

Abstract

The increased availability of X-ray image archives (e.g. the ChestX-ray14 dataset from the NIH Clinical Center) has triggered a growing interest in deep learning techniques. To provide better insight into the different approaches, and their applications to chest X-ray classification, we investigate a powerful network architecture in detail: the ResNet-50. Building on prior work in this domain, we consider transfer learning with fine-tuning as well as the training of a dedicated X-ray network from scratch. To leverage the high spatial resolutions of X-ray data, we also include an extended ResNet-50 architecture, and a network integrating non-image data (patient age, gender and acquisition type) in the classification process.

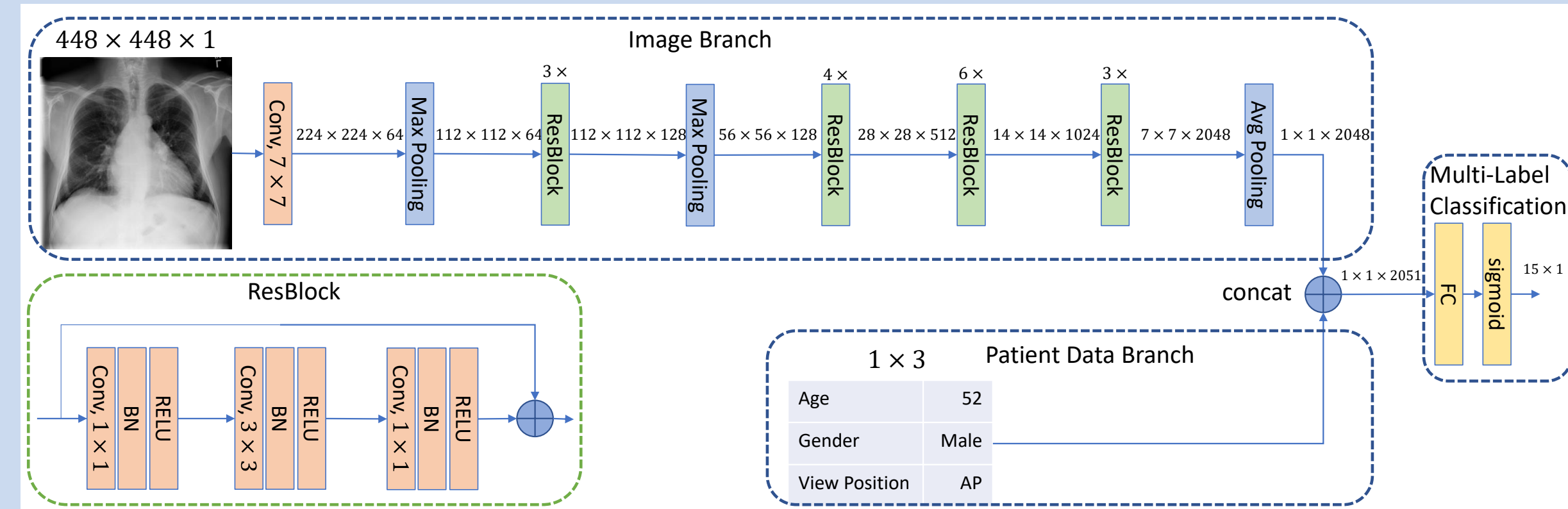
In a systematic evaluation, using 5-fold re-sampling and a multi-label loss function, we evaluate the performance of the different approaches for pathology classification by ROC statistics. We observe a considerable spread in the achieved performance and conclude that the X-ray-specific ResNet-50, integrating non-image data yields the best overall results.

Motivation and Challenges

- In the UK, 23,000 chest X-rays (CXRs) were not formally reviewed by a radiologist over a one year period
- It is expected that the demand for CXR readings will increase
- Providing an system with high sensitivity to reduce the workload is desired
- Pathologies in CXRs have a very diverse appearance (i.e. size, location, etc.)

