

Detecting
abnormali-
ties on
chest
X-rays
using deep
neural
networks

Swaroop
Kumar M
L

Department
of Studies
in
Computer
Science
University
of Mysore

Introduction

Data

Baselines

Experiments

Results

Examples

References

References

References

References

Detecting abnormalities on chest X-rays using deep neural networks

Swaroop Kumar M L
Department of Studies in Computer Science
University of Mysore

June 14, 2019

Abstract

Detecting
abnormali-
ties on
chest
X-rays
using deep
neural
networks

Swaroop
Kumar M
L

Department
of Studies
in
Computer
Science
University
of Mysore

Introduction

Data

Baselines

Experiments

Results

Examples

References

References

References

References

Inspired by previous work, we develop algorithms that can detect abnormalities on the x-ray. The algorithm explains these detections by generating heatmaps pointing out areas of the image that most influenced it.

Abstract

Detecting
abnormali-
ties on
chest
X-rays
using deep
neural
networks

Swaroop
Kumar M
L

Department
of Studies
in
Computer
Science
University
of Mysore

Introduction

Data

Baselines

Experiments

Results

Examples

References

References

References

References

Inspired by previous work, we develop algorithms that can detect abnormalities on the x-ray. The algorithm explains these detections by generating heatmaps pointing out areas of the image that most influenced it. We establish baselines, benchmark against previous work and show that

Abstract

Detecting
abnormali-
ties on
chest
X-rays
using deep
neural
networks

Swaroop
Kumar M
L

Department
of Studies
in
Computer
Science
University
of Mysore

Introduction

Data

Baselines

Experiments

Results

Examples

References

References

References

References

Inspired by previous work, we develop algorithms that can detect abnormalities on the x-ray. The algorithm explains these detections by generating heatmaps pointing out areas of the image that most influenced it. We establish baselines, benchmark against previous work and show that

- Transfer-learning from a large non-TB dataset dramatically improves TB detection

Abstract

Detecting abnormalities on chest X-rays using deep neural networks

Swaroop Kumar M L

Department of Studies in Computer Science University of Mysore

Introduction

Data

Baselines

Experiments

Results

Examples

References

References

References

References

Inspired by previous work, we develop algorithms that can detect abnormalities on the x-ray. The algorithm explains these detections by generating heatmaps pointing out areas of the image that most influenced it. We establish baselines, benchmark against previous work and show that

- Transfer-learning from a large non-TB dataset dramatically improves TB detection
- Models in the domain show inferior performance on external data from a different hospital system, but

Abstract

Detecting abnormalities on chest X-rays using deep neural networks

Swaroop Kumar M L

Department of Studies in Computer Science University of Mysore

Introduction

Data

Baselines

Experiments

Results

Examples

References

References

References

References

Inspired by previous work, we develop algorithms that can detect abnormalities on the x-ray. The algorithm explains these detections by generating heatmaps pointing out areas of the image that most influenced it. We establish baselines, benchmark against previous work and show that

- Transfer-learning from a large non-TB dataset dramatically improves TB detection
- Models in the domain show inferior performance on external data from a different hospital system, but
- Recent techniques such as mixup and progressive resizing improve performance and generalization.

Abstract

Detecting abnormalities on chest X-rays using deep neural networks

Swaroop Kumar M L

Department of Studies in Computer Science University of Mysore

Introduction

Data

Baselines

Experiments

Results

Examples

References

References

References

References

Inspired by previous work, we develop algorithms that can detect abnormalities on the x-ray. The algorithm explains these detections by generating heatmaps pointing out areas of the image that most influenced it. We establish baselines, benchmark against previous work and show that

- Transfer-learning from a large non-TB dataset dramatically improves TB detection
- Models in the domain show inferior performance on external data from a different hospital system, but
- Recent techniques such as mixup and progressive resizing improve performance and generalization.

We achieve performance competitive with previous work in detecting pneumonia-like and other abnormalities on the NIH chestX-ray14 dataset and

Abstract

Detecting abnormalities on chest X-rays using deep neural networks

Swaroop Kumar M L

Department of Studies in Computer Science University of Mysore

Introduction

Data

Baselines

Experiments

Results

Examples

References

References

References

References

Inspired by previous work, we develop algorithms that can detect abnormalities on the x-ray. The algorithm explains these detections by generating heatmaps pointing out areas of the image that most influenced it. We establish baselines, benchmark against previous work and show that

- Transfer-learning from a large non-TB dataset dramatically improves TB detection
- Models in the domain show inferior performance on external data from a different hospital system, but
- Recent techniques such as mixup and progressive resizing improve performance and generalization.

We achieve performance competitive with previous work in detecting pneumonia-like and other abnormalities on the NIH chestX-ray14 dataset and in detecting tuberculosis on the Shenzhen hospital dataset, and achieve state-of-the-art performance on the Montgomery county tuberculosis dataset.

Abstract

Detecting abnormalities on chest X-rays using deep neural networks

Swaroop Kumar M L

Department of Studies in Computer Science University of Mysore

Introduction

Data

Baselines

Experiments

Results

Examples

References

References

References

References

Inspired by previous work, we develop algorithms that can detect abnormalities on the x-ray. The algorithm explains these detections by generating heatmaps pointing out areas of the image that most influenced it. We establish baselines, benchmark against previous work and show that

- Transfer-learning from a large non-TB dataset dramatically improves TB detection
- Models in the domain show inferior performance on external data from a different hospital system, but
- Recent techniques such as mixup and progressive resizing improve performance and generalization.

We achieve performance competitive with previous work in detecting pneumonia-like and other abnormalities on the NIH chestX-ray14 dataset and in detecting tuberculosis on the Shenzhen hospital dataset, and achieve state-of-the-art performance on the Montgomery county tuberculosis dataset. We look for potential sources of bias and evaluate our baseline with respect to gender, age and view position.

Problem definition

Detecting abnormalities on chest X-rays using deep neural networks

Swaroop
Kumar M
L

Department of Studies in Computer Science University of Mysore

Introduction

Data

Baselines

Experiments

Results

Examples

References

References

References

References

References

When air in the alveoli is replaced with pus, blood and other fluids, referred to as consolidation and commonly caused by pneumonia, or when abscesses in the lung rupture forming cavities, indicating a tuberculosis infection, these are visible on the chest x-ray.

Problem definition

Detecting abnormalities on chest X-rays using deep neural networks

Swaroop Kumar M L

Department of Studies in Computer Science University of Mysore

Introduction

Data

Baselines

Experiments

Results

Examples

References

References

References

References

When air in the alveoli is replaced with pus, blood and other fluids, referred to as consolidation and commonly caused by pneumonia, or when abscesses in the lung rupture forming cavities, indicating a tuberculosis infection, these are visible on the chest x-ray.

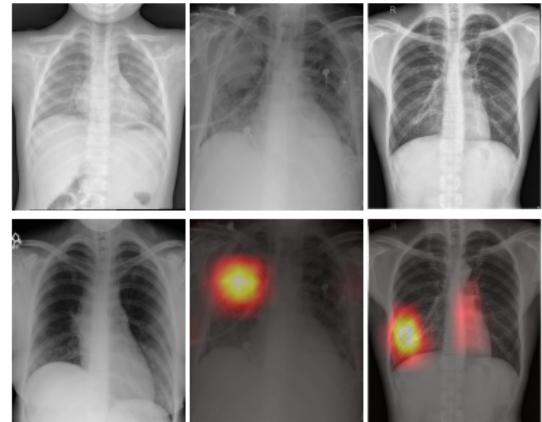


Figure: From left to right: The first column shows two *Normal* images. Columns 2 and 3 show images with *Pneumonia* and *Tuberculosis* respectively, the first row showing the original images and the second showing the same overlaid with heatmaps localizing the abnormalities, which we call *examples*

Detecting
abnormali-
ties on
chest
X-rays
using deep
neural
networks

Swaroop
Kumar M

L
Department
of Studies
in
Computer
Science
University
of Mysore

Introduction

Data

Baselines

Experiments

Results

Examples

References

References

References

References

An algorithm that can automatically detect these abnormalities can be useful in numerous ways.

An algorithm that can automatically detect these abnormalities can be useful in numerous ways.

- ① For active case finding in high-risk populations, for example, with mobile x-ray vans[1]

An algorithm that can automatically detect these abnormalities can be useful in numerous ways.

- ① For active case finding in high-risk populations, for example, with mobile x-ray vans[1]
- ② As an initial screening test before or along with other tests such as a sputum examination

An algorithm that can automatically detect these abnormalities can be useful in numerous ways.

- ① For active case finding in high-risk populations, for example, with mobile x-ray vans[1]
- ② As an initial screening test before or along with other tests such as a sputum examination
- ③ To aid a radiologist in her workflow by sorting her queue based on severity, suggesting areas to consider in an image, providing a second opinion, etc.

An algorithm that can automatically detect these abnormalities can be useful in numerous ways.

- ① For active case finding in high-risk populations, for example, with mobile x-ray vans[1]
- ② As an initial screening test before or along with other tests such as a sputum examination
- ③ To aid a radiologist in her workflow by sorting her queue based on severity, suggesting areas to consider in an image, providing a second opinion, etc.

Our work focuses on the core algorithm. We divide the problem into, and explore, four sub-problems.

Classification

Detecting
abnormali-
ties on
chest
X-rays
using deep
neural
networks

Swaroop
Kumar M
L

Department
of Studies
in
Computer
Science
University
of Mysore

Introduction

Data

Baselines

Experiments

Results

Examples

References

References

References

References

For our primary dataset, a large collection of chest x-rays annotated with multiple abnormalities including pneumonia [2], we formulate the problem as a multi-class multi-label classification problem.

Classification

Detecting
abnormali-
ties on
chest
X-rays
using deep
neural
networks

Swaroop
Kumar M
L
Department
of Studies
in
Computer
Science
University
of Mysore

Introduction

Data

Baselines

Experiments

Results

Examples

References

References

References

References

For our primary dataset, a large collection of chest x-rays annotated with multiple abnormalities including pneumonia [2], we formulate the problem as a multi-class multi-label classification problem.

The task is to learn a function $f : X \rightarrow 2^L$

Classification

Detecting abnormalities on chest X-rays using deep neural networks

Swaroop Kumar M L

Department of Studies in Computer Science University of Mysore

Introduction

Data

Baselines

Experiments

Results

Examples

References

References

References

References

For our primary dataset, a large collection of chest x-rays annotated with multiple abnormalities including pneumonia [2], we formulate the problem as a multi-class multi-label classification problem.

The task is to learn a function $f : X \rightarrow 2^L$

For the Shenzhen hospital tuberculosis dataset[3], we formulate the problem as a binary classification problem.

Classification

Detecting abnormalities on chest X-rays using deep neural networks

Swaroop Kumar M L

Department of Studies in Computer Science University of Mysore

Introduction

Data

Baselines

Experiments

Results

Examples

References

References

References

References

For our primary dataset, a large collection of chest x-rays annotated with multiple abnormalities including pneumonia [2], we formulate the problem as a multi-class multi-label classification problem.

The task is to learn a function $f : X \rightarrow 2^L$

For the Shenzhen hospital tuberculosis dataset[3], we formulate the problem as a binary classification problem.

The task is to learn a function $f : X \rightarrow \{0, 1\}$

Explainability

Detecting
abnormali-
ties on
chest
X-rays
using deep
neural
networks

Swaroop
Kumar M

L

Department
of Studies
in
Computer
Science
University
of Mysore

Introduction

Data

Baselines

Experiments

Results

Examples

References

References

References

References

Deep neural networks have outperformed previous methods in several domains. However, they remain black-boxes with millions of parameters, leading to a lack of trust and limiting their use in routine clinical practice.

