

Noisy한 메디컬 이미지의 이해를 위한 최신 딥러닝 모델 적용에 관한 연구

전자공학과 201820963 나두연
소프트웨어학과 202020809 백종호
소프트웨어학과 201920779 양태규

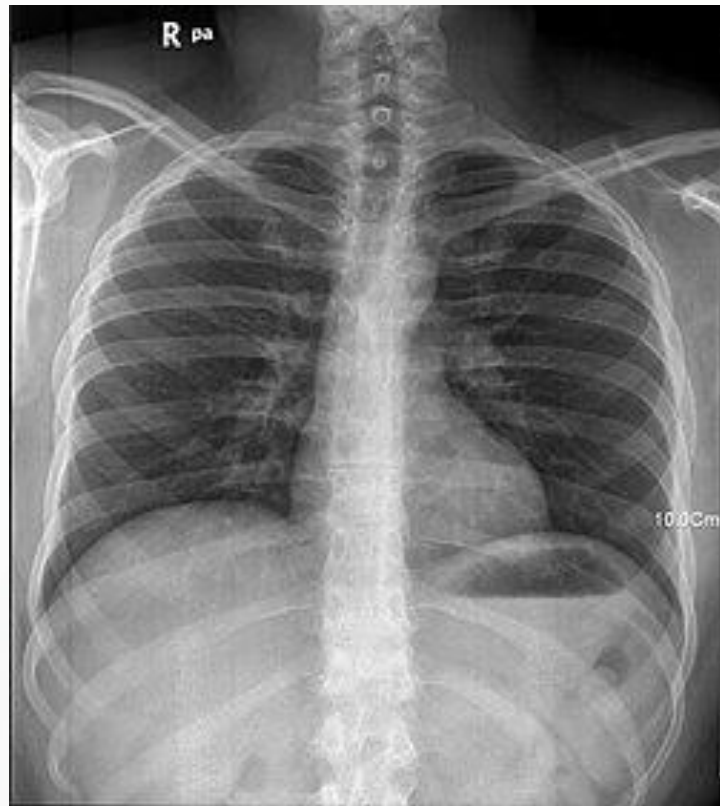


Table of contents

01

연구 배경 및 목표

02

연구 개요

03

연구 내용

04

결론

연구 배경 및 목표

We need Clean Data for Deep Learning

- Advantages of Deep Learning!
 - 의료 영상 및 의학 분야에서 질병 분류와 치료 결과 예측과 같은 다양한 분야에서 활용
- Limited Labeled Data in Medical Field
 - 전문가를 통해 의료 이미지 데이터 셋에 주석을 달거나 자동 시스템 등을 활용한 대량의 데이터 셋에 레이블을 지정

=> Cost effective X

What we want.....

- *“Noisy Label Medical Image dataset”*
 - Robustness to Noisy Label
- Great Performance



02

연구 개요

CheXpert

Dataset

Training Set: 223,414

Validation Set: 234

Test Set: 668

14 Labels

Train data distribution					
	label	Nan	-1	0	1
0	No Finding	201033	0	0	22381
1	Enlarged Cardiomeastinum	178575	12403	21638	10798
2	Cardiomegaly	177211	8087	11116	27000
3	Lung Opacity	105636	5598	6599	105581
4	Lung Lesion	211470	1488	1270	9186
5	Edema	137458	12984	20726	52246
6	Consolidation	152792	27742	28097	14783
7	Pneumonia	195806	18770	2799	6039
8	Atelectasis	154971	33739	1328	33376
9	Pneumothorax	144480	3145	56341	19448
10	Pleural Effusion	90203	11628	35396	86187
11	Pleural Other	216922	2653	316	3523
12	Fracture	211220	642	2512	9040
13	Support Devices	100197	1079	6137	116001

Chest X-ray Dataset

Validation Set: 234

- annotated by certified radiologists

Test Set: 668

- annotated by certified radiologists

Training Set: 223,414 => **Noise Labels**

- *automatically detect the disease by **CheXpert Labeler** with **radiology reports**.*

1. Basically, Report is split and tokenized into sentence using NLTK
2. Positive(1) / Negative(0) / **Uncertain(-1)** / No finding(Nan)

Because of Noisy Labels.....

