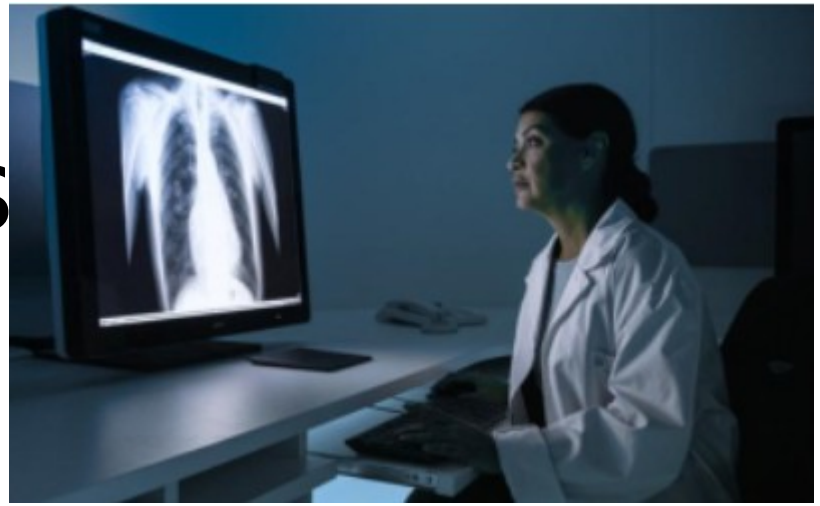
Medical Imaging Diagnosis: Classification, Localization, and Interpretation

Medical Imaging

- Ultrasound Imaging
- MRI (Magnetic Resonance Imaging)
- Pediatric X-ray Imaging
- Medical X-ray Imaging
- Radiography
- Computed Tomography (CT)
- Dental Cone-beam Computed Tomography
- Fluoroscopy
- Mammography

https://www.fda.gov/radiation-emitting-products/radiation-emitting-products-and-procedures/medical-imaging

Radiologists



- Radiologists are physicians who are specialized in diagnosing and treating diseases using a variety of medical imaging techniques such as x-rays, computed tomography (CT), magnetic resonance imaging (MRI), Positron emission tomography (PET), and ultrasound.
- Expertise in radiology is defined by refined visual search patterns and diagnostic accuracy.
- Expert radiologists not only perceive abnormalities that non-experts do not, but they also better understand what to attend to and what to ignore.

https://www.frontiersin.org/articles/10.3389/fnhum.2019.00213/full

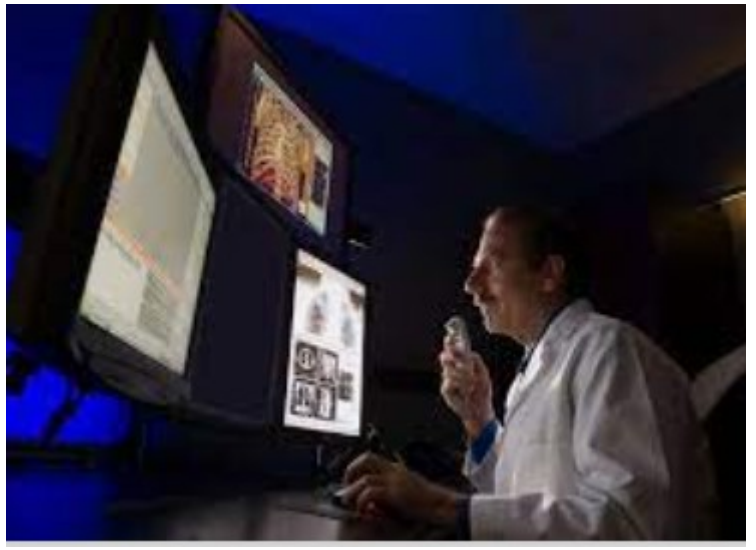
Radiologists and Images

- 2 billion chest x-rays are done each year worldwide.
- In US, 800 million medical scans a year, which amounts to 600 billion images,

- A radiologist reads about 20,000 images a year, roughly 50-100 per day, and the number is increasing
- The U.S. has **approximately 34,000** radiologists, or 100 per million according to the Journal of Nuclear Medicine (2015)
- 31% of American radiologists have experienced at least one malpractice claim, often missed diagnoses.

Eric Topol, Deep Medicine, 2019

How radiologists work



- Broadly, diagnostic radiology entails
  - (1) detection—noting a potentially significant finding is present that merits further analysis;
  - (2) recognition—deciding that the finding is pathologic;
  - (3) discrimination—characterizing the lesion as a specific type; and
  - (4) diagnosis -- The first task, detection, has primary importance, because all following steps leading to diagnosis rely on detection efficacy

Findings:
There is no focal consolidation, effusion or pneumothorax. The cardiomediastinal silhouette is normal. There has been interval resolution of pulmonary vascular congestion since DATE.
Impression:
No pneumonia or pulmonary vascular congestion. Telephone notification to dr. NAME at TIME on DATE per request

The error rate of radiologists scanning images is around 33% and has not been changed for the past seven decades.

https://arxiv.org/pdf/1904.02633.pdf

https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6603246/pdf/fnhum-13-00213.pdf