Explainability

Detecting abnormalities on chest X-rays using deep neural networks

Swaroop Kumar M

L
Department
of Studies
in
Computer
Science
University
of Mysore

Introduction

Data

Baselines

Experiments

Results

Examples

References

References

References

References

Deep neural networks have outperformed previous methods in several domains. However, they remain black-boxes with millions of parameters, leading to a lack of trust and limiting their use in routine clinical practice.

Several methods have been proposed to make these models more interpretable, broadly falling into two categories:

- ① Methods that create a proxy model

Explainability

Detecting abnormalities on chest X-rays using deep neural networks

Swaroop Kumar M

L
Department
of Studies
in
Computer
Science
University
of Mysore

Introduction

Data

Baselines

Experiments

Results

Examples

References

References

References

References

Deep neural networks have outperformed previous methods in several domains. However, they remain black-boxes with millions of parameters, leading to a lack of trust and limiting their use in routine clinical practice.

Several methods have been proposed to make these models more interpretable, broadly falling into two categories:

- ① Methods that create a proxy model
- ② Methods that generate a saliency map

Explainability

Detecting abnormalities on chest X-rays using deep neural networks

Swaroop Kumar M

L
Department
of Studies
in

Computer
Science
University
of Mysore

Introduction

Data

Baselines

Experiments

Results

Examples

References

References

References

References

Deep neural networks have outperformed previous methods in several domains. However, they remain black-boxes with millions of parameters, leading to a lack of trust and limiting their use in routine clinical practice.

Several methods have been proposed to make these models more interpretable, broadly falling into two categories:

- ① Methods that create a proxy model
- ② Methods that generate a saliency map

Given an input $x = \{x_1, x_2, x_3 \dots x_n\} \in X$, and a set of class labels $L = \{l_1, l_2, l_3 \dots l_c\}$, the task is to compute the attribution

$$A_j = \{a_{j1}, a_{j2}, a_{j3} \dots a_{jn}\} \in \mathbb{R}^n \quad (1)$$

for each class $l_j \in L$ where $a_{ji} \in A_j$ is a measure of the relevance of the i^{th} feature to the model's inference regarding the j^{th} class.

Generalizability

Detecting
abnormali-
ties on
chest
X-rays
using deep
neural
networks

Swaroop
Kumar M
L

Department
of Studies
in
Computer
Science
University
of Mysore

Introduction

Data

Baselines

Experiments

Results

Examples

References

References

References

References

A test set is considered representative of data that will be encountered in the external world and is used exclusively to evaluate a model. However, true generalization to new datasets may be lower than expected.

Generalizability

Detecting
abnormali-
ties on
chest
X-rays
using deep
neural
networks

Swaroop
Kumar M
L

Department
of Studies
in
Computer
Science
University
of Mysore

Introduction

Data

Baselines

Experiments

Results

Examples

References

References

References

References

A test set is considered representative of data that will be encountered in the external world and is used exclusively to evaluate a model. However, true generalization to new datasets may be lower than expected.

Two datasets may have different distributions. In the context of biomedical imaging, datasets may be collected from different hospital systems and machines.

Generalizability

Detecting
abnormali-
ties on
chest
X-rays
using deep
neural
networks

Swaroop
Kumar M
L

Department
of Studies
in
Computer
Science
University
of Mysore

Introduction

Data

Baselines

Experiments

Results

Examples

References

References

References

References

A test set is considered representative of data that will be encountered in the external world and is used exclusively to evaluate a model. However, true generalization to new datasets may be lower than expected.

Two datasets may have different distributions. In the context of biomedical imaging, datasets may be collected from different hospital systems and machines.

The dataset used to train a model may have confounding variables that do not exist in other datasets. For example, in [4], Zech et al. show that CNNs were able to directly detect the hospital system and department within a hospital system from a chest radiograph where saliency maps showed high activation in image corners.

Weighted average of saliency maps

Detecting abnormalities on chest X-rays using deep neural networks

Swaroop Kumar M L

Department of Studies in Computer Science University of Mysore

Introduction

Data

Baselines

Experiments

Results

Examples

References

References

References

References

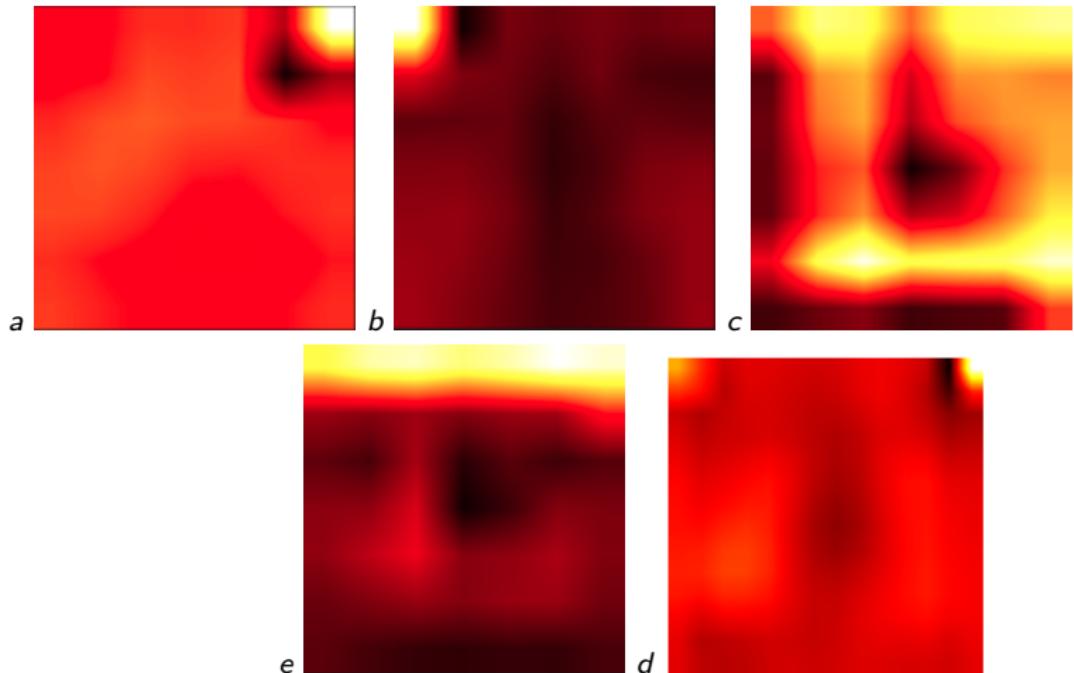


Figure: Average saliency maps for *pneumonia*. Clockwise from the top-left: a) Baseline, b) Baseline with more data augmentation, c) Trained with margins cropped, d) Trained on NIH CXR-14, tested on Guangzhou and e) Trained without pre-training on ImageNet

Fairness

Detecting
abnormali-
ties on
chest
X-rays
using deep
neural
networks

Swaroop
Kumar M
L

Department
of Studies
in
Computer
Science
University
of Mysore

Introduction

Data

Baselines

Experiments

Results

Examples

References

References

References

References

Algorithms may inadvertently learn and leverage biases in the datasets, discriminate based on race, gender, etc. and amplify existing social inequities.

Fairness

Detecting
abnormali-
ties on
chest
X-rays
using deep
neural
networks

Swaroop
Kumar M
L

Department
of Studies
in
Computer
Science
University
of Mysore

Introduction

Data

Baselines

Experiments

Results

Examples

References

References

References

References

Algorithms may inadvertently learn and leverage biases in the datasets, discriminate based on race, gender, etc. and amplify existing social inequities.

In [5], Buolamwini et al. show that facial recognition datasets are overwhelmingly composed of light-skinned individuals and that commercial gender classification systems performed worse for dark-skinned people and females, with a difference in accuracy of more than 30% between light-skinned males and dark-skinned females.

Fairness

Detecting abnormalities on chest X-rays using deep neural networks

Swaroop Kumar M L

Department of Studies in Computer Science University of Mysore

Introduction

Data

Baselines

Experiments

Results

Examples

References

References

References

References

Algorithms may inadvertently learn and leverage biases in the datasets, discriminate based on race, gender, etc. and amplify existing social inequities.

In [5], Buolamwini et al. show that facial recognition datasets are overwhelmingly composed of light-skinned individuals and that commercial gender classification systems performed worse for dark-skinned people and females, with a difference in accuracy of more than 30% between light-skinned males and dark-skinned females.

A simple measure against bias could be to hide variables like gender and race from a model, but complex machine learning models learn to use other correlated variables as proxy for hidden ones (for example, zipcode as correlated with race and the word 'women' in an institution's name as correlated with the gender of its students[6]).

We measure the potential for discrimination by training models with architectures similar to the abnormality detection model, to identify gender and age group from images alone. We then test our baseline model's performance across genders and age groups.

Data

Detecting
abnormali-
ties on
chest
X-rays
using deep
neural
networks

Swaroop
Kumar M

L
Department
of Studies
in
Computer
Science
University
of Mysore

Introduction

Data

Baselines

Experiments

Results

Examples

References

References

References

References

We use the NIH ChestX-ray14 dataset[7] and the Shenzhen hospital tuberculosis dataset[3] to train models to detect pneumonia and other abnormalities, and tuberculosis respectively.

Data

Detecting
abnormali-
ties on
chest
X-rays
using deep
neural
networks

Swaroop
Kumar M

L
Department
of Studies
in
Computer
Science
University
of Mysore

Introduction

Data

Baselines

Experiments

Results

Examples

References

References

References

References

We use the NIH ChestX-ray14 dataset[7] and the Shenzhen hospital tuberculosis dataset[3] to train models to detect pneumonia and other abnormalities, and tuberculosis respectively.

We then use two external datasets, the Guangzhou medical center pediatric pneumonia dataset[8] and the Montgomery county tuberculosis dataset[3] as *external* datasets to test the ability of these models to generalize to other hospital systems.

Detecting abnormalities on chest X-rays using deep neural networks

Swaroop Kumar M

L
Department of Studies in Computer Science University of Mysore

Introduction

Data

Baselines

Experiments

Results

Examples

References

References

References

References

	Training	Testing	Different Hospital system?
Pneumonia	NIH CXR-14	NIH CXR-14	No
	NIH CXR-14	Guangzhou	Yes
	Guangzhou	Guangzhou	No
Tuberculosis	Shenzhen	Shenzhen	No
	Shenzhen	Montgomery	Yes
	Montgomery	Montgomery	No

Table: Datasets used for training and testing

NIH CXR-14

Detecting
abnormali-
ties on
chest
X-rays
using deep
neural
networks

Swaroop
Kumar M
L

Department
of Studies
in
Computer
Science
University
of Mysore

Introduction

Data

Baselines

Experiments

Results

Examples

References

References

References

References

The NIH chestX-ray14 dataset consists of 112,120 chest x-ray images of 30,805 unique patients which was collected using the PAC system of the National Institutes of Health. Each image was labeled with up-to 14 abnormalities algorithmically using the associated radiology report in natural language.

NIH CXR-14

Detecting
abnormali-
ties on
chest
X-rays
using deep
neural
networks

Swaroop
Kumar M
L

Department
of Studies
in
Computer
Science
University
of Mysore

Introduction

Data

Baselines

Experiments

Results

Examples

References

References

References

References

The NIH chestX-ray14 dataset consists of 112,120 chest x-ray images of 30,805 unique patients which was collected using the PAC system of the National Institutes of Health. Each image was labeled with up-to 14 abnormalities algorithmically using the associated radiology report in natural language.

About half or 60,361 of these are labeled as *No finding* and the rest are labeled with one of more of the 14 abnormalities. Some abnormalities are more common than others, the most common being *Infiltration*, which is present in 19,894 images, and the least common being *Hernia*, which is present in only 227 images.