- Supervised Learning ?
 - 학습데이터 모두 정확한 label이 붙어있음을 가정함
 - 입력 이미지와 label이 정확할 때, 올바른 분류 및 예측이 가능
- Convergence ?
 - 수렴 속도 느림
 - 학습이 진행될수록 잘못된 정보에 Overfitting

CheXpert

A Large Chest X-Ray Dataset And Competition

What is CheXpert?

CheXpert is a large dataset of chest X-rays and competition for automated chest x-ray interpretation, which features uncertainty labels and radiologist-labeled reference standard evaluation sets.

[READ THE PAPER \(IRVIN & RAJPURKAR ET AL.\)](#)

Why CheXpert?

Chest radiography is the most common imaging examination globally, critical for screening, diagnosis, and management of many life threatening diseases. Automated chest radiograph interpretation at the level of practicing radiologists could provide substantial benefit in many medical settings, from improved workflow prioritization and clinical decision support to large-scale screening and global population health initiatives. For progress in both development and validation of automated algorithms, we realized there was a need for a labeled dataset that (1) was large, (2) had strong reference standards, and (3) provided expert human performance metrics for comparison.

Leaderboard

Will your model perform as well as radiologists in detecting different pathologies in chest X-rays?

Rank	Date	Model	AUC	Num Rads Below Curve
1	Aug 31, 2020	DeepAUC-v1 <i>ensemble</i> https://arxiv.org/abs/2012.03173	0.930	2.8
2	Sep 01, 2019	Hierarchical-Learning-V1 (ensemble) <i>Vingroup</i> <i>Big Data Institute</i> https://arxiv.org/abs/1911.06475	0.930	2.6
3	Oct 15, 2019	Conditional-Training-LSR <i>ensemble</i>	0.929	2.6
4	Dec 04, 2019	Hierarchical-Learning-V4 (ensemble) <i>Vingroup</i> <i>Big Data Institute</i> https://arxiv.org/abs/1911.06475	0.929	2.6

03

연구 내용

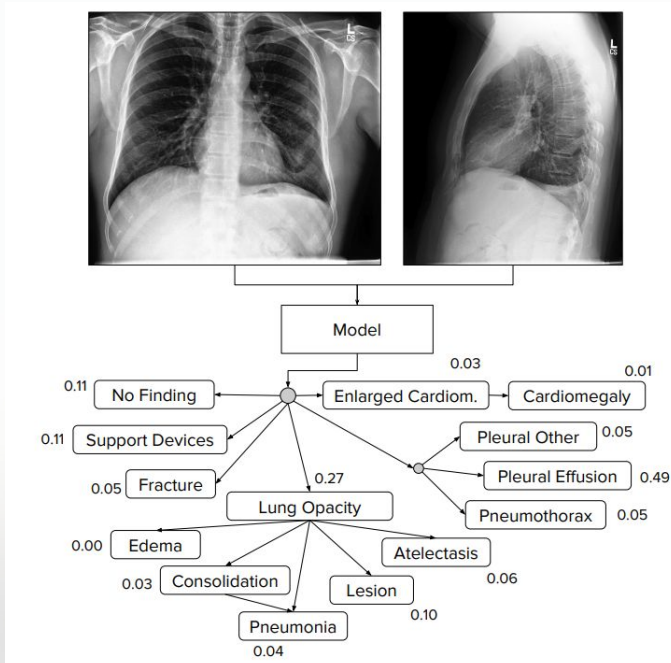
How?

1. ImageNet Pretrained - Transfer Learning

- Better Results
- Fast Convergence

2. Disease Hierarchy

- 병들을 판단하는데 있어서 상관 관계 고려
- 대표적으로 계층 관계
- Cardiomegaly => Enlarged Cardiomegaly
- Pneumonia => Consolidation
- Lung Lesion, Edema, Consolidation, Pneumonia, Atelectasis
=> Lung Opacity



How?