AI also makes mistake

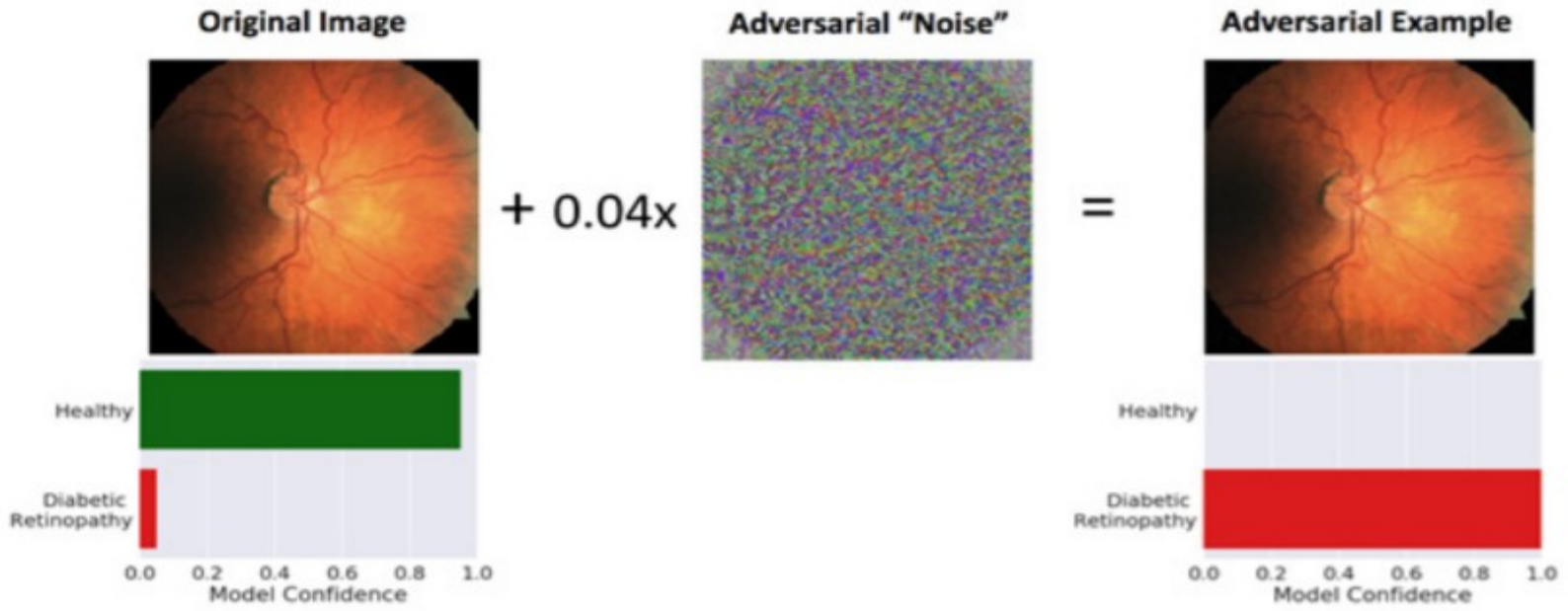


FIGURE 4-2 | Construction of an "adversarial example." Left: An unaltered fundus image of a healthy retina. The AI system (bottom left) correctly identifies it as a healthy eye. Middle: Adversarial "noise" that is constructed with knowledge of the AI system is added to the original image. Right: Resulting adversarial image that superimposes the original image and the adversarial noise. Though the original image is indistinguishable from the adversarial example to human eyes, the AI system has now changed the diagnosis to diabetic retinopathy with essentially 100 percent confidence.
SOURCE: Image was provided by Samuel Finlayson.

https://nap.nationalacademies.org/read/27111/chapter/6#109

JAMA Network Open

JAMA Network Open. 2021;4(12):e2141096. doi:10.1001/jamanetworkopen.2021.41096

Original Investigation | Imaging

An Artificial Intelligence–Based Chest X-ray Model on Human Nodule Detection Accuracy From a Multicenter Study

AI-Rad Companion
SIEMENS Healthineers

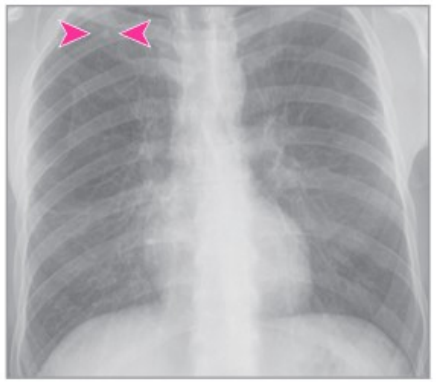
Fatemeh Homayounieh, MD; Subba Digumarthy, MD; Shadi Ebrahimian, MD; Johannes Rueckel, MD; Boj Friedrich Hoppe, MD; Bastian Oliver Sabel, MD; Sailesh Conjeti, PhD; Karsten Ridder, MD; Markus Sistermanns, MD; Lei Wang, PhD; Alexander Preuhs, PhD; Florin Ghesu, PhD; Awais Mansoor, PhD; Mateen Moghbel, MD; Ariel Botwin, MD; Ramandeep Singh, MD; Samuel Cartmell, MD; John Patti, MD; Christian Huemmer, PhD; Andreas Fieselmann, PhD; Clemens Joerger, PhD; Negar Mirshahzadeh, MSc; Victorine Muse, MD; Mannudeep Kalra, MD

DESIGN, SETTING, AND PARTICIPANTS This diagnostic study included 100 posteroanterior chest radiograph images taken between 2000 and 2010 of adult patients from an ambulatory health care center in Germany and a lung image database in the US. Included images were selected to represent nodules with different levels of detection difficulties (from easy to difficult), and comprised both normal and nonnormal control.

