Convolutional Neural Network to Detect Thorax Diseases from Multi-View Chest X-Rays

THE UNIVERSITY OF SYDNEY

ICONIP 2019

Maram Monshi, Josiah Poon, Vera Chung

Introduction

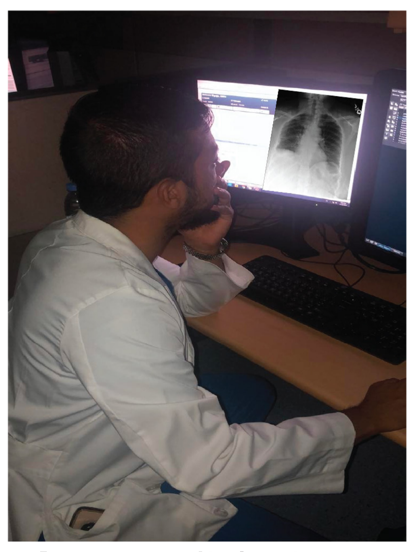

Fig 1. Radiologist Examins a Chest X-Ray

Chest X-Ray: It is the most common radiologist exams in the world that demands correct and immediate diagnosis of a patient's thorax to avoid life threatening diseases.

Problem: Certified radiologists are hard to find. Stress, fatigue and experience contribute to the quality of an examination.

Solution: Automated & precise system that can flag potentially life-threatening diseases to handle emergency cases efficiently.

Example: Convolution neural network (CNN) is a supervised deep learning model that is able to learn useful features which are beyond the limit of radiology detection.

Related Work

Dataset: IU X-Ray (2015), ChestX-ray14 (2017), CheXpert, PadChest & MIMIC-CXR (2019).

Model: CheXNet, text-image embedding network (TieNet) & attention guided convolutional neural network (AG-CNN).

Approach: CNN such as AlexNet, VGG-16, DenseNet & ResNet.

Gap: Using only frontal view, long training time & low accuracy.

Consistent with recent models: We focus on training CNN models to detect 12 common thorax diseases.

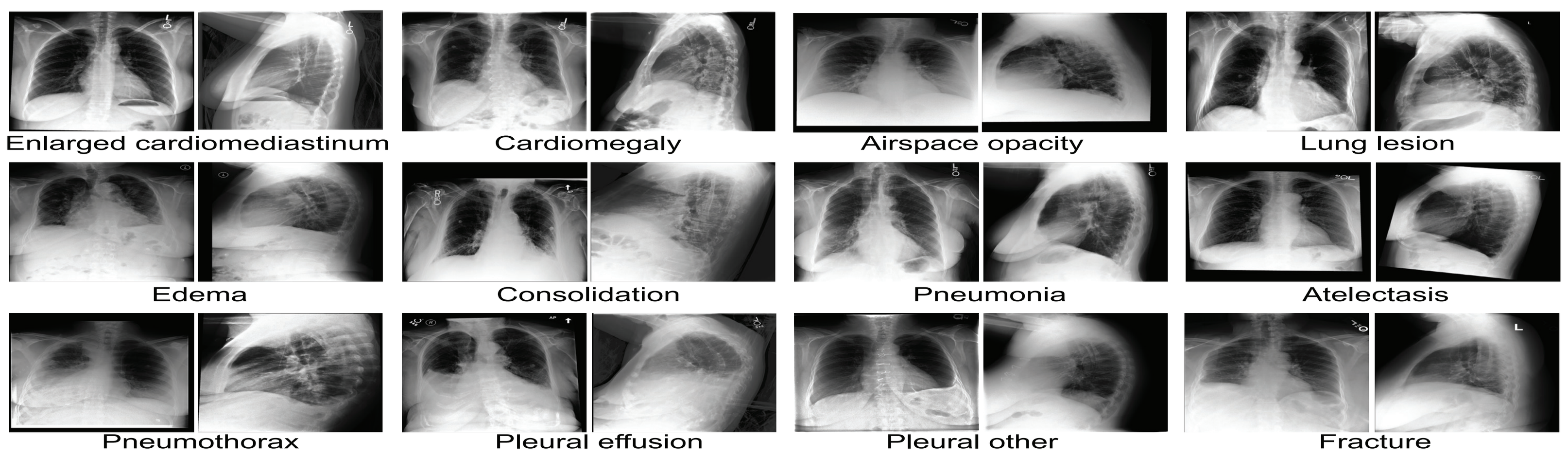

Fig 2. Examples of 12 Thoracic Diseases from MIMIC-CXR Dataset

Unique from past works: We propose a novel **stage-wise training** approach to observe the model's performance => <u>reduce training time</u> & <u>increase accuracy</u>. We adopt a combination of **recent techniques on multi-view chest X-rays** including Res-Net-50, transfer learning, fine tuning, <u>fit one cycle function</u> & <u>discriminative learning rates.</u>

Proposed Model

Structure Overview: We divided the task of detecting thorax diseases into 12 sub-tasks. Each task considers the presence/absence of a disease. For each binary label problem, ResNet-50 is used as the baseline CNN architecture. ResNet-50 consists of 49 convolution layers & ends with 1 fully connected layer.