NIH CXR-14

Detecting abnormalities on chest X-rays using deep neural networks

Swaroop Kumar M L

Department of Studies in

Computer Science
University of Mysore

Introduction

Data

Baselines

Experiments

Results

Examples

References

References

References

References

The NIH chestX-ray14 dataset consists of 112,120 chest x-ray images of 30,805 unique patients which was collected using the PAC system of the National Institutes of Health. Each image was labeled with up-to 14 abnormalities algorithmically using the associated radiology report in natural language.

About half or 60,361 of these are labeled as *No finding* and the rest are labeled with one of more of the 14 abnormalities. Some abnormalities are more common than others, the most common being *Infiltration*, which is present in 19,894 images, and the least common being *Hernia*, which is present in only 227 images.

We split the dataset into train, validation and test sets roughly in the ratio 70:10:20. We make sure that there is no patient overlap, that is, all images of a patient are in the same subset since patient overlap may lead to overfitting.

Abnormality	Number of images			
	Train	Validation	Test	Total
Atelectasis	7996	1119	2420	11559
Cardiomegaly	1950	240	582	2776
Effusion	9261	1292	2754	13317
Infiltration	13914	2018	3938	19894
Mass	3988	625	1133	5782
Nodule	4375	613	1335	6331
Pneumonia	978	133	242	1431
Pneumothorax	3705	504	1089	5302
Consolidation	3263	447	957	4667
Edema	1690	200	413	2303
Emphysema	1799	208	509	2516
Fibrosis	1158	166	362	1686
Pleural Thickening	2279	372	734	3385
Hernia	144	41	42	227
No Finding	42405	6079	11928	60361
Total	78468	11219	22433	112120

Table: For the NIH CXR-14 dataset, the number of images in the train, validation and test sets per abnormality

① Noisy labels

Labels were extracted using NLP from radiology reports in natural language text. This may lead to label noise. [7] show that these labels are about 90% accurate. Although deep neural networks have been shown to be robust to label noise in general[9], structured noise can be especially detrimental to performance.

① Noisy labels

Labels were extracted using NLP from radiology reports in natural language text. This may lead to label noise. [7] show that these labels are about 90% accurate. Although deep neural networks have been shown to be robust to label noise in general[9], structured noise can be especially detrimental to performance.

② Labeling schema

Some of these abnormalities are sub-types of others, but labels are provided as a non-hierarchical list. For example, *Pneumonia* on the x-ray is a form of *Consolidation*

① Noisy labels

Labels were extracted using NLP from radiology reports in natural language text. This may lead to label noise. [7] show that these labels are about 90% accurate. Although deep neural networks have been shown to be robust to label noise in general[9], structured noise can be especially detrimental to performance.

② Labeling schema

Some of these abnormalities are sub-types of others, but labels are provided as a non-hierarchical list. For example, *Pneumonia* on the x-ray is a form of *Consolidation*

③ Class imbalance

The majority of images do not contain abnormalities, and some abnormalities are common while others are rare. For example, *Infiltration* appears in 17% of the images while *Hernia* appears in only 0.2% of the images.

Guangzhou

Detecting
abnormali-
ties on
chest
X-rays
using deep
neural
networks

Swaroop
Kumar M
L
Department
of Studies
in
Computer
Science
University
of Mysore

Introduction

Data

Baselines

Experiments

Results

Examples

References

References

References

References

The Guangzhou pediatric pneumonia dataset consists of 8,497 chest x-ray images of children under the age of 5 years collected from the Guangzhou Women and Children's Medical Center, Guangzhou. Each image has been labeled as either *Normal* (4232 images) or *Pneumonia* (4265 images). Images labeled as *Pneumonia* have been further tagged as *Bacterial* (2772 images) or *Viral* (1493 images) based on the cause of pneumonia.

Guangzhou

Detecting
abnormali-
ties on
chest
X-rays
using deep
neural
networks

Swaroop
Kumar M
L
Department
of Studies
in
Computer
Science
University
of Mysore

Introduction
Data
Baselines
Experiments
Results
Examples
References
References
References
References
References

The Guangzhou pediatric pneumonia dataset consists of 8,497 chest x-ray images of children under the age of 5 years collected from the Guangzhou Women and Children's Medical Center, Guangzhou. Each image has been labeled as either *Normal* (4232 images) or *Pneumonia* (4265 images). Images labeled as *Pneumonia* have been further tagged as *Bacterial* (2772 images) or *Viral* (1493 images) based on the cause of pneumonia.

We use the standard test set and split the rest of the dataset into train and validation sets in the ratio 80:20. However, when used as an external dataset, we use the entire dataset.

Label	Number of images			
	Train	Validation	Test	Total
Normal	3199	799	234	4232
Viral pneumonia	1076	269	148	1493
Bacterial pneumonia	2024	506	242	2772
Total	6299	1574	624	8497

Table: For the Guangzhou pediatric pneumonia dataset, the number of images in the train, validation and test sets per label

Shenzhen

Detecting
abnormali-
ties on
chest
X-rays
using deep
neural
networks

Swaroop
Kumar M
L
Department
of Studies
in
Computer
Science
University
of Mysore

Introduction

Data

Baselines

Experiments

Results

Examples

References

References

References

References

References

The Shenzhen tuberculosis dataset consists of 615 chest x-ray images collected from the Shenzhen No.3 Hospital in Shenzhen, Guangdong providence, China. Each of the images is labeled as either *Normal* or *Tuberculosis*. 340 of these are normal and 275 show manifestations of tuberculosis.

Shenzhen

Detecting
abnormali-
ties on
chest
X-rays
using deep
neural
networks

Swaroop
Kumar M
L

Department
of Studies
in
Computer
Science
University
of Mysore

Introduction

Data

Baselines

Experiments

Results

Examples

References

References

References

References

The Shenzhen tuberculosis dataset consists of 615 chest x-ray images collected from the Shenzhen No.3 Hospital in Shenzhen, Guangdong providence, China. Each of the images is labeled as either *Normal* or *Tuberculosis*. 340 of these are normal and 275 show manifestations of tuberculosis.

Considering the small size of the dataset, we create 9 folds of the dataset and report average and standard deviation of metrics. Each fold contains all the images split into train, validation and test sets in the ratio 70:10:20.

Montgomery

Detecting abnormalities on chest X-rays using deep neural networks

Swaroop Kumar M L

Department of Studies in Computer Science University of Mysore

Introduction

Data

Baselines

Experiments

Results

Examples

References

References

References

References

The Montgomery county tuberculosis dataset consists of 138 chest x-ray images collected from the tuberculosis control program of the Department of Health and Human Services of Montgomery County, MD, USA. Each of the images is labeled as either *Normal* or *Tuberculosis*. 80 of these are normal and 58 show manifestations of tuberculosis. The dataset also contains manually segmented lung masks for each image.

Montgomery

Detecting
abnormali-
ties on
chest
X-rays
using deep
neural
networks

Swaroop
Kumar M
L

Department
of Studies
in
Computer
Science
University
of Mysore

Introduction

Data

Baselines

Experiments

Results

Examples

References

References

References

References

The Montgomery county tuberculosis dataset consists of 138 chest x-ray images collected from the tuberculosis control program of the Department of Health and Human Services of Montgomery County, MD, USA. Each of the images is labeled as either *Normal* or *Tuberculosis*. 80 of these are normal and 58 show manifestations of tuberculosis. The dataset also contains manually segmented lung masks for each image.

Similar to the Shenzhen dataset, we create 9 folds and report average and standard deviation of metrics. Each fold contains all the images split into train, validation and test sets in the ratio 70:10:20. However, when used as an external dataset, we ignore these folds and test on the entire dataset.

Model architecture

Detecting abnormalities on chest X-rays using deep neural networks

Swaroop Kumar M

L
Department
of Studies
in
Computer
Science
University
of Mysore

Introduction

Data

Baselines

Experiments

Results

Examples

References

References

References

References

We replace the final fully connected layer of 121-layer dense convolutional neural network with one that has either 14 outputs (for the NIH CXR-14 dataset) or 2 outputs (for all other datasets) after which we apply a sigmoid non-linearity.

Model architecture

Detecting abnormalities on chest X-rays using deep neural networks

Swaroop Kumar M

L
Department
of Studies
in
Computer
Science
University
of Mysore

Introduction

Data

Baselines

Experiments

Results

Examples

References

References

References

References

We replace the final fully connected layer of 121-layer dense convolutional neural network with one that has either 14 outputs (for the NIH CXR-14 dataset) or 2 outputs (for all other datasets) after which we apply a sigmoid non-linearity.

The fully convolutional backbone of the network results in $k w \times h$ feature maps and is followed by a global-average-pooling layer where the k feature maps are averaged along the width and height to form a k dimensional vector.

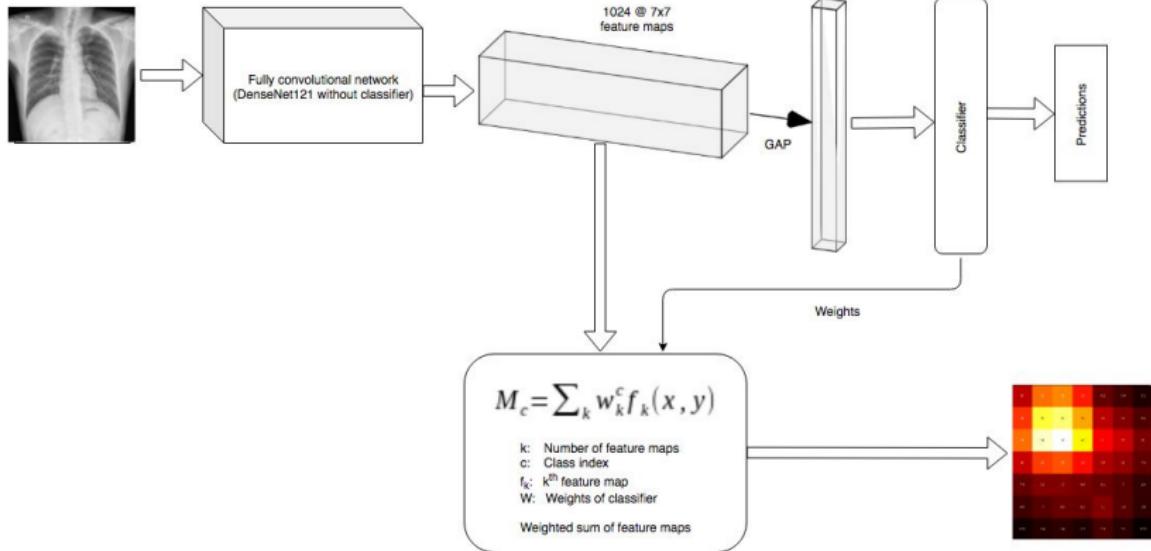


Figure: Basic architecture of the model

Inference

Detecting
abnormali-
ties on
chest
X-rays
using deep
neural
networks

Swaroop
Kumar M
L

Department
of Studies
in
Computer
Science
University
of Mysore

Introduction

Data

Baselines

Experiments

Results

Examples

References

References

References

References

At the multi-class multi-label classification task (for the NIH CXR-14 dataset), we merge the train and validation sets after training and use it to compute optimal thresholds for each abnormality by optimizing for the class-specific F1-score.

Inference

Detecting abnormalities on chest X-rays using deep neural networks

Swaroop Kumar M L

Department of Studies in Computer Science University of Mysore

Introduction

Data

Baselines

Experiments

Results

Examples

References

References

References

References

At the multi-class multi-label classification task (for the NIH CXR-14 dataset), we merge the train and validation sets after training and use it to compute optimal thresholds for each abnormality by optimizing for the class-specific F1-score.

