced the final fully connected layer with one that has a single output(1/0)
- The weights of the network are initialized with weights from a model pretrained on ImageNet, using mini- batches of size 16
- Initial learning rate of 0.001 that is decayed by a factor of 10 each time, validation loss plateaus after an epoch, and pick the model with the lowest validation loss.



## Data set : Chest X-ray 14

- ChestX-ray14 dataset released by Wang et al. (2017) which contains 112,120 frontal-view X-ray images of 30,805 unique patients.
- Each image with up to 14 different thoracic pathology labels using automatic extraction methods on radiology reports
- Label images that have pneumonia as one of the annotated pathologies as positive examples and label all other images as negative examples.
- Randomly split the dataset into:  
training (28744 patients, 98637 images),  
validation (1672 patients, 6351 images),  
test (389 patients, 420 images),  
no patient overlap between the sets.



## Data set : pre-processing

- Randomly splited the dataset into:  
training (28744 patients, 98637 images),  
validation (1672 patients, 6351 images),  
test (389 patients, 420 images),  
no patient overlap between the sets.
- Downscale the images to  $224 \times 224$ , and normalized based on mean and standard deviation
- Augmented the training data with random horizontal flipping





# 1. Comparison with Radiologist

	F1 Score (95% CI)
Radiologist 1	0.383 (0.309, 0.453)
Radiologist 2	0.356 (0.282, 0.428)
Radiologist 3	0.365 (0.291, 0.435)
Radiologist 4	0.442 (0.390, 0.492)
Radiologist Avg.	0.387 (0.330, 0.442)
CheXNet	0.435 (0.387, 0.481)

- Computed the F1 score (harmonic avg.)
- Used the bootstrap to construct 95% confidence intervals (CIs) on 10,000 bootstrap samples
- The difference in F1 scores, the 95% CI on the difference does not contain 0, CheXNet is statistically significantly higher





## 2. Comparison with Previous CNNs on ChestX-ray14

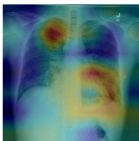
Pathology	Wang et al. (2017)	Yao et al. (2017)	CheXNet (ours)
Atelectasis	0.716	0.772	<b>0.8094</b>
Cardiomegaly	0.807	0.904	<b>0.9248</b>
Effusion	0.784	0.859	<b>0.8638</b>
Infiltration	0.609	0.695	<b>0.7345</b>
Mass	0.706	0.792	<b>0.8676</b>
Nodule	0.671	0.717	<b>0.7802</b>
<b>Pneumonia</b>	<b>0.633</b>	<b>0.713</b>	<b>0.7680</b>
Pneumothorax	0.806	0.841	<b>0.8887</b>
Consolidation	0.708	0.788	<b>0.7901</b>
Edema	0.835	0.882	<b>0.8878</b>
Emphysema	0.815	0.829	<b>0.9371</b>
Fibrosis	0.769	0.767	<b>0.8047</b>
Pleural Thickening	0.708	0.765	<b>0.8062</b>
Hernia	0.767	0.914	<b>0.9164</b>

- CheXNet outperforms the best published results on all 14 pathologies in the ChestX-ray14 dataset. Especially on Mass, Nodule, Pneumonia, and Emphysema, CheXNet has a margin  $>0.05$  AUROC over previous state of the art results.

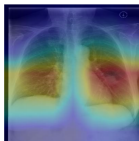


# Achievements

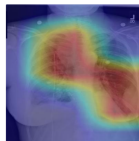
- localizes pathologies it identifies using Class Activation Maps, which highlight the areas of the X-ray that are most important for making a particular pathology classification.



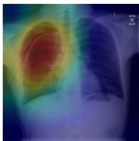
(a) Patient with multifocal community-acquired pneumonia. The model correctly detects the airspace disease in the left lower and right upper lobes to arrive at the pneumonia diagnosis.



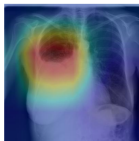
(b) Patient with a left lung nodule. The model identifies the left lower lobe lung nodule and correctly classifies the pathology.



(c) Patient with primary lung malignancy and two large masses, one in the left lower lobe and one in the right upper lobe adjacent to the mediastinum. The model correctly identifies both masses in the X-ray.



(d) Patient with a right-sided pneumothorax and chest tube. The model detects the abnormal lung



(e) Patient with a large right pleural effusion (fluid in the pleural space). The model correctly labels the effu-



(f) Patient with congestive heart failure and cardiomegaly (enlarged heart). The model correctly identi-



# Achievements

- 1. Instead of outputting one binary label, ChexNet outputs a vector  $t$  of binary labels indicating the absence or presence of each of the 14 pathology
- 2. The final output is the predicted probability of the presence of each pathology class, since replaced with the final fully connected layer producing a 14 dimensional output
- 3. Modified loss function optimize the sum of unweighted binary cross entropy losses

$$L(X, y) = \sum_{c=1}^{14} [-y_c \log p(Y_c = 1|X) - (1 - y_c) \log p(Y_c = 0|X)]$$



## Limitation:

- 1. Only frontal X-ray images for training and testing, only 4 doctors label images
- 2. Neither the model nor the radiologists were not permitted to use patient history, the CXR14 dataset has labels that don't really match the images, thus the accuracy for training and testing are hard to say ChexNet is better
- 3. Original:  $1024 \times 1034$  pixel, downsampling to  $224 \times 224$ , which means they downsampled with a reduced dynamic range, So there is something like a 20-50 fold reduction in image information compared to clinical images.
- The system learns to match good quality image labels despite the flawed training data. This is very interesting, but without seeing the results in the other classes it is hard to be sure that the pneumonia result is not an outlier. Even if it is true, there is no doubt in my mind that the performance would be better with cleaner training labels.



## My suggestion

- 1. Add ROC curve and a few metrics
- 2. Do some outlier analysis
- 3. More doctors



## References

- 1. Huang, Gao, Liu, Zhuang, Weinberger, Kilian Q, and van der Maaten, Laurens. Densely connected convolutional networks. arXiv preprint arXiv:1608.06993, 2016.
- 2. Yao, Li, Poblenz, Eric, Dagunts, Dmitry, Covington, Ben, Bernard, Devon, and Lyman, Kevin. Learning to diagnose from scratch by exploiting dependencies among labels. arXiv preprint arXiv:1710.10501, 2017.
- 3. Ioffe, Sergey and Szegedy, Christian. Batch normalization: Accelerating deep network training by reducing internal covariate shift. In International Conference on Machine Learning, 2015.
- 4. <https://stanfordmlgroup.github.io/>
- 5. <https://medium.com/@BalintBotz/a-few-thoughts-about-chexnet-and-the-way-human-performance-should-and-should-not-be-measured-68031dca7bf>