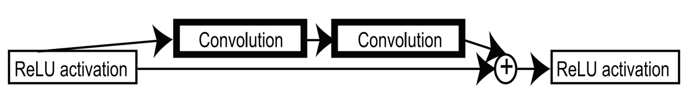

Fig 3. A Basic Residual Block

last output   convolution operations   generated output

$$\boxed{x_l} = \boxed{F}(x_{l-1}) + \boxed{x_{l-1}}$$

Training Stages:

Embrace transfer learning (PyTorch & fastai)

Observe the model's performance (fit-one-cycle method)

Use the optimal learning rate finder

parameters   learning rate   gradient

$$\boxed{\theta_t^l} = \theta_{t-1}^l - \boxed{\eta^l} \cdot \boxed{\nabla_{\theta^l}} J(\theta)$$

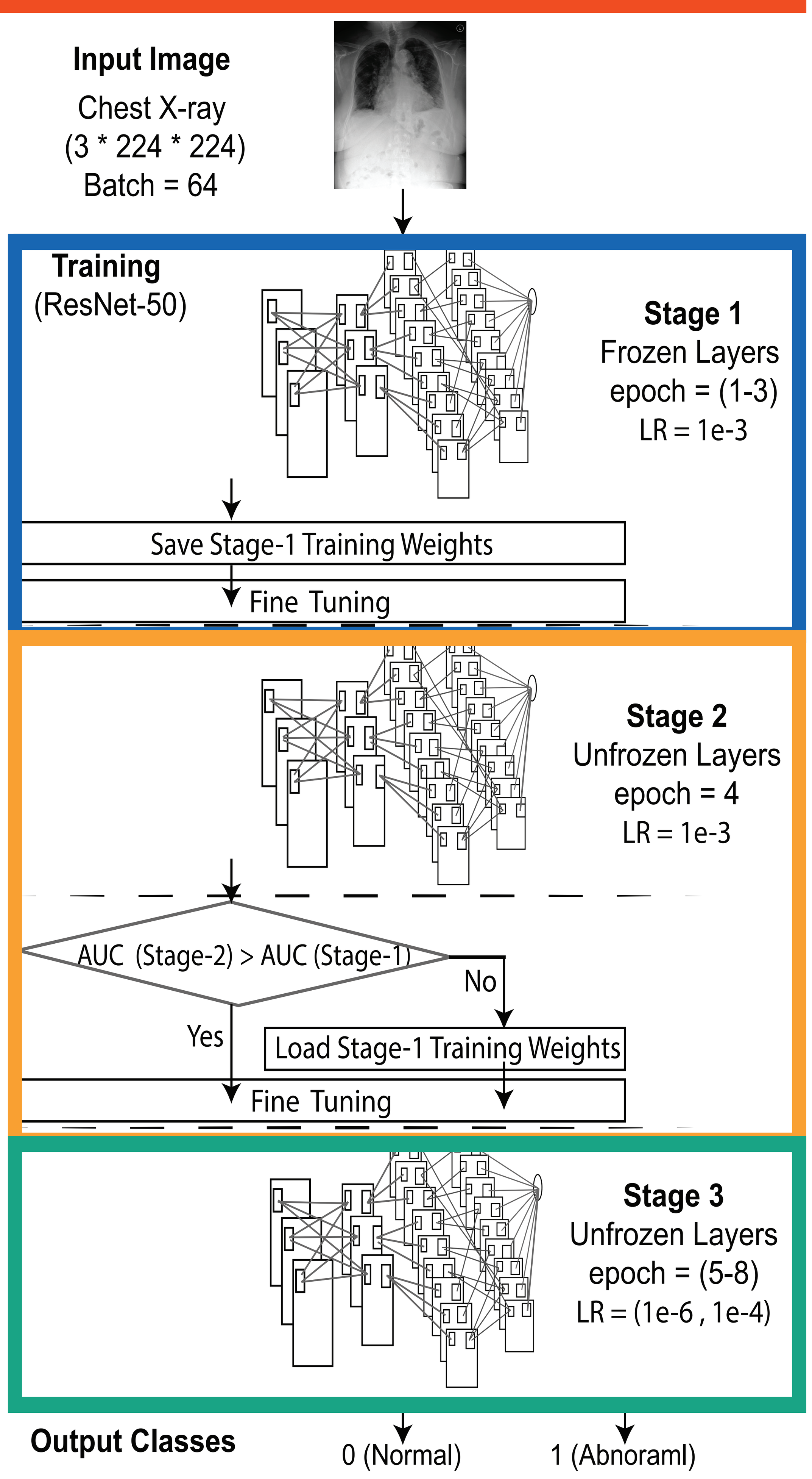

Fig 4. Overall Illustration of Our Model

Experiment

Dataset: We organized a subset of 10% of the MIMIC-CXR dataset into training set (33,195) and validation set (3,688). We dropped uncertain & unknown labels.

For example, our subset includes 6932 images with Cardiomegaly.

Pathology	Positive (%)	Negative (%)
Enlarged Cardiom.	1019 (2.8)	35367 (97.19)
Cardiomegaly	6932 (18.79)	29951 (81.2)
Airspace Opacity	7582 (20.42)	29542 (79.57)
Lung Lesion	1060 (2.82)	36472 (97.17)
Edema	3964 (11.06)	31859 (88.93)
Consolidation	1410 (3.8)	35634 (96.19)
Pneumonia	2738 (7.83)	32202 (92.16)
Atelectasis	6356 (17.54)	29876 (82.45)
Pneumothorax	1523 (4.05)	36059 (95.94)
Pleural Effusion	7869 (21.34)	28994 (78.65)
Pleural Other	425 (1.13)	37132 (98.86)
Fracture	805 (2.13)	36829 (97.86)

Pre-Processing: We employed several augmentation strategies. Training using 224 px reduces training time without worsening AUC.

Parameter	Value
Size	224
Flip (horizontally)	True
Lighting	0.3
Affine	0.5

Image Size (pixels)	Epoch								Avg. AUC per Epoch