Architecture

- CXR-Net based on ResNet-50 architecture:
 1. Larger input size (i.e. for the detection of small structures, which could be indicative of a pathology)
 2. Included non-image features (i.e. radiologists also use information beyond the image for their diagnosis)
 3. Trained from scratch on ChestXray14 [1] (CXR14) (i.e. learning X-ray specific representation)



- Class-averaged binary cross entropy as loss function:

$$\ell(\vec{y}, \vec{f}) = \frac{1}{M} \sum_{m=1}^M H[y_m, f_m], \text{ with } H[y, f] = -y \log f - (1 - y) \log(1 - f).$$

Results

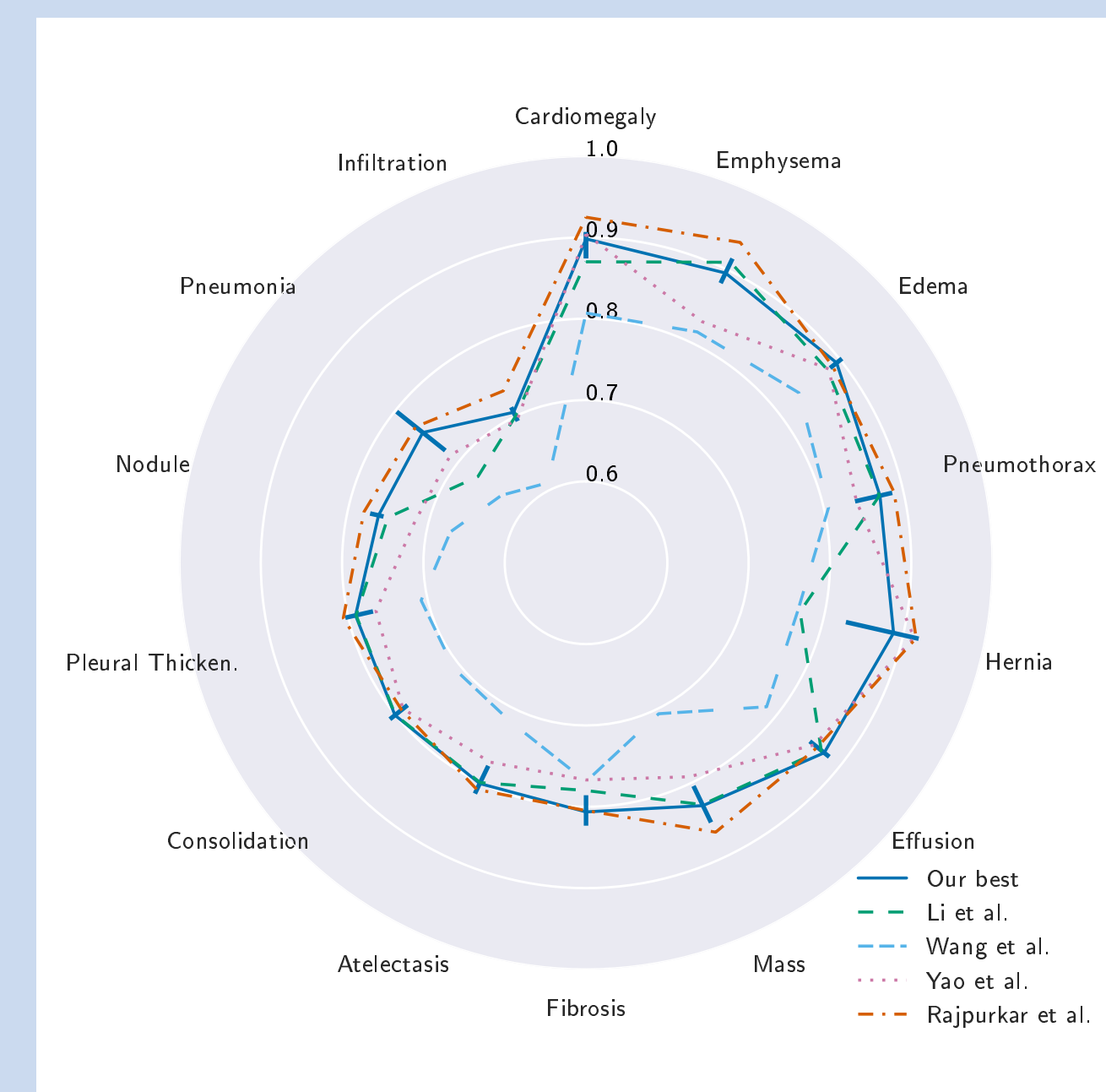
- Evaluation on entire corpus of CXR14 and comparison to baseline experiments
- 5 times re-sampling scheme with patient stratification for training, validation and testing (70%, 10%, and 30% respectively)