Although it is possible to compute optimal thresholds for the classification task since we have ground truth labels, it is not possible to do so for the localization task since we only have weak labels and not precise location.

Inference

Detecting abnormalities on chest X-rays using deep neural networks

Swaroop Kumar M L

Department of Studies in Computer Science University of Mysore

Introduction

Data

Baselines

Experiments

Results

Examples

References

References

References

References

At the multi-class multi-label classification task (for the NIH CXR-14 dataset), we merge the train and validation sets after training and use it to compute optimal thresholds for each abnormality by optimizing for the class-specific F1-score.

Although it is possible to compute optimal thresholds for the classification task since we have ground truth labels, it is not possible to do so for the localization task since we only have weak labels and not precise location.

At binary classification tasks, since the network has two output nodes, we do not compute thresholds but simply consider as the proper output the class whose corresponding output node has higher activation.

Computing saliency maps

Detecting
abnormali-
ties on
chest
X-rays
using deep
neural
networks

Swaroop
Kumar M
L

Department
of Studies
in
Computer
Science
University
of Mysore

Introduction

Data

Baselines

Experiments

Results

Examples

References

References

References

References

To compute explanations, we save the feature maps resulting from the final convolutional layer during a forward pass and perform a weighted sum of these feature maps using the weights of the final fully-connected layer between each of the feature maps and the desired output node, as follows.

Computing saliency maps

Detecting abnormalities on chest X-rays using deep neural networks

Swaroop Kumar M L

Department of Studies in Computer Science University of Mysore

Introduction
Data

Baselines
Experiments
Results

Examples
References
References
References
References

To compute explanations, we save the feature maps resulting from the final convolutional layer during a forward pass and perform a weighted sum of these feature maps using the weights of the final fully-connected layer between each of the feature maps and the desired output node, as follows.

If f_i is the i^{th} feature map and w_i^j is the weight between the i^{th} input node and the j^{th} output node in the fully-connected layer, the saliency map for the j^{th} class M_j is

$$M_j = \sum_i w_i^j f_i \quad (2)$$

M_j is a $w \times h$ saliency map which we interpolate to the size of the input image and visualize as a heatmap

We use a region-growing algorithm to determine bounding boxes given a saliency map. Specifically, we first threshold the saliency map and using the maximum element as a seed point, grow a region around it, including all non-zero neighbours, and repeat the same until all non-zero elements are included in a region, each time choosing as seed the maximum element not included a region. For each region, we determine a bounding box as the smallest rectangle which encloses the entire region.

Detecting abnormalities on chest X-rays using deep neural networks

Swaroop Kumar M L

Department of Studies in Computer Science University of Mysore

Introduction

Data

Baselines

Experiments

Results

Examples

References

References

References

References

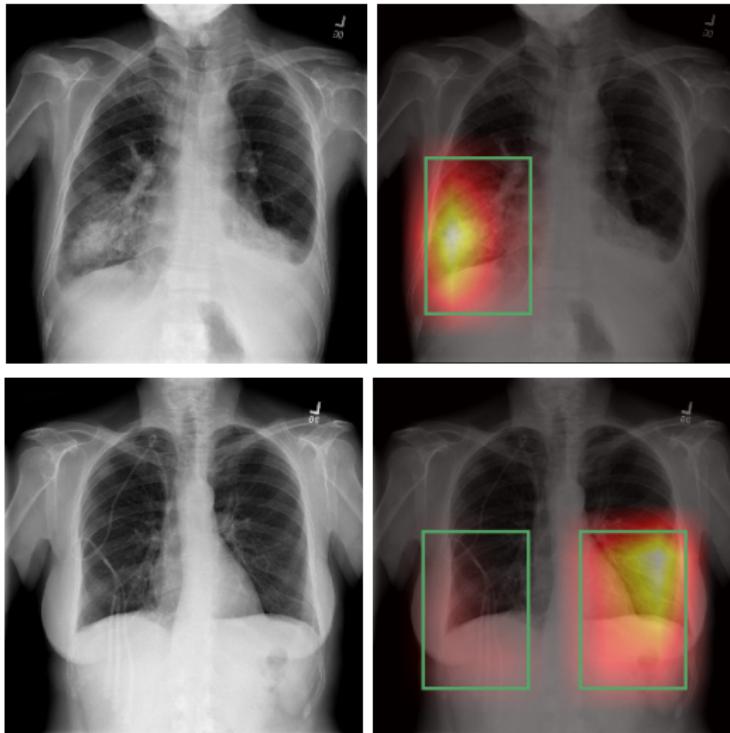


Figure: Examples of saliency maps with corresponding bounding boxes drawn. The first column shows original x-ray images and the second row shows saliency map and bounding boxes overlaid on the image. Row 1: Pneumonia. Row 2: Atelectasis.

Evaluation metrics

Detecting
abnormali-
ties on
chest
X-rays
using deep
neural
networks

Swaroop
Kumar M
L

Department
of Studies
in
Computer
Science
University
of Mysore

Introduction

Data

Baselines

Experiments

Results

Examples

References

References

References

References

We primarily use the area under the receiver-operator-characteristic curve (AUROC) to measure the performance of a model. AUROC is not affected by the class distribution, does not need thresholds to be set, and is commonly used in the literature. Apart from AUROC and accuracy, we also use

Evaluation metrics

Detecting abnormalities on chest X-rays using deep neural networks

Swaroop Kumar M L

Department of Studies in

Computer Science University of Mysore

Introduction

Data

Baselines

Experiments

Results

Examples

References

References

References

References

We primarily use the area under the receiver-operator-characteristic curve (AUROC) to measure the performance of a model. AUROC is not affected by the class distribution, does not need thresholds to be set, and is commonly used in the literature. Apart from AUROC and accuracy, we also use

① Specificity

$$\text{Specificity} = \frac{|TN|}{|TN| + |FP|} \quad (3)$$

② Sensitivity

$$\text{Sensitivity} = \frac{|TP|}{|TP| + |FN|} \quad (4)$$

Baseline results

Abnormality	AUROC
Atelectasis	0.823
Cardiomegaly	0.905
Effusion	0.879
Infiltration	0.712
Mass	0.839
Nodule	0.778
Pneumonia	0.760
Pneumothorax	0.869
Consolidation	0.806
Edema	0.891
Emphysema	0.923
Fibrosis	0.831
Pleural_Thickening	0.785
Hernia	0.929
Average	0.838

Table: Baseline results on the NIH CXR-14 dataset

	AUROC	Accuracy	Specificity	Sensitivity
Mean	0.956	0.902	0.902	0.899
Standard deviation	0.009	0.019	0.050	0.045

Table: Baseline results on the Shenzhen hospital tuberculosis dataset

Experiments

Detecting
abnormali-
ties on
chest
X-rays
using deep
neural
networks

Swaroop
Kumar M

L

Department
of Studies
in
Computer
Science
University
of Mysore

Introduction

Data

Baselines

Experiments

Results

Examples

References

References

References

References

① Data augmentation

Experiments

Detecting
abnormali-
ties on
chest
X-rays
using deep
neural
networks

Swaroop
Kumar M
L

Department
of Studies
in
Computer
Science
University
of Mysore

Introduction

Data

Baselines

Experiments

Results

Examples

References

References

References

References

- ➊ Data augmentation
- ➋ Test-time augmentation

Experiments

Detecting
abnormali-
ties on
chest
X-rays
using deep
neural
networks

Swaroop
Kumar M
L

Department
of Studies
in
Computer
Science
University
of Mysore

Introduction

Data

Baselines

Experiments

Results

Examples

References

References

References

References

- ➊ Data augmentation
- ➋ Test-time augmentation
- ➌ Mixup

Experiments

Detecting
abnormali-
ties on
chest
X-rays
using deep
neural
networks

Swaroop
Kumar M
L

Department
of Studies
in
Computer
Science
University
of Mysore

Introduction

Data

Baselines

Experiments

Results

Examples

References

References

References

References

- ① Data augmentation
- ② Test-time augmentation
- ③ Mixup
- ④ Transfer-learning from ImageNet

Experiments

Detecting
abnormali-
ties on
chest
X-rays
using deep
neural
networks

Swaroop
Kumar M
L

Department
of Studies
in
Computer
Science
University
of Mysore

Introduction

Data

Baselines

Experiments

Results

Examples

References

References

References

References

- ➊ Data augmentation
- ➋ Test-time augmentation
- ➌ Mixup
- ➍ Transfer-learning from ImageNet
- ➎ Transfer-learning from NIH CXR-14

Experiments

Detecting
abnormali-
ties on
chest
X-rays
using deep
neural
networks

Swaroop
Kumar M
L

Department
of Studies
in
Computer
Science
University
of Mysore

Introduction

Data

Baselines

Experiments

Results

Examples

References

References

References

References

- ① Data augmentation
- ② Test-time augmentation
- ③ Mixup
- ④ Transfer-learning from ImageNet
- ⑤ Transfer-learning from NIH CXR-14
- ⑥ Over-diagnosis of TB

Experiments

Detecting
abnormali-
ties on
chest
X-rays
using deep
neural
networks

Swaroop
Kumar M
L

Department
of Studies
in
Computer
Science
University
of Mysore

Introduction

Data

Baselines

Experiments

Results

Examples

References

References

References

References

- ① Data augmentation
- ② Test-time augmentation
- ③ Mixup
- ④ Transfer-learning from ImageNet
- ⑤ Transfer-learning from NIH CXR-14
- ⑥ Over-diagnosis of TB
- ⑦ Progressive resizing

Experiments

Detecting
abnormali-
ties on
chest
X-rays
using deep
neural
networks

Swaroop
Kumar M
L

Department
of Studies
in
Computer
Science
University
of Mysore

Introduction

Data

Baselines

Experiments

Results

Examples

References

References

References

References

- ➊ Data augmentation
- ➋ Test-time augmentation
- ➌ Mixup
- ➍ Transfer-learning from ImageNet
- ➎ Transfer-learning from NIH CXR-14
- ➏ Over-diagnosis of TB
- ➐ Progressive resizing
- ➑ Ensembling predictions

Experiments

Detecting
abnormali-
ties on
chest
X-rays
using deep
neural
networks

Swaroop
Kumar M
L

Department
of Studies
in
Computer
Science
University
of Mysore

Introduction

Data

Baselines

Experiments

Results

Examples

References

References

References

References

- ➊ Data augmentation
- ➋ Test-time augmentation
- ➌ Mixup
- ➍ Transfer-learning from ImageNet
- ➎ Transfer-learning from NIH CXR-14
- ➏ Over-diagnosis of TB
- ➐ Progressive resizing
- ➑ Ensembling predictions
- ➒ Ensembling saliency maps

Experiments

Detecting
abnormali-
ties on
chest
X-rays
using deep
neural
networks

Swaroop
Kumar M
L

Department
of Studies
in
Computer
Science
University
of Mysore

Introduction

Data

Baselines

Experiments

Results

Examples

References

References

References

References

- ① Data augmentation
- ② Test-time augmentation
- ③ Mixup
- ④ Transfer-learning from ImageNet
- ⑤ Transfer-learning from NIH CXR-14
- ⑥ Over-diagnosis of TB
- ⑦ Progressive resizing
- ⑧ Ensembling predictions
- ⑨ Ensembling saliency maps
- ⑩ Cropping of image margins