	1	2	3	4	5	6	7	8	
299	0.565	0.733	0.758	0.791	0.798	0.804	0.804	**0.807**	0.757
224	0.725	0.733	0.747	0.785	0.793	0.799	**0.802**	0.802	0.773

Training: We used 4 NVIDIA Tesla P4 GPUs to reduce training time.

No. of GPUs	Epoch								Avg. Time per Epoch (min)
	1	2	3	4	5	6	7	8	
1	32:42	32:26	32:36	34:34	33:40	33:52	33:58	34:00	33:28
4	13:32	12:54	13:01	13:05	13:07	13:08	13:07	13:06	13:07

Result: We computed the Area Under Curve (AUC) of each pathology on the validation set for each of the eight training epochs.

Stage 3 results in larger AUC values than **stage-1** & **stage 2** due to the discriminative learning rates.

Pathology	Epoch							
	1	2	3	4	5	6	7	8
Enlarged Cardiom.	0.670	0.694	0.700	0.544	0.702	0.705	0.708	0.710
Cardiomegaly	0.725	0.733	0.747	0.785	0.793	0.799	0.802	0.802
Airspace Opacity	0.621	0.687	0.694	0.712	0.730	0.730	0.733	0.737
Lung Lesion	0.520	0.638	0.612	0.638	0.651	0.688	0.730	0.729
Edema	0.816	0.848	0.857	0.887	0.892	0.894	0.896	0.897
Consolidation	0.748	0.758	0.769	0.778	0.788	0.797	0.797	0.799
Pneumonia	0.556	0.531	0.545	0.497	0.550	0.585	0.580	0.587
Atelectasis	0.706	0.706	0.743	0.827	0.830	0.835	0.837	0.838
Pneumothorax	0.710	0.786	0.817	0.839	0.853	0.862	0.868	0.860
Pleural Effusion	0.837	0.869	0.881	0.891	0.903	0.906	0.905	0.899
Pleural Other	0.585	0.637	0.676	0.533	0.707	0.736	0.739	0.727
Fracture	0.546	0.563	0.576	0.606	0.636	0.648	0.711	0.741
Average	0.670	0.704	0.718	0.711	0.753	0.765	0.776	0.777

In 5 out of 7 overlap pathologies, our model performs better than both DualNet models.