Pathology	Without non-image features		With non-image features	
	FT	CXR-Net	FT	CXR-Net
Cardiomegaly	88.5 ± 0.7	89.7 ± 0.3	88.4 ± 0.8	89.8 ± 0.8
Emphysema	89.2 ± 1.0	88.3 ± 1.3	89.4 ± 1.2	89.1 ± 1.2
Edema	89.1 ± 0.4	88.8 ± 0.5	89.1 ± 0.7	88.9 ± 0.3
Hernia	85.5 ± 3.8	87.5 ± 4.5	88.2 ± 3.2	89.6 ± 4.4
Pneumothorax	87.0 ± 0.8	85.9 ± 0.9	86.5 ± 0.6	85.9 ± 1.1
Effusion	87.1 ± 0.2	87.6 ± 0.2	87.2 ± 0.3	87.3 ± 0.3
Mass	82.2 ± 1.0	83.9 ± 0.9	82.2 ± 1.0	83.2 ± 0.3
Fibrosis	80.0 ± 0.9	79.2 ± 1.6	80.0 ± 0.9	78.9 ± 0.5
Atelectasis	80.3 ± 0.7	79.2 ± 0.7	80.1 ± 0.6	79.1 ± 0.4
Consolidation	79.5 ± 0.5	80.0 ± 0.3	79.6 ± 0.5	80.0 ± 0.7
Pleural Thicken.	79.0 ± 0.7	78.0 ± 1.1	78.6 ± 1.1	77.1 ± 1.3
Nodule	72.6 ± 0.9	75.1 ± 1.3	74.7 ± 0.6	75.8 ± 1.4
Pneumonia	74.4 ± 1.6	75.3 ± 2.2	73.3 ± 1.3	76.7 ± 1.5
Infiltration	69.9 ± 0.6	70.2 ± 0.5	70.2 ± 0.2	70.0 ± 0.7
Average	81.7 ± 1.0	82.1 ± 1.2	82.0 ± 0.9	82.2 ± 1.1
No Findings	76.9 ± 0.5	77.1 ± 0.4	76.8 ± 0.4	77.1 ± 0.3

- All our experiments with non-image features slightly increase the AUC on average to its counterpart