Experiments

Detecting
abnormali-
ties on
chest
X-rays
using deep
neural
networks

Swaroop
Kumar M
L

Department
of Studies
in
Computer
Science
University
of Mysore

Introduction

Data

Baselines

Experiments

Results

Examples

References

References

References

References

- ① Data augmentation
- ② Test-time augmentation
- ③ Mixup
- ④ Transfer-learning from ImageNet
- ⑤ Transfer-learning from NIH CXR-14
- ⑥ Over-diagnosis of TB
- ⑦ Progressive resizing
- ⑧ Ensembling predictions
- ⑨ Ensembling saliency maps
- ⑩ Cropping of image margins
- ⑪ Fairness

Experiments

Detecting
abnormali-
ties on
chest
X-rays
using deep
neural
networks

Swaroop
Kumar M
L

Department
of Studies
in
Computer
Science
University
of Mysore

Introduction

Data

Baselines

Experiments

Results

Examples

References

References

References

References

- ➊ Data augmentation
- ➋ Test-time augmentation
- ➌ Mixup
- ➍ Transfer-learning from ImageNet
- ➎ Transfer-learning from NIH CXR-14
- ➏ Over-diagnosis of TB
- ➐ Progressive resizing
- ➑ Ensembling predictions
- ➒ Ensembling saliency maps
- ➓ Cropping of image margins
- ➔ Fairness
- ➕ Generalization

Experiments

Detecting
abnormali-
ties on
chest
X-rays
using deep
neural
networks

Swaroop
Kumar M
L

Department
of Studies
in
Computer
Science
University
of Mysore

Introduction

Data

Baselines

Experiments

Results

Examples

References

References

References

References

- ➊ Data augmentation
- ➋ Test-time augmentation
- ➌ Mixup
- ➍ Transfer-learning from ImageNet
- ➎ Transfer-learning from NIH CXR-14
- ➏ Over-diagnosis of TB
- ➐ Progressive resizing
- ➑ Ensembling predictions
- ➒ Ensembling saliency maps
- ➓ Cropping of image margins
- ➔ Fairness
- ➕ Generalization

Comparison to previous work on NIH CXR-14

Detecting abnormalities on chest X-rays using deep neural networks

Swaroop Kumar M L
Department of Studies in Computer Science University of Mysore

Introduction

Data

Baselines

Experiments

Results

Examples

References

References

References

References

Authors	Average AUROC
Wang et al. (2017)	0.738
Y. Shen et al.	0.775
H. Wang et al. (ChestNet)	0.781
P. Kumar et al.	0.792
Yao et al. (2017)	0.803
Y. Tang et al.	0.805
S. Guendel et al.	0.807
Yan et al.	0.83
X. Xu et al. (DeepCXray)	0.832
Rajpurkar et al. (CheXNet)	0.841
B. Zhou et al.	0.842
Rajpurkar et al. (ChexNext)	0.849
Our model	0.856
Q. Guan et al.	0.871

Table: Comparison to previous work on the NIH CXR-14 dataset

Comparison to human radiologists

Detecting abnormalities on chest X-rays using deep neural networks

Swaroop Kumar M L

Department of Studies in Computer Science University of Mysore

Introduction

Data

Baselines

Experiments

Results

Examples

References

References

References

References

Rajpurkar et al. in [10] measured human performance in terms of AUROC for each disease, using the majority vote of 3 independent board-certified cardiothoracic specialist radiologists (average experience 15 years) as ground truth, and measure the the performance of 6 BC radiologists from 3 academic institutions (average experience 12 years) and 3 senior radiology residents by fitting a curve to these 9 radiologists' operating points and calculating the area under it.

Abnormality	AUROC			
	Baseline	Ensemble	Radiologist	Difference (%)
Atelectasis	0.823	0.839	0.808	-3.06
Cardiomegaly	0.899	0.916	0.888	-2.79
Effusion	0.881	0.89	0.9	0.96
Infiltration	0.705	0.72	0.734	1.39
Mass	0.857	0.868	0.886	1.76
Nodule	0.779	0.817	0.899	8.17
Pneumonia	0.767	0.765	0.823	5.83
Pneumothorax	0.881	0.895	0.94	4.46
Consolidation	0.822	0.819	0.841	2.22
Edema	0.911	0.902	0.91	0.83
Emphysema	0.913	0.944	0.911	-3.33
Fibrosis	0.824	0.854	0.897	4.31
Pleural_Thickening	0.81	0.805	0.779	-2.59
Hernia	0.906	0.944	0.985	4.1
Average	0.841	0.856	0.8715	1.59

Table: Comparison to human radiologists on the NIH CXR-14 dataset

Caveats with comparison to human radiologists

Detecting abnormalities on chest X-rays using deep neural networks

Swaroop Kumar M L

Department of Studies in Computer Science University of Mysore

Introduction

Data

Baselines

Experiments

Results

Examples

References

References

References

References

Overestimating model performance

- ① Noisy labels. Model may be overfitting to noise.
- ② Model may be overfitting to certain hospitals and machines.
- ③ Model may be exploiting spurious correlations such as using presence of chest-drains to infer Pneumothorax.

Caveats with comparison to human radiologists

Detecting abnormalities on chest X-rays using deep neural networks

Swaroop Kumar M L

Department of Studies in Computer Science University of Mysore

Introduction

Data

Baselines

Experiments

Results

Examples

References

References

References

References

Overestimating model performance

- ① Noisy labels. Model may be overfitting to noise.
- ② Model may be overfitting to certain hospitals and machines.
- ③ Model may be exploiting spurious correlations such as using presence of chest-drains to infer Pneumothorax.

Underestimating human performance

- ① Unconventional labelling schema. Radiologists usually do not diagnose many of these abnormalities from x-rays. They have access to other clinical data
- ② Radiologists work with higher resolution images with more grey-levels than the dataset provides

Comparison to previous work on Shenzhen

Detecting abnormalities on chest X-rays using deep neural networks

Swaroop Kumar M L
Department of Studies in Computer Science University of Mysore

Introduction

Data

Baselines

Experiments

Results

Examples

References

References

References

References

Authors	AUROC	Accuracy
Jaeger et al	0.9	0.841
Hwang et al	0.93	0.837
Lopez and Valiati	0.926	0.846
MT Islam et al	0.94	0.9
Haloi et al	0.949	
Liu et al (ResNet-152)	0.967	0.923
Liu et al (Inception-ResNet-v2)	0.983	0.917
Vajda et al	0.99	0.957
Our baseline	0.956	0.902
Our best model		
Pretrained on NIH CXR-14 with mixup $\alpha = 0.4$	0.985	0.949

Table: Comparison to previous work on the Shenzhen tuberculosis dataset

Comparison to previous work on Montgomery

Detecting abnormalities on chest X-rays using deep neural networks

Swaroop Kumar M L

Department of Studies in Computer Science University of Mysore

Introduction

Data

Baselines

Experiments

Results

Examples

References

References

References

References

Authors	AUROC	Accuracy
Jaeger et al	0.869	0.783
Lopez and Valiati	0.926	0.826
Liu et al (Inception-ResNet-v2)	0.957	0.844
Liu et al (ResNet-152)	0.951	0.890
Vajda et al	0.870	0.783
Our baseline	0.871	0.755
Our best model		
Pre-trained on NIH CXR-14 (480 x 480)	0.957	0.89

Table: Comparison to previous work on the Montgomery tuberculosis dataset

Examples

Detecting abnormalities on chest X-rays using deep neural networks

Swaroop Kumar M L

Department of Studies in Computer Science University of Mysore

Introduction

Data

Baselines

Experiments

Results

Examples

References

References

References

References

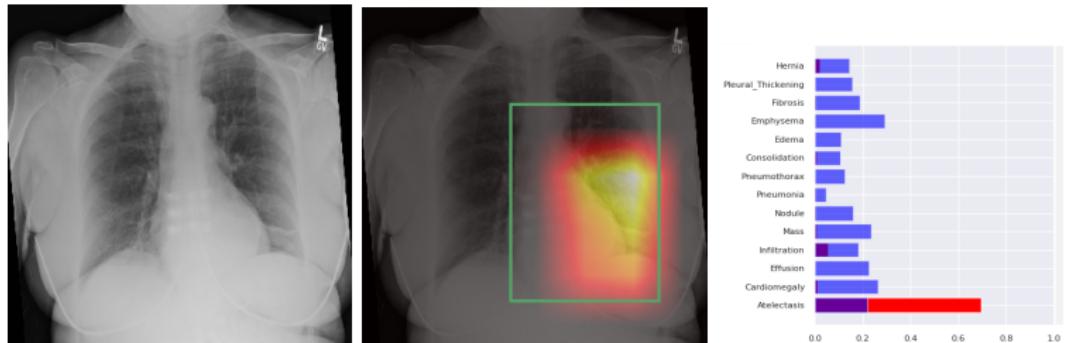


Figure: Original image, image overlaid with saliency map and bounding boxes for *Atelectasis*, and predicted probabilities for an x-ray image.

Detecting abnormalities on chest X-rays using deep neural networks

Swaroop Kumar M L

Department of Studies in Computer Science University of Mysore

Introduction

Data

Baselines

Experiments

Results

Examples

References

References

References

References

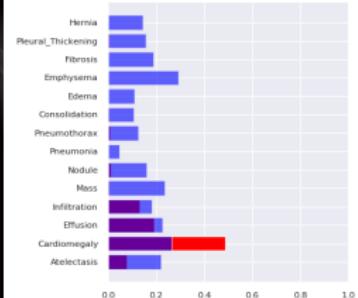
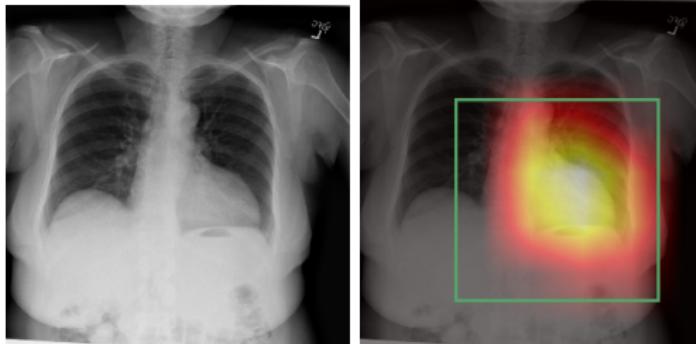


Figure: Original image, image overlaid with saliency map and bounding boxes for *Cardiomegaly*, and predicted probabilities for an x-ray image.

Detecting abnormalities on chest X-rays using deep neural networks

Swaroop Kumar M L

Department of Studies in Computer Science University of Mysore

Introduction

Data

Baselines

Experiments

Results

Examples

References

References

References

References

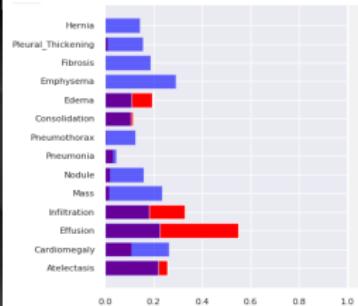
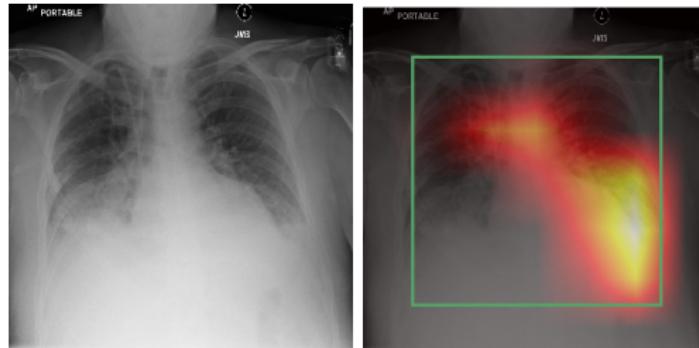


Figure: Original image, image overlaid with saliency map and bounding boxes for *Effusion*, and predicted probabilities for an x-ray image.