EXPOSURES All images were processed with a novel AI algorithm, the AI Rad Companion Chest X-ray. Two thoracic radiologists established the ground truth and 9 test radiologists from Germany and the US independently reviewed all images in 2 sessions (unaided and AI-aided mode) with at least a 1-month washout period.

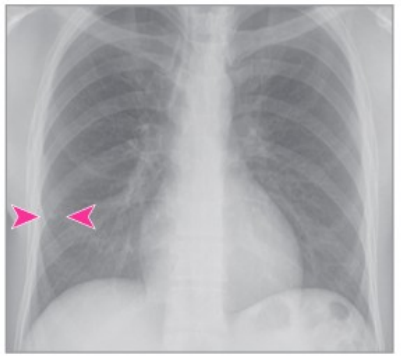
CONCLUSIONS AND RELEVANCE In this diagnostic study, an AI algorithm was associated with improved detection of pulmonary nodules on chest radiographs compared with unaided interpretation for different levels of detection difficulty and for readers with different experience.

A AI score: 9



Source	Unaided and aided scores	
	Unaided	Aided
J5	0	9
J1	0	6
J2	0	7
S2	8	10

C AI score: 10



Source	Unaided and aided scores	
	Unaided	Aided
J6	0	4
S3	0	6
J5	0	8
S2	0	10
J4	0	10

Labeled Medical Image Data

Table 2: Large Open-Source Medical Imaging Data Sets

Data Set Description	Image Types	No. of Patients	Ground Truth	Single or Multiple Institutions
American College of Radiology Imaging Network National CT Colonography Trial (ACRIN 6664) (102)	CT	825	Pathology (biopsies)	Multiple
Alzheimer's Disease Neuroimaging Initiative (103)	MRI, PET	>1700	Clinical (follow-up)	Multiple
Curated Breast Imaging Subset of the Digital Database for Screening Mammography (36)	Mammography	6671	Pathology (biopsies)	Multiple
ChestX-ray8, National Institutes of Health chest x-ray database (41)	Radiography	30 805	Imaging reports	Single
CheXpert, chest radiographs (79)	Radiography	65 240	Imaging reports	Single
Collaborative Informatics and Neuroimaging Suite (104)	MRI		Clinical (follow-up)	Multiple
DeepLesion, body CT (60)	CT	4427	Imaging	Single
Head and neck PET/CT (105)	PET/CT, CT	298	Pathology (biopsies), clinical (follow-up)	Multiple
Lung Image Database Consortium image collection (106)	CT, radiography	1010	Imaging, clinical for a subset	Multiple
MRNet, knee MRI (80)	MRI	1370	Imaging reports	Single
Musculoskeletal bone radiographs, or MURA (107)	Radiography	14 863	Imaging reports	Single
National Lung Screening Trial (108)	CT, pathology	26 254	Clinical (follow-up)	Multiple
PROSTATEx Challenge, SPIE-AAPM-NCI Prostate MR Classification Challenge (109)	MRI	346	Pathology (biopsies), imaging	Multiple
Radiological Society of North America Intracranial Hemorrhage Detection (110)	CT	25 000	Imaging	Multiple
Cancer Genome Atlas Kidney Renal Clear Cell Carcinoma data collection (111)	CT, MRI	267	Pathology (biopsies), clinical (follow-up)	Multiple
Virtual Imaging Clinical Trial for Regulatory Evaluation (112)	Mammography, digital breast tomosynthesis	2994	Imaging	Multiple

Note.—AAPM = American Association of Physicists in Medicine, NCI = National Cancer Institute, SPIE = Society of Photo-Optical Instrumentation Engineers.

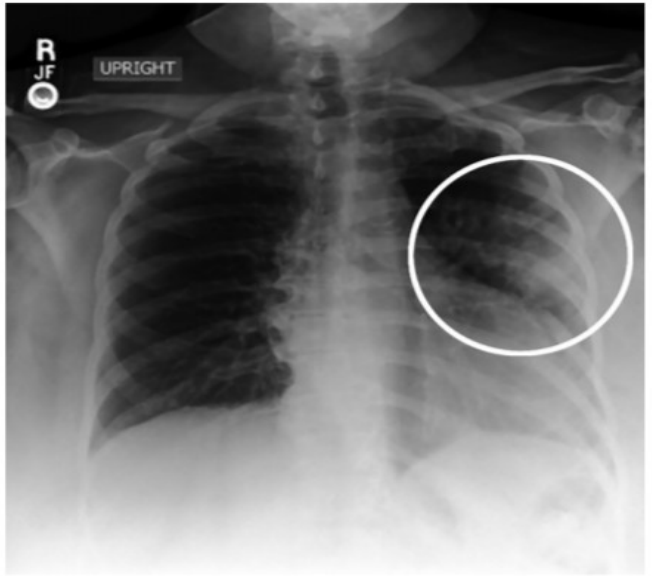


Figure 3: Image in posterior-anterior direction shows nonspecific abnormality on chest radiograph. Application of most accurate label for nonspecific finding such as opacity in left lung (circle) is challenging in absence of other clinical and laboratory data.