3. Mixup augmentation & Label Smoothing

- Better Results
- Uncertain(-1) => 0.5 : Model의 판단 자율성 ↑

4. Not Just Transformer(ViT). Hybrid CNN+ViT

- 일반 ViT로는 ResNet보다 결과가 훨씬 좋지 않음
- Vision Transformers focus more on modeling the global relationship
- CNNs pays more attention to the local texture.
- **We need specific feature(disease) in Image => CNN**

	ResNet50_base	ViT_base
Atelectasis	0.78	0.75
Cardiomegaly	0.85	0.78
Consolidation	0.91	0.78
Edema	0.89	0.85
Pleural Effusion	0.95	0.87
...		
mean AUC	0.832	0.781

But.....

	ResNet (~3)	Hybrid
No Finding	0.902	0.910
Enlarged Cardio.	0.874	0.876
Cardiomegaly	0.896	0.919
Lung Opacity	0.933	0.926
Lung Lesion	0.912	0.917
Edema	0.907	0.916
Consolidation	0.865	0.882
Pneumonia	0.785	0.753
Atelectasis	0.849	0.859
Pneumothorax	0.950	0.969
Pleural Effusion	0.943	0.959
Pleural Other	0.969	0.965
Fracture	0.800	0.676
Support Devices	0.926	0.940

Additional approach?

1. Other Dataset?

- 각 의료기관의 다른 촬영기기로 인한 Domain의 차이가 Distribution Shift를 야기. 즉, 노이즈한 라벨을 처리한다고 해도 다른 문제가 발생할 가능성이 높음

2. Multimodal?

- 현재, 대용량 자연 이미지 데이터를 대상으로 Multi Modal의 연구가 활발
- CheXpert 데이터셋의 radiology report의 부재로 접목 불가

Additional approach?

3. MentorNet

(Jiang, L., Zhou, Z., Leung, T., Li, L. J., & Fei-Fei, L. (2018, July).

Mentornet: Learning data-driven curriculum for very deep neural networks on corrupted labels. In International Conference on Machine Learning)

- 전체 데이터 중 쉬운 패턴의 데이터를 먼저 학습
- training loss 값이 클수록 noisy한 데이터라고 생각하여 가중치 부여.
- But Chexpert Dataset에서 한 명의 데이터 라벨 14개가 모두 완전한 경우가 낮음
14개 항목 중 5개 이상 labeling된 사람 : below 10%

4. Ensemble

- Noisy label에서는 전통적으로 ensemble이 좋은 성과를 보임.
- 여러 모델의 ensemble을 통해 성능 향상을 이루는 것보다는 single model을 통해 classification 성능 향상을 목표로 함.

Idea

1. MentorNet 아이디어

- Loss 작은 데이터를 학습에서 제외하는 것이 아닌, loss 큰 데이터도 학습에 사용할 수 있도록 weight 조정
- 일정 threshold 값을 넘는 데이터의 경우 weight 부여
- Nan값 처리할 때, label에 weight 부여

2. 라벨 재 생성

- 위 Idea들과 방법을 적용한 모델을 이용
- 높은 Accuracy가 나온 모델을 토대로 라벨을 재생성
- 재 학습 - FC layer만 학습, 나머지는 Freeze

MentorNet 아이디어

	ResNet50_base	ResNet50_0.3	ResNet50_0.5
Lung Lesion	0.69	0.87	0.90
Pneumonia	0.73	0.77	0.75
Fracture	0.71	0.75	0.74
...			
mean AUC	0.832	0.841	0.843

04

결론

	ResNet (~3)	Hybrid	Hybrid + Idea
No Finding	0.902	0.910	0.912
Enlarged Cardio.	0.874	0.876	0.882
Cardiomegaly	0.896	0.919	0.921
Lung Opacity	0.933	0.926	0.924
Lung Lesion	0.912	0.917	0.915
Edema	0.907	0.916	0.914
Consolidation	0.865	0.882	0.892
Pneumonia	0.785	0.753	0.832
Atelectasis	0.849	0.859	0.889
Pneumothorax	0.950	0.969	0.965
Pleural Effusion	0.943	0.959	0.954
Pleural Other	0.969	0.965	0.967
Fracture	0.800	0.676	0.872
Support Devices	0.926	0.940	0.938
MEAN AUC	0.894	0.891	0.913

For Robustness



Labelling

Most Important Factors



Transformer

But with CNN



New Idea

**From MentorNet,
ReLabelling**

Thanks!