Detecting abnormalities on chest X-rays using deep neural networks

Swaroop Kumar M L

Department of Studies in Computer Science University of Mysore

Introduction

Data

Baselines

Experiments

Results

Examples

References

References

References

References

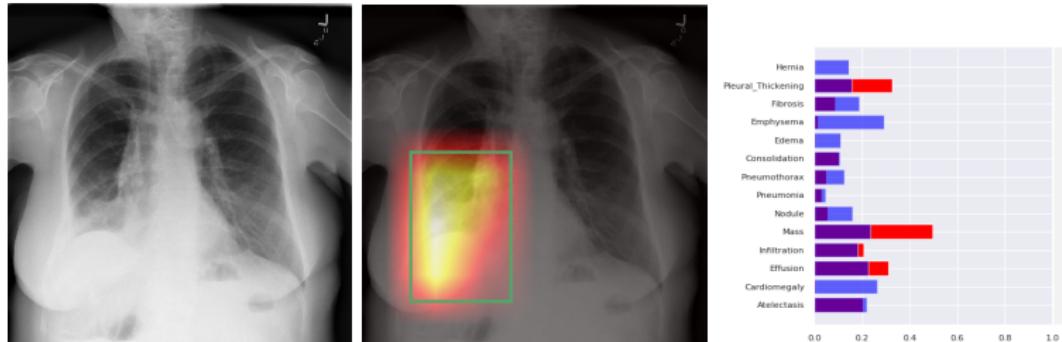


Figure: Original image, image overlaid with saliency map and bounding boxes for *Infiltration*, and predicted probabilities for an x-ray image.

Detecting abnormalities on chest X-rays using deep neural networks

Swaroop Kumar M L

Department of Studies in Computer Science University of Mysore

Introduction

Data

Baselines

Experiments

Results

Examples

References

References

References

References

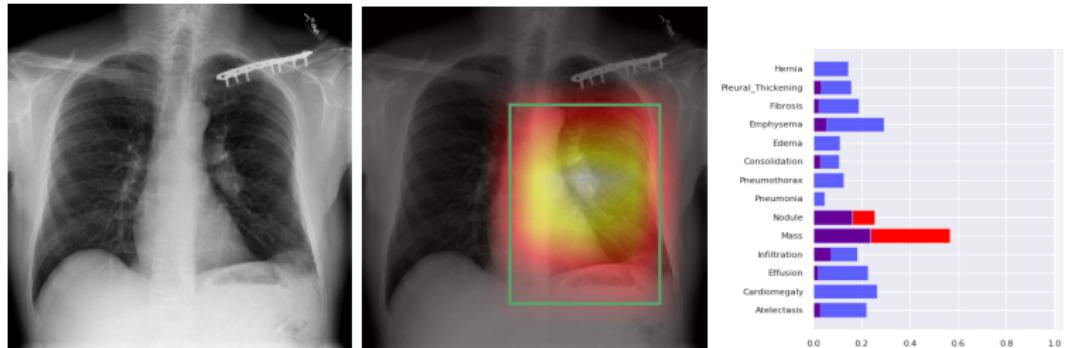


Figure: Original image, image overlaid with saliency map and bounding boxes for *Mass*, and predicted probabilities for an x-ray image.

Detecting abnormalities on chest X-rays using deep neural networks

Swaroop Kumar M L

Department of Studies in Computer Science University of Mysore

Introduction

Data

Baselines

Experiments

Results

Examples

References

References

References

References

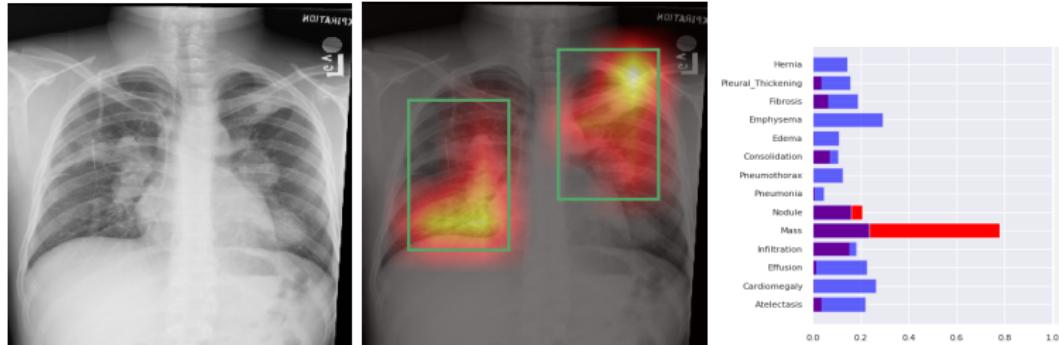


Figure: Original image, image overlaid with saliency map and bounding boxes for *Nodule*, and predicted probabilities for an x-ray image.

Detecting abnormalities on chest X-rays using deep neural networks

Swaroop Kumar M L

Department of Studies in Computer Science University of Mysore

Introduction

Data

Baselines

Experiments

Results

Examples

References

References

References

References

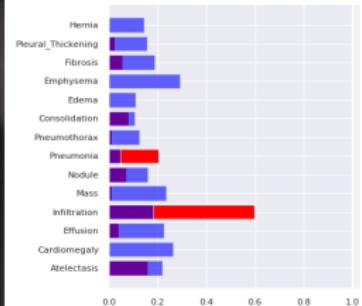
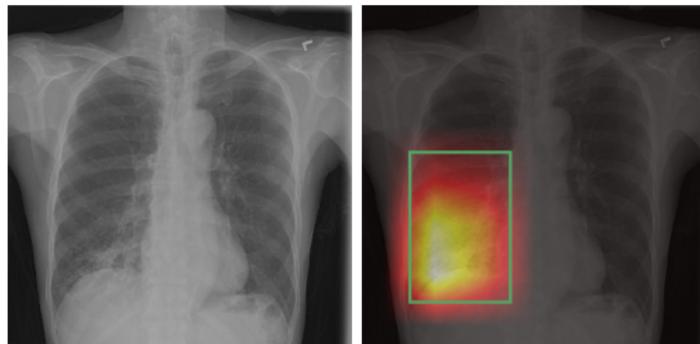


Figure: Original image, image overlaid with saliency map and bounding boxes for *Pneumonia*, and predicted probabilities for an x-ray image.

Detecting abnormalities on chest X-rays using deep neural networks

Swaroop Kumar M L

Department of Studies in Computer Science University of Mysore

Introduction

Data

Baselines

Experiments

Results

Examples

References

References

References

References

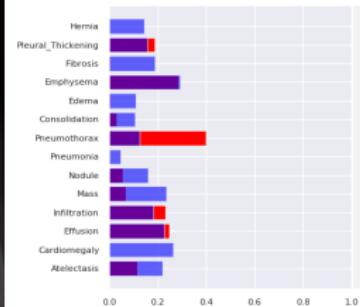
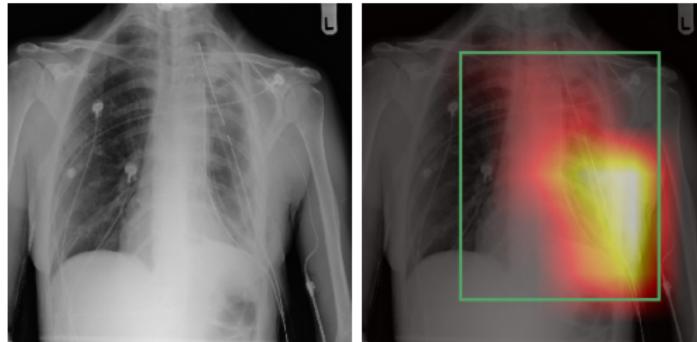


Figure: Original image, image overlaid with saliency map and bounding boxes for *Pneumothorax*, and predicted probabilities for an x-ray image.

Detecting abnormalities on chest X-rays using deep neural networks

Swaroop Kumar M L

Department of Studies in Computer Science University of Mysore

Introduction

Data

Baselines

Experiments

Results

Examples

References

References

References

References

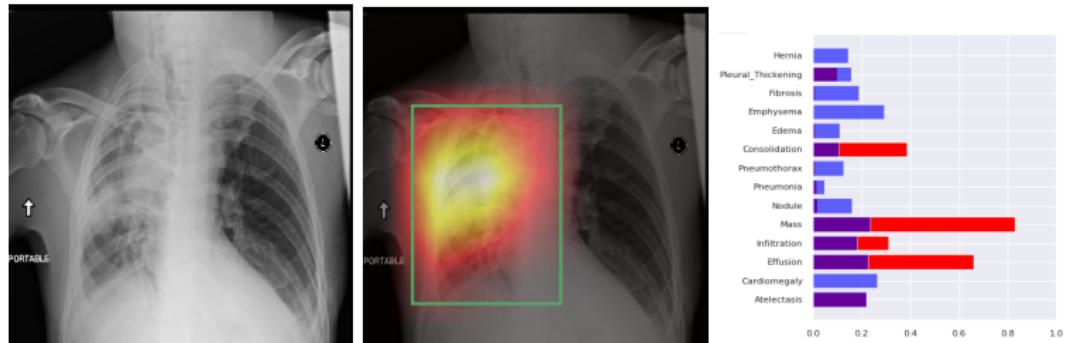


Figure: Original image, image overlaid with saliency map and bounding boxes for *Consolidation*, and predicted probabilities for an x-ray image.

Detecting abnormalities on chest X-rays using deep neural networks

Swaroop Kumar M L

Department of Studies in Computer Science University of Mysore

Introduction

Data

Baselines

Experiments

Results

Examples

References

References

References

References

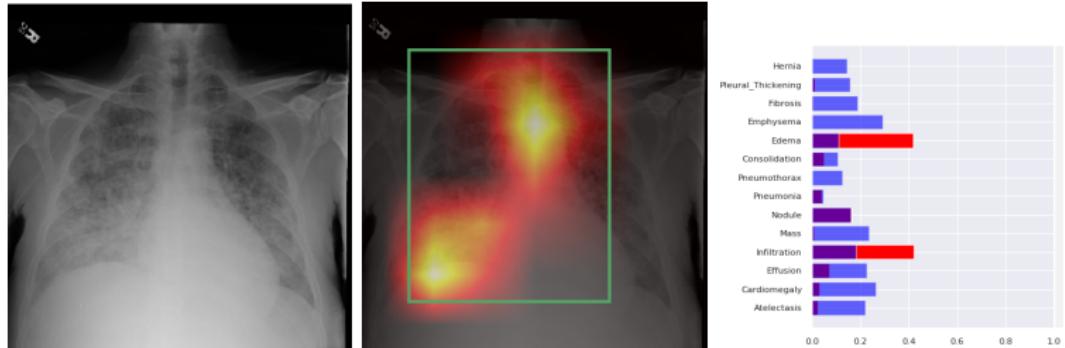


Figure: Original image, image overlaid with saliency map and bounding boxes for *Edema*, and predicted probabilities for an x-ray image.