Willemink, M., Koszek, W.A., Hardell, C., Wu, J., Fleischmann, D., Harvey, H., Folio, L. R., Summers, R. M., Rubin, D. L., & Lungren, M P. (2020). Preparing medical imaging data for machine learning. Radiology, 295, 4-15.

Better Data Augmentation for Contrastive Learning

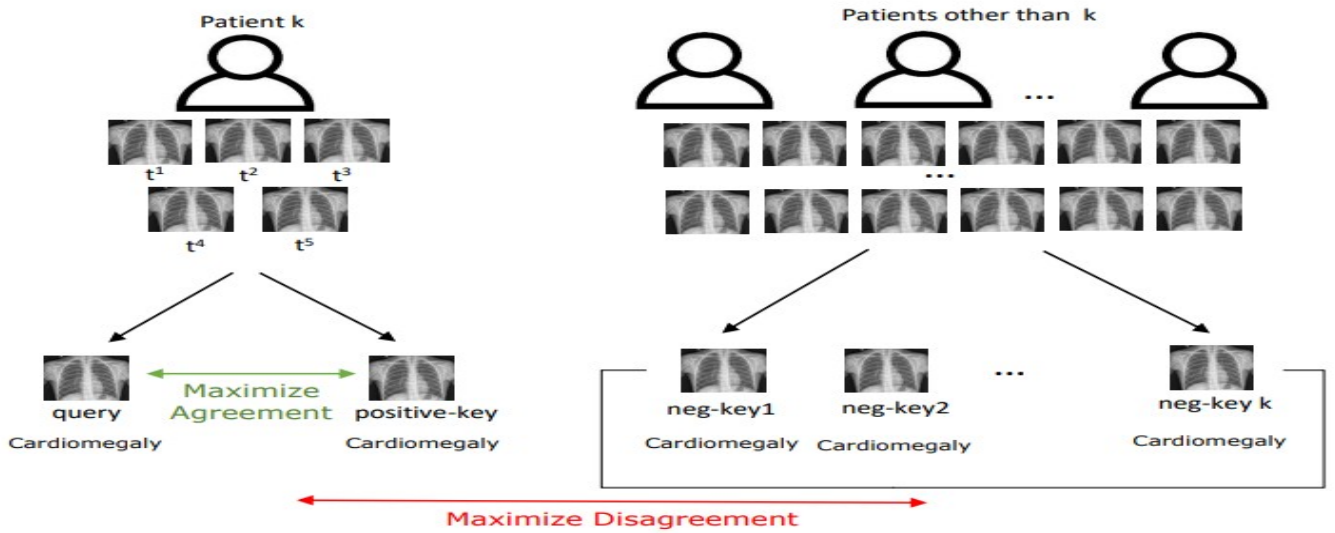


Fig. 2. Overview of our Supervised Patient-Metadata Based Contrastive Augmentation. Positive and Negative Keys are constructed using Patient Metadata (association of patient id with chest X-rays) and Supervised disease labels of chest X-rays.

Jaiswal, A., Li, H., Zander, C., Han, Y., Rousseau, J., Peng, Y., & Ding, Y. (2021). SCALP - Supervised Contrastive Learning for Cardiopulmonary Disease Classification and Localization in Chest X-rays using Patient Metadata. IEEE International Conference on Data Mining (ICDM 2021), Dec 7-10, 2021, Auckland, New Zealand.

Better Data Augmentation for Contrastive Learning

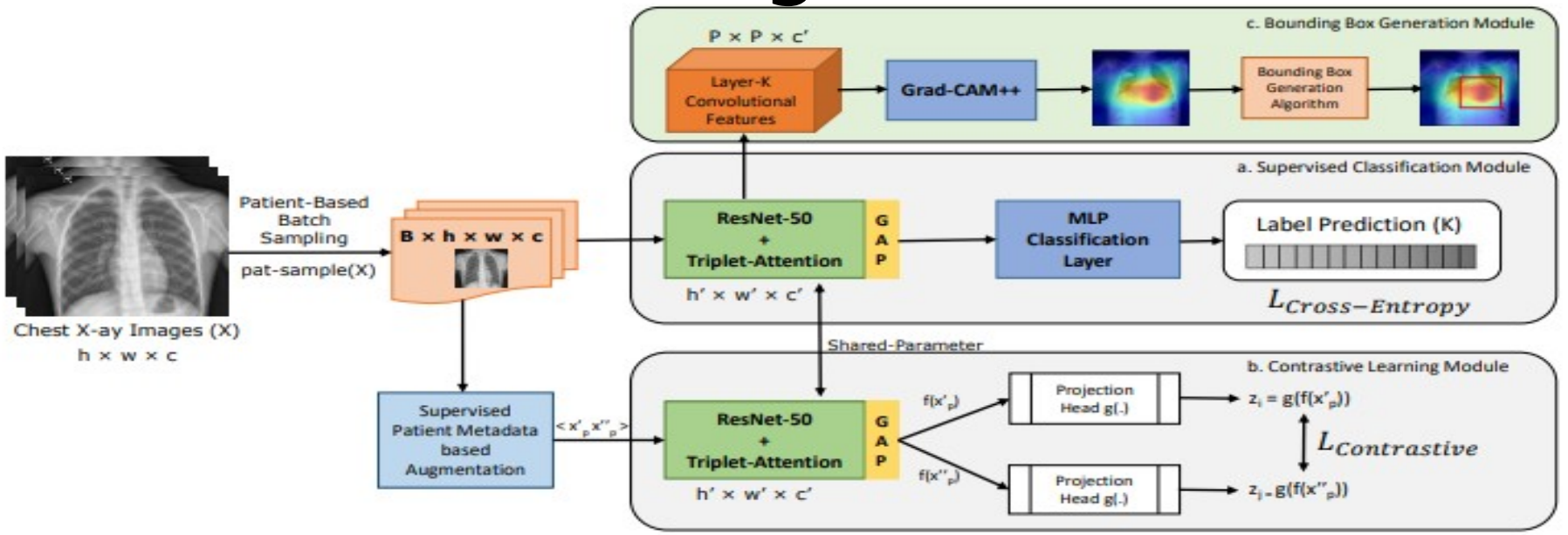
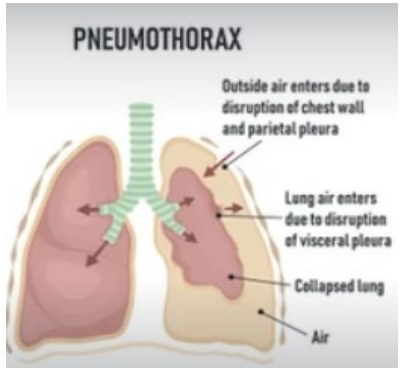