Pathology	DualNet		Our Model
	PA + Lateral	AP + Lateral	Multi-View
Enlarged Cardiom.	-	-	0.710
Cardiomegaly	0.840	0.755	0.802
Airspace Opacity	-	-	0.737
Lung Lesion	-	-	0.730
Edema	0.734	0.749	0.897
Consolidation	0.632	0.623	0.799
Pneumonia	0.625	0.593	0.587
Atelectasis	0.766	0.671	0.838
Pneumothorax	0.706	0.621	0.868
Pleural Effusion	0.757	0.733	0.906
Pleural Other	-	-	0.739
Fracture	-	-	0.741
Average	0.722	0.677	0.779

Analysis: Our model is trained on a better annotated chest X-rays (CheXpert labeler) than DualNet (NegBio labeler). We reach improved results over those achieved by DualNet using small image sizes 224 by 224 pixels instead of 512 by 512.

Limitation: Class labels in the training set are noisy. The positive-negative subsets ratio was highly imbalanced in some pathologies. Yet, our model's AUC is above 0.7.

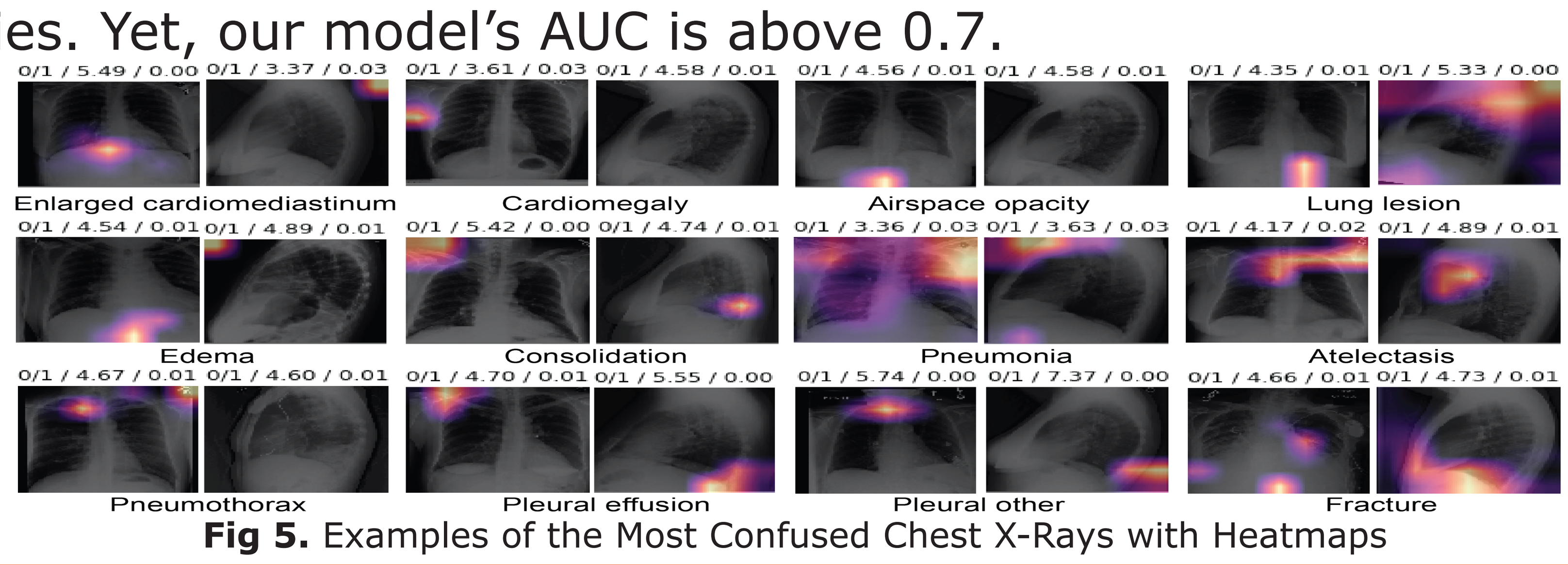

Fig 5. Examples of the Most Confused Chest X-Rays with Heatmaps

Conclusion

Contribution: We proposed ResNet-50 CNN based stage-wise models to detect 12 thorax diseases on 10% of the largest chest X-rays dataset to date, MIMIC-CXR. The absolute labelling performance with an average weighted AUC of 0.779 is encouraging.

Future Work: We plan to improve our CNN model performance through data augmentation. We will incorporate useful information from the free-text radiology reports like patient's history to accurately recognize the presence/absence of thorax diseases.