Detecting abnormalities on chest X-rays using deep neural networks

Swaroop Kumar M L

Department of Studies in Computer Science University of Mysore

Introduction

Data

Baselines

Experiments

Results

Examples

References

References

References

References

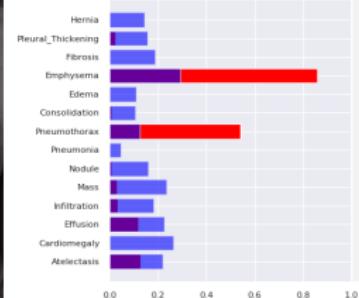
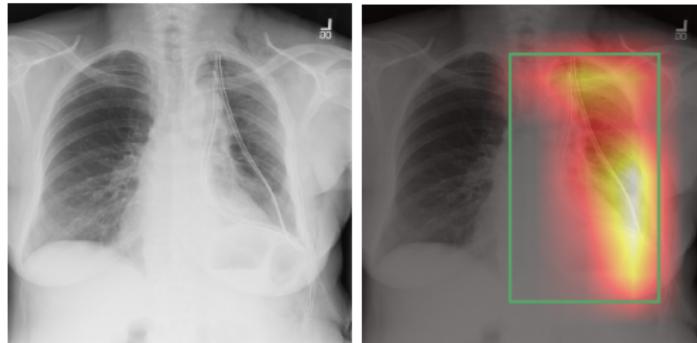


Figure: Original image, image overlaid with saliency map and bounding boxes for *Emphysema*, and predicted probabilities for an x-ray image.

Detecting abnormalities on chest X-rays using deep neural networks

Swaroop Kumar M L

Department of Studies in Computer Science University of Mysore

Introduction

Data

Baselines

Experiments

Results

Examples

References

References

References

References

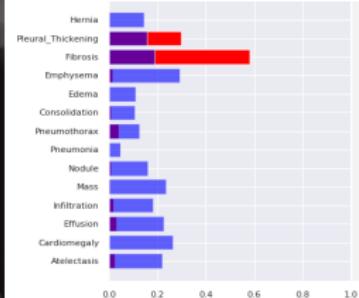
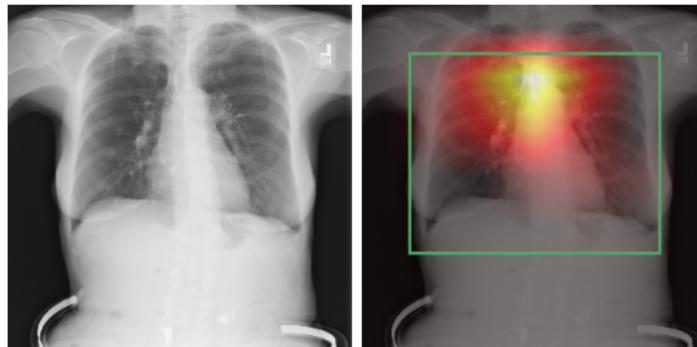


Figure: Original image, image overlaid with saliency map and bounding boxes for *Fibrosis*, and predicted probabilities for an x-ray image.

Detecting abnormalities on chest X-rays using deep neural networks

Swaroop Kumar M

L

Department of Studies in Computer Science University of Mysore

Introduction

Data

Baselines

Experiments

Results

Examples

References

References

References

References

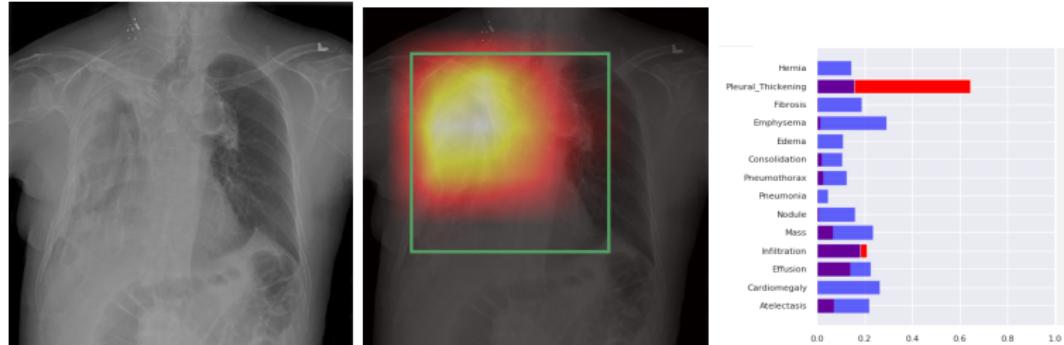


Figure: Original image, image overlaid with saliency map and bounding boxes for *Pleural Thickening*, and predicted probabilities for an x-ray image.

Detecting abnormalities on chest X-rays using deep neural networks

Swaroop Kumar M L

Department of Studies in Computer Science University of Mysore

Introduction

Data

Baselines

Experiments

Results

Examples

References

References

References

References

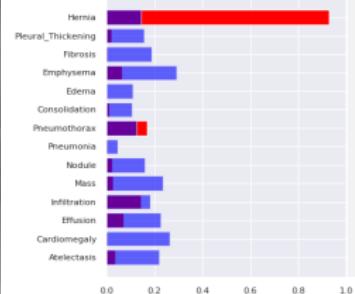
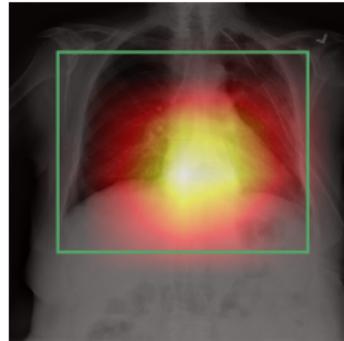


Figure: Original image, image overlaid with saliency map and bounding boxes for *Hernia*, and predicted probabilities for an x-ray image.

Detecting abnormalities on chest X-rays using deep neural networks

Swaroop Kumar M L

Department of Studies in Computer Science University of Mysore

Introduction

Data

Baselines

Experiments

Results

Examples

References

References

References

References

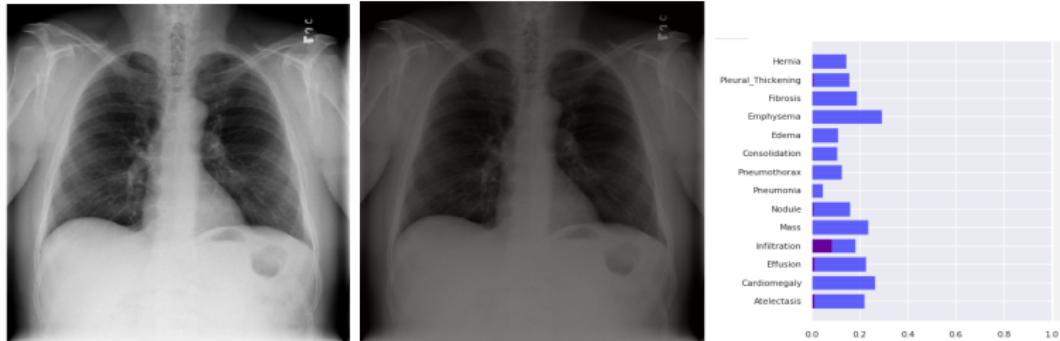


Figure: Original image, image overlaid with saliency map and bounding boxes, and predicted probabilities for an x-ray image showing no abnormalities.

Detecting abnormalities on chest X-rays using deep neural networks

Swaroop Kumar M L

Department of Studies in Computer Science University of Mysore

Introduction

Data

Baselines

Experiments

Results

Examples

References

References

References

References

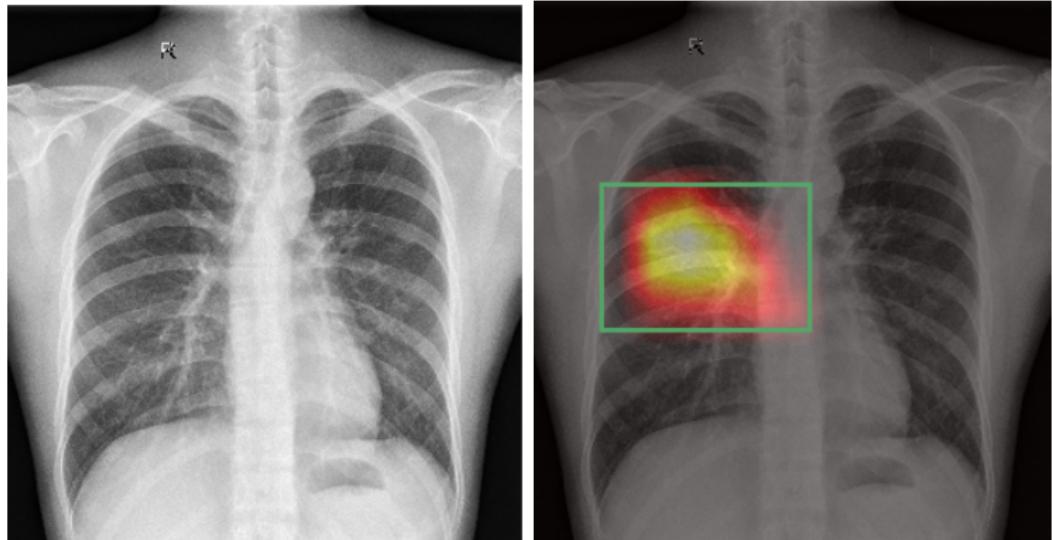


Figure: Original image, image overlaid with saliency map and bounding boxes for *Tuberculosis*

Detecting
abnormali-
ties on
chest
X-rays
using deep
neural
networks

Swaroop
Kumar M
L

Department
of Studies
in
Computer
Science
University
of Mysore

Introduction

Data

Baselines

Experiments

Results

Examples

References

References

References

References

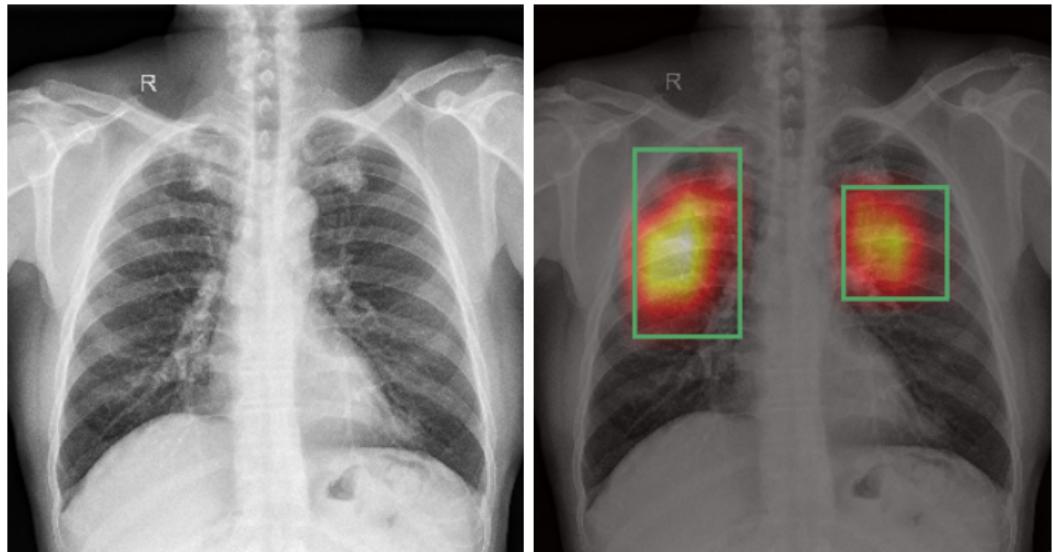


Figure: Original image, image overlaid with saliency map and bounding boxes for *Tuberculosis*