Fig. 1. Model Overview. The input images are sampled in batches with a constraint that no two images in a batch are from the same patient. Learning is performed using a shared encoder (Resent-50) with triplet attention and joint loss from the supervised classification and contrastive learning module.

$$L_{Total} = \lambda \times L_{Cross-Entropy} + (1 - \lambda) \times L_{Contrastive}$$

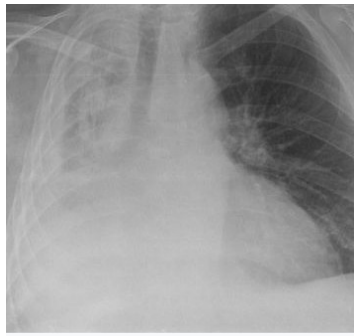
Disease Classification



Method	Atelectasis	Cardiomegaly	Effusion	Infiltration	Mass	Nodule	Pneumonia	Pneumothorax	Mean
Wang et. al. [9]	0.72	0.81	0.78	0.61	0.71	0.67	0.63	0.81	0.718
Wang et. al. [21]	0.73	0.84	0.79	0.67	0.73	0.69	0.72	0.85	0.753
Yao et. al. [22]	0.77	0.90	0.86	0.70	0.79	0.72	0.71	0.84	0.786
Raj. et. al. [4]	0.82	0.91	0.88	0.72	0.86	0.78	0.76	0.89	0.828
Kum. et. al. [23]	0.76	0.91	0.86	0.69	0.75	0.67	0.72	0.86	0.778
Liu et. al. [13]	0.79	0.87	0.88	0.69	0.81	0.73	0.75	0.89	0.801
Seyed et. al. [11]	0.81	0.92	0.87	0.72	0.83	0.78	0.76	0.88	0.821
Our model (std)	0.79 ±0.01	0.92 ±0.00	0.79 ±0.01	0.89 ±0.01	0.88 ±0.02	0.87 ±0.00	0.77 ±0.01	0.81 ±0.02	0.839

TABLE I

COMPARISON WITH THE BASELINE MODELS FOR AUC OF EACH CLASS AND AVERAGE AUC. FOR EACH COLUMN, RED VALUES DENOTE THE BEST RESULTS. NOTE THAT THE BEST BASELINE WITH MEAN AUC 0.828 USES THE DENSENET-121 ARCHITECTURE, WHILE OUR MODEL IS TRAINED USING A COMPARATIVELY SIMPLE AND LIGHT-WEIGHT RESNET-50 ARCHITECTURE.

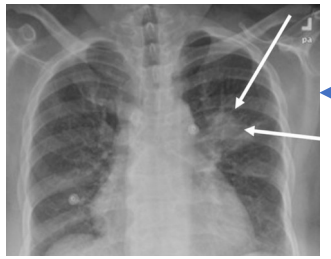


Atelectasis of a person's right lung

Mass, Nodule = A lung nodule (or mass) is a small abnormal area

Pneumothorax = total lung collapse
Atelectasis = partial lung collapse

Cardiomegaly = an enlarged heart



Infiltration = fluid or some unusual substance in the lung tissue.
Effusion = too much pleural fluid around the lungs.

Pneumonia = lung inflammation caused by bacterial or viral infection, in which the air sacs fill with pus, fluid, and may become solid

Method	Atelectasis	Cardiomegaly	Effusion	Infiltration	Mass	Nodule	Pneumonia	Pneumothorax	Mean
Wang et. al. [9]	0.72	0.81	0.78	0.61	0.71	0.67	0.63	0.81	0.718
Wang et. al. [21]	0.73	0.84	0.79	0.67	0.73	0.69	0.72	0.85	0.753
Yao et. al. [22]	0.77	0.90	0.86	0.70	0.79	0.72	0.71	0.84	0.786
Raj. et. al. [4]	0.82	0.91	0.88	0.72	0.86	0.78	0.76	0.89	0.828
Kum. et. al. [23]	0.76	0.91	0.86	0.69	0.75	0.67	0.72	0.86	0.778
Liu et. al. [13]	0.79	0.87	0.88	0.69	0.81	0.73	0.75	0.89	0.801
Seyed et. al. [11]	0.81	0.92	0.87	0.72	0.83	0.78	0.76	0.88	0.821
Our model	0.79	0.92	0.79	0.89	0.88	0.87	0.77	0.81	0.839
(std)	±0.01	±0.00	±0.01	±0.01	±0.02	±0.00	±0.01	±0.02	