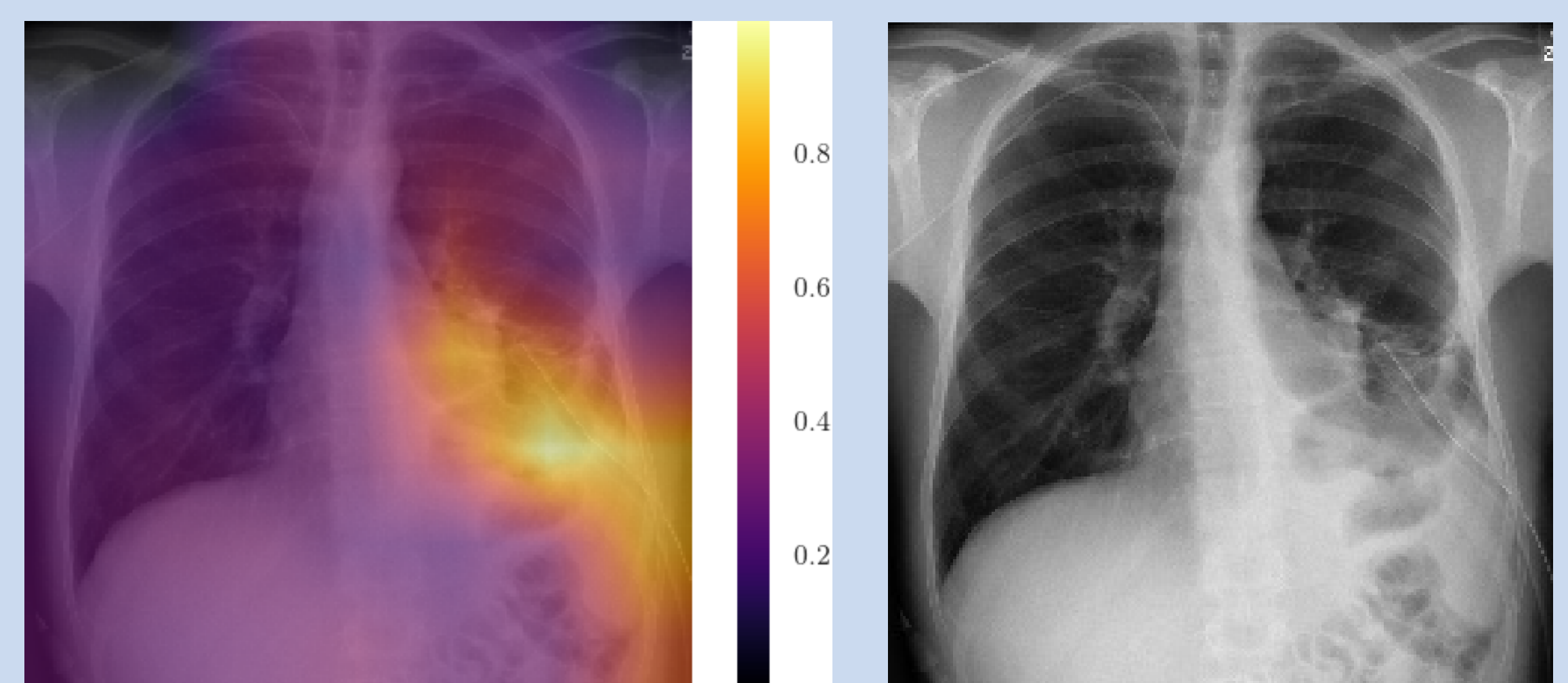
Comparison to Others

- AUC for different pathologies (min and max over folds as error bars)



Visual Example of Grad-CAM Results

- Example image of "pneumothorax" with a drain from the testset



- As discussed by [2] the quality of the labels, and their precise medical interpretation might be a limiting factor

Conclusion and Outlook

- Best overall results can be reported for the model that is exclusively trained with CXRs and incorporating non-image data
- Our fine-tuned ResNet-50 model achieves state-of-the-art results in four out of fourteen classes compared to [3]

Future work:

- Investigation of practical use of deep learning in clinical practice
- New architectures for leveraging label dependencies
- Incorporating segmentation information

References

1. Wang, X. et al. *ChestX-Ray8: Hospital-Scale Chest X-Ray Database and Benchmarks on Weakly-Supervised Classification and Localization of Common Thorax Diseases in CVPR* (2017).
2. Oakden-Rayner, L. Exploring the ChestXray14 dataset: problems. <https://lukeoakdenrayner.wordpress.com/2017/12/18/> (2018) (2017).
3. Rajpurkar, P. et al. *CheXNet: Radiologist-Level Pneumonia Detection on Chest X-Rays with Deep Learning in CoRR* (2017).