Prototype GUI

Detecting abnormalities on chest X-rays using deep neural networks

Swaroop Kumar M

L

Department of Studies in Computer Science University of Mysore

Introduction

Data

Baselines

Experiments

Results

Examples

References

References

References

References

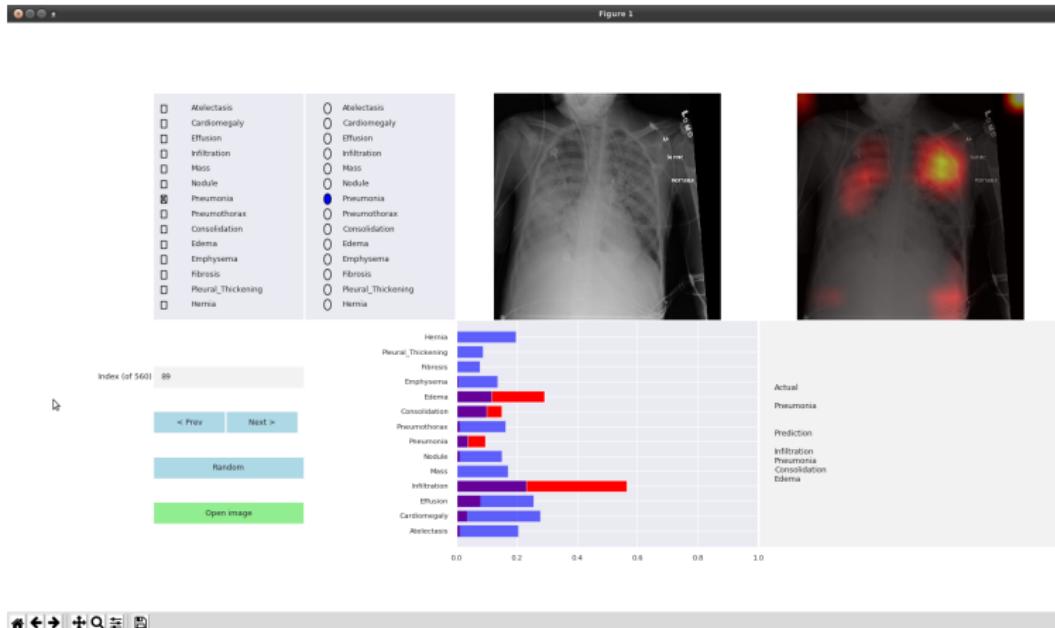


Figure: A simple graphical user interface we created to let radiologists qualitatively evaluate predictions.

Future enhancements

Detecting abnormalities on chest X-rays using deep neural networks

Swaroop Kumar M L
Department of Studies in Computer Science University of Mysore

Introduction

Data

Baselines

Experiments

Results

Examples

References

References

References

References

- ① Training on images from multiple hospital systems to improve the model's ability to generalize to other hospital systems and machine types, perhaps using techniques in domain adaptation or deep domain confusion [11] to prevent the model from learning features that are necessary to identify the hospital system.

Future enhancements

Detecting abnormalities on chest X-rays using deep neural networks

Swaroop Kumar M L
Department of Studies in Computer Science University of Mysore

Introduction

Data

Baselines

Experiments

Results

Examples

References

References

References

References

- ① Training on images from multiple hospital systems to improve the model's ability to generalize to other hospital systems and machine types, perhaps using techniques in domain adaptation or deep domain confusion [11] to prevent the model from learning features that are necessary to identify the hospital system.
- ② Exploring ways to infuse domain knowledge into the algorithm to exploit correlations between the abnormalities.

Future enhancements

Detecting abnormalities on chest X-rays using deep neural networks

Swaroop Kumar M L
Department of Studies in Computer Science University of Mysore

Introduction

Data

Baselines

Experiments

Results

Examples

References

References

References

References

- ① Training on images from multiple hospital systems to improve the model's ability to generalize to other hospital systems and machine types, perhaps using techniques in domain adaptation or deep domain confusion [11] to prevent the model from learning features that are necessary to identify the hospital system.
- ② Exploring ways to infuse domain knowledge into the algorithm to exploit correlations between the abnormalities.
- ③ Using attention to allow the network to focus on pathological areas.

Future enhancements

Detecting abnormalities on chest X-rays using deep neural networks

Swaroop Kumar M L
Department of Studies in Computer Science University of Mysore

Introduction

Data

Baselines

Experiments

Results

Examples

References

References

References

References

- ① Training on images from multiple hospital systems to improve the model's ability to generalize to other hospital systems and machine types, perhaps using techniques in domain adaptation or deep domain confusion [11] to prevent the model from learning features that are necessary to identify the hospital system.
- ② Exploring ways to infuse domain knowledge into the algorithm to exploit correlations between the abnormalities.
- ③ Using attention to allow the network to focus on pathological areas.
- ④ Training models end-to-end to generate radiology reports in natural language from x-ray images.

Future enhancements

- ① Training on images from multiple hospital systems to improve the model's ability to generalize to other hospital systems and machine types, perhaps using techniques in domain adaptation or deep domain confusion [11] to prevent the model from learning features that are necessary to identify the hospital system.
- ② Exploring ways to infuse domain knowledge into the algorithm to exploit correlations between the abnormalities.
- ③ Using attention to allow the network to focus on pathological areas.
- ④ Training models end-to-end to generate radiology reports in natural language from x-ray images.
- ⑤ Segmentation of lung regions and bone shadow suppression to reduce the number of false-positives.

Detecting abnormalities on chest X-rays using deep neural networks

Swaroop Kumar M L
Department of Studies in Computer Science University of Mysore

Introduction

Data

Baselines

Experiments

Results

Examples

References

References

References

References

References

Future enhancements

- ① Training on images from multiple hospital systems to improve the model's ability to generalize to other hospital systems and machine types, perhaps using techniques in domain adaptation or deep domain confusion [11] to prevent the model from learning features that are necessary to identify the hospital system.
- ② Exploring ways to infuse domain knowledge into the algorithm to exploit correlations between the abnormalities.
- ③ Using attention to allow the network to focus on pathological areas.
- ④ Training models end-to-end to generate radiology reports in natural language from x-ray images.
- ⑤ Segmentation of lung regions and bone shadow suppression to reduce the number of false-positives.
- ⑥ Using recently released datasets such as CheXPert[12] and PadChest[13] which have more images and a hierarchical labeling schema.

Future enhancements

- ① Training on images from multiple hospital systems to improve the model's ability to generalize to other hospital systems and machine types, perhaps using techniques in domain adaptation or deep domain confusion [11] to prevent the model from learning features that are necessary to identify the hospital system.
- ② Exploring ways to infuse domain knowledge into the algorithm to exploit correlations between the abnormalities.
- ③ Using attention to allow the network to focus on pathological areas.
- ④ Training models end-to-end to generate radiology reports in natural language from x-ray images.
- ⑤ Segmentation of lung regions and bone shadow suppression to reduce the number of false-positives.
- ⑥ Using recently released datasets such as CheXPert[12] and PadChest[13] which have more images and a hierarchical labeling schema.

References |

Detecting abnormalities on chest X-rays using deep neural networks

Swaroop Kumar M L
Department of Studies in Computer Science University of Mysore

Introduction

Data

Baselines

Experiments

Results

Examples

References

References

References

References



A. Modi and R. Suresh, *Scaling up tb screening with ai: Deploying automated x-ray screening in remote regions*, Apr. 2019. [Online]. Available: <http://blog.qure.ai/notes/scaling-up-tb-screening-with-ai>.



X. Wang, Y. Peng, L. Lu, Z. Lu, M. Bagheri, and R. M. Summers, *ChestX-ray8: Hospital-Scale Chest X-Ray Database and Benchmarks on Weakly-Supervised Classification and Localization of Common Thorax Diseases*, 2017.



S. Jaeger, S. Candemir, S. Antani, Y.-X. J. Wang, P.-X. Lu, and G. Thoma, “Two public chest x-ray datasets for computer-aided screening of pulmonary diseases,” *Quantitative imaging in medicine and surgery*, vol. 4, no. 6, p. 475, 2014.



J. R. Zech, M. A. Badgeley, M. Liu, A. B. Costa, J. J. Titano, and E. K. Oermann, “Variable generalization performance of a deep learning model to detect pneumonia in chest radiographs: A cross-sectional study,” *PLoS medicine*, vol. 15, no. 11, e1002683, 2018.



J. Buolamwini and T. Gebru, “Gender shades: Intersectional accuracy disparities in commercial gender classification,” in *Conference on Fairness, Accountability and Transparency*, 2018, pp. 77–91.

References II

Detecting abnormalities on chest X-rays using deep neural networks

Swaroop Kumar M L

Department of Studies in Computer Science University of Mysore

Introduction

Data

Baselines

Experiments

Results

Examples

References

References

References

References



J. Dastin, *Insight - amazon scraps secret ai recruiting tool that showed bias...* Oct. 2018. [Online]. Available: <https://in.reuters.com/article/amazon-com-jobs-automation/insight-amazon-scaps-secret-ai-recruiting-tool-that-showed-bias-against-women-idINKCN1MK0AH>.



X. Wang, Y. Peng, L. Lu, Z. Lu, M. Bagheri, and R. M. Summers, "ChestX-ray8: Hospital-scale Chest X-ray Database and Benchmarks on Weakly-Supervised Classification and Localization of Common Thorax Diseases," May 2017. DOI: 10.1109/CVPR.2017.369. arXiv: 1705.02315. [Online]. Available: <http://arxiv.org/abs/1705.02315%20http://dx.doi.org/10.1109/CVPR.2017.369>.



D. S. Kermany, M. Goldbaum, W. Cai, C. C. Valentim, H. Liang, S. L. Baxter, A. McKeown, G. Yang, X. Wu, F. Yan, et al., "Identifying medical diagnoses and treatable diseases by image-based deep learning," *Cell*, vol. 172, no. 5, pp. 1122–1131, 2018.



D. Rolnick, A. Veit, S. Belongie, and N. Shavit, "Deep learning is robust to massive label noise," *arXiv preprint arXiv:1705.10694*, 2017.

References III

Detecting abnormalities on chest X-rays using deep neural networks

Swaroop Kumar M L

Department of Studies in Computer Science University of Mysore

Introduction

Data

Baselines

Experiments

Results

Examples

References

References

References

References



P. Rajpurkar, J. Irvin, R. L. Ball, K. Zhu, B. Yang, H. Mehta, T. Duan, D. Ding, A. Bagul, C. P. Langlotz, B. N. Patel, K. W. Yeom, K. Shpanskaya, F. G. Blankenberg, J. Seekins, T. J. Amrhein, D. A. Mong, S. S. Halabi, E. J. Zucker, A. Y. Ng, and M. P. Lungren, "Deep learning for chest radiograph diagnosis: A retrospective comparison of the CheXNeXt algorithm to practicing radiologists," *PLoS Medicine*, vol. 15, no. 11, A. Sheikh, Ed., e1002686, Nov. 2018, ISSN: 15491676. DOI: 10.1371/journal.pmed.1002686. [Online]. Available: <http://dx.plos.org/10.1371/journal.pmed.1002686>.



E. Tzeng, J. Hoffman, N. Zhang, K. Saenko, and T. Darrell, "Deep domain confusion: Maximizing for domain invariance," *arXiv preprint arXiv:1412.3474*, 2014.



J. Irvin, P. Rajpurkar, M. Ko, Y. Yu, S. Ciurea-IIcus, C. Chute, H. Marklund, B. Haghgoo, R. Ball, K. Shpanskaya, J. Seekins, D. A. Mong, S. S. Halabi, J. K. Sandberg, R. Jones, D. B. Larson, C. P. Langlotz, B. N. Patel, M. P. Lungren, and A. Y. Ng, "CheXpert: A Large Chest Radiograph Dataset with Uncertainty Labels and Expert Comparison," Jan. 2019. arXiv: 1901.07031. [Online]. Available: <http://arxiv.org/abs/1901.07031>.



A. Bustos, A. Pertusa, J.-M. Salinas, and M. de la Iglesia-Vayá, "PadChest: A large chest x-ray image dataset with multi-label annotated reports," Jan. 2019. arXiv: 1901.07441. [Online]. Available: <http://arxiv.org/abs/1901.07441>.