TABLE I

COMPARISON WITH THE BASELINE MODELS FOR AUC OF EACH CLASS AND AVERAGE AUC. FOR EACH COLUMN, RED VALUES DENOTE THE BEST RESULTS. NOTE THAT THE BEST BASELINE WITH MEAN AUC 0.828 USES THE DENSENET-121 ARCHITECTURE, WHILE OUR MODEL IS TRAINED USING A COMPARATIVELY SIMPLE AND LIGHT-WEIGHT RESNET-50 ARCHITECTURE.

T(IoU)	Model	**Atelectasis**	**Cardiomegaly**	**Effusion**	**Infiltration**	**Mass**	**Nodule**	**Pneumonia**	**Pneumothorax**	**Mean**
0.1	Baseline [9]	0.69	0.94	0.66	0.71	0.40	0.14	0.63	0.38	0.569
	Our Model	0.62	0.97	0.64	0.81	0.51	0.12	0.8	0.29	0.595
0.2	Baseline [9]	0.47	0.68	0.45	0.48	0.26	0.05	0.35	0.23	0.371
	Our Model	0.42	0.92	0.42	0.6	0.25	0.04	0.56	0.18	0.434
0.3	Baseline [9]	0.24	0.46	0.30	0.28	0.15	0.04	0.17	0.13	0.221
	Our Model	0.29	0.78	0.23	0.37	0.13	0.01	0.4	0.05	0.283
0.4	Baseline [9]	0.09	0.28	0.20	0.12	0.07	0.01	0.08	0.07	0.115
	Our Model	0.18	0.55	0.12	0.19	0.09	0.01	0.25	0.02	0.176
0.5	Baseline [9]	0.05	0.18	0.11	0.07	0.01	0.01	0.03	0.03	0.061
	Our Model	0.07	0.33	0.04	0.10	0.04	0.0	0.14	0.10	0.102
0.6	Baseline [9]	0.02	0.08	0.05	0.02	0.00	0.01	0.02	0.03	0.029
	Our Model	0.02	0.14	0.02	0.04	0.03	0.00	0.07	0.00	0.040
0.7	Baseline [9]	0.01	0.03	0.02	0.00	0.00	0.00	0.01	0.02	0.011
	Our Model	0.01	0.04	0.01	0.03	0.01	0.00	0.02	0.00	0.015

TABLE II

DISEASE LOCALIZATION UNDER VARYING IOU ON THE NIH CHEST X-RAY DATASET. NOTE THAT SINCE OUR MODEL DOESN'T USE ANY GROUND TRUTH BOUNDING BOX INFORMATION, TO FAIRLY EVALUATE THE PERFORMANCE OF OUR MODEL, WE ONLY CONSIDER THE PREVIOUS METHODS' RESULTS WITH THE SAME SETTINGS AS SCALP.

	SCALP w/o Contrastive	SCALP
Atelectasis	0.751	0.79
Cardiomegaly	0.850	0.92
Effusion	0.833	0.79
Infiltration	0.670	0.89
Mass	0.694	0.88
Nodule	0.640	0.87
Pneumonia	0.700	0.77
Pneumothorax	0.792	0.81
Mean	0.7413	0.839

TABLE III
AUC COMPARISON OF SCALP WITH AND WITHOUT CONTRASTIVE LEARNING MODULE.

λ	0.99	0.90	0.85	0.80	0.75	0.70
AUC	0.766	0.794	0.822	**0.839**	0.818	0.785

TABLE IV
AUC COMPARISON OF SCALP FOR VARYING λ IN EQUATION 7. A HIGHER VALUE OF λ IMPLIES LOWER WEIGHT TO THE CONTRASTIVE LOSS. SCALP ACHIEVES THE BEST PERFORMANCE WHEN 80% WEIGHT IS GIVEN TO CLASSIFICATION LOSS AND 20% WEIGHT IS GIVEN TO CONTRASTIVE LOSS.

$$L_{Total} = \lambda \times L_{Cross-Entropy} + (1 - \lambda) \times L_{Contrastive}$$

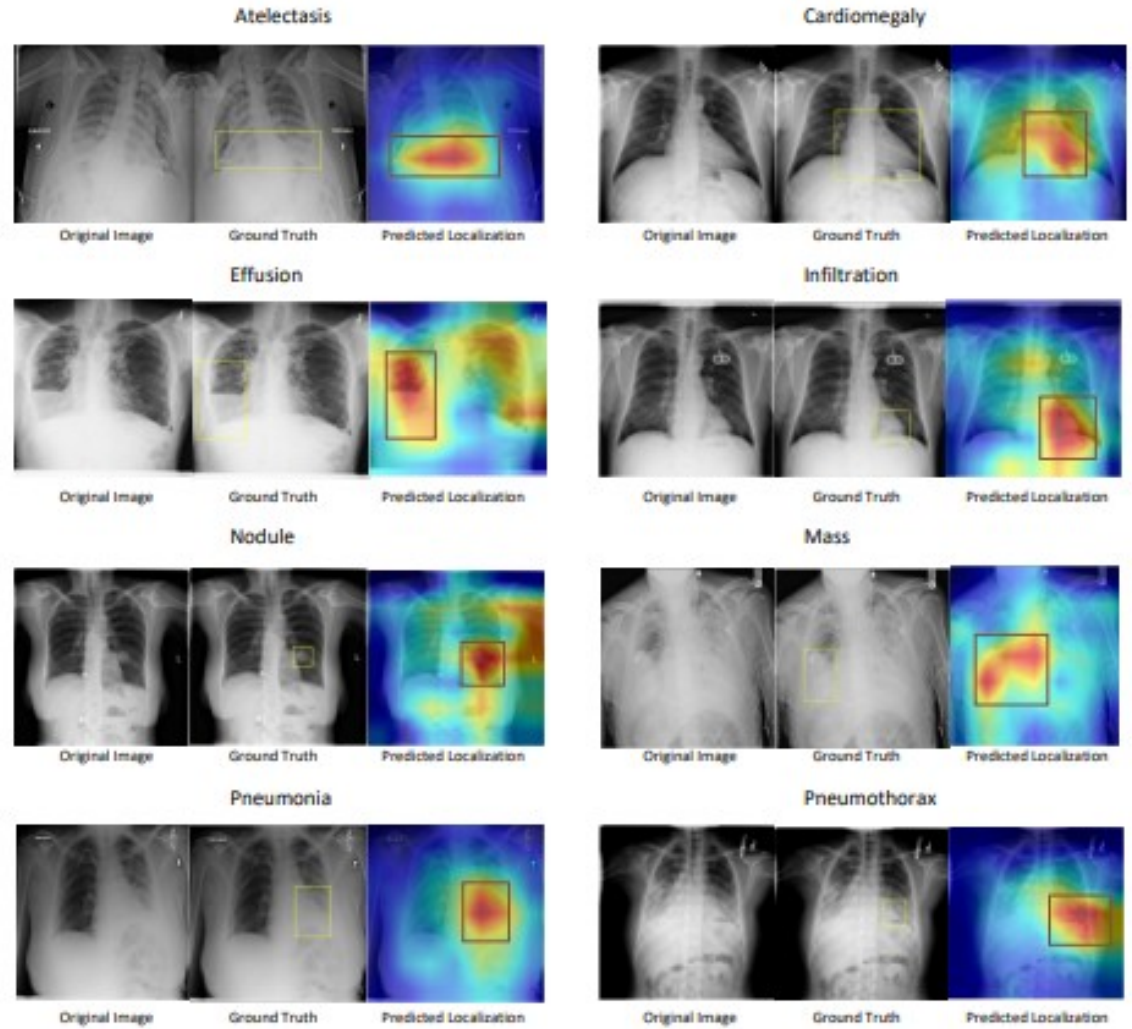


Fig. 3. Examples of visualization of localization on the test images. We plot the results of diseases near the thoracic. The attention maps are generated from the fourth layer of SCALP's encoder and overlapped with its corresponding original radiology image. The ground-truth and the predicted bounding boxes are shown in yellow and red color respectively.

Explainability: Grad-CAM

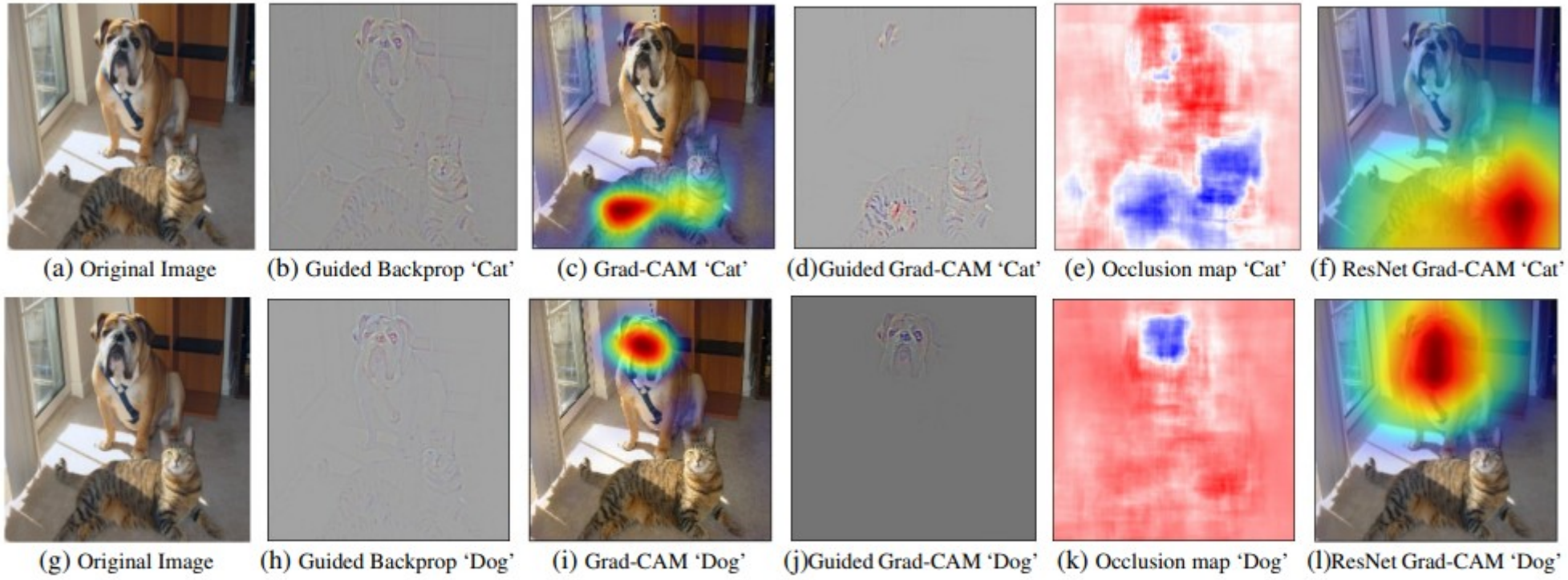


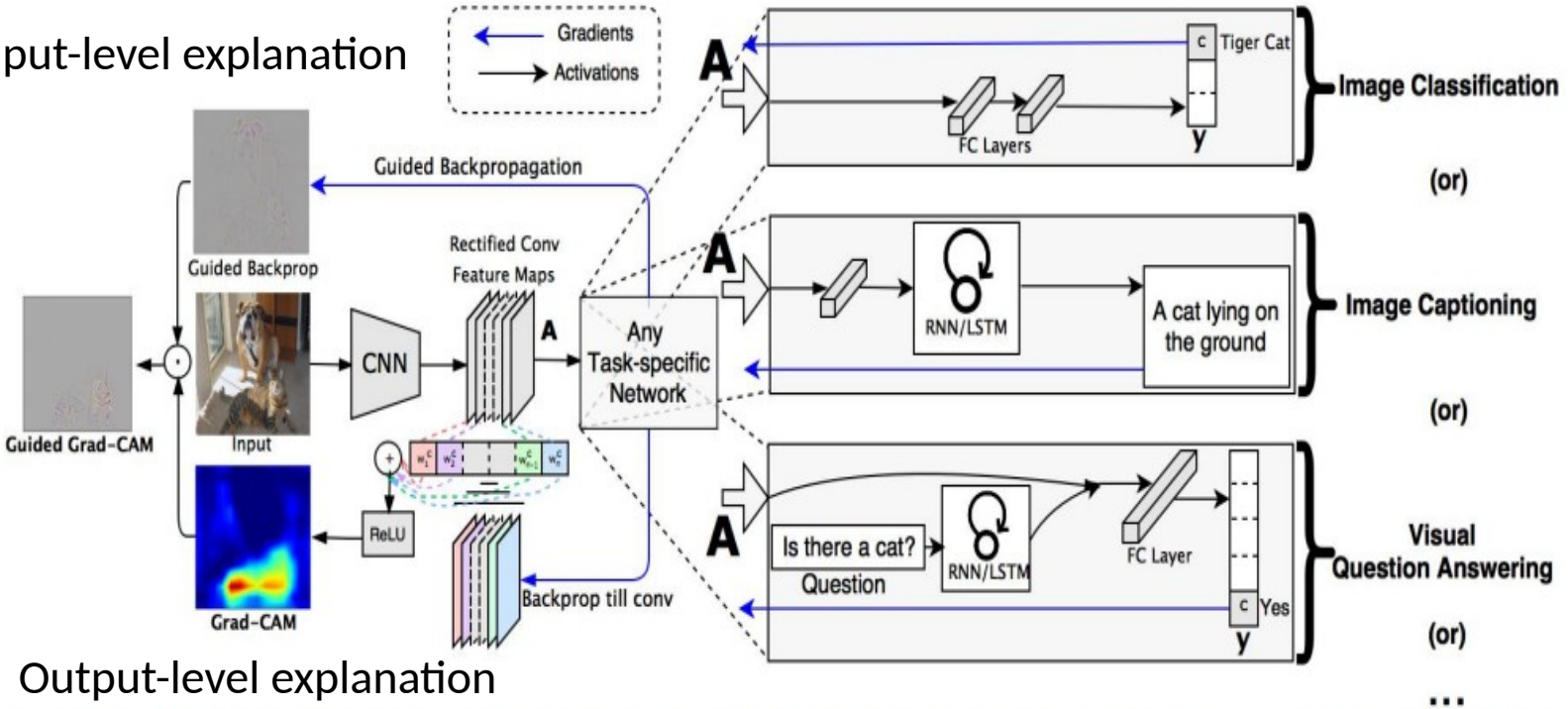
Fig. 1: (a) Original image with a cat and a dog. (b-f) Support for the cat category according to various visualizations for VGG-16 and ResNet. (b) Guided Backpropagation [53]: highlights all contributing features. (c, f) Grad-CAM (Ours): localizes class-discriminative regions, (d) Combining (b) and (c) gives Guided Grad-CAM, which gives high-resolution class-discriminative visualizations. Interestingly, the localizations achieved by our Grad-CAM technique, (c) are very similar to results from occlusion sensitivity (e), while being orders of magnitude cheaper to compute. (f, l) are Grad-CAM visualizations for ResNet-18 layer. Note that in (c, f, i, l), red regions corresponds to high score for class, while in (e, k), blue corresponds to evidence for the class. Figure best viewed in color.

Grad-CAM: Visual Explanations from Deep Networks via Gradient-based Localization

Ramprasaath R. Selvaraju, Michael Cogswell, Abhishek Das, Ramakrishna Vedantam, Devi Parikh, Dhruv Batra

Grad-CAM

Input-level explanation



Output-level explanation

Fig. 2: Grad-CAM overview: Given an image and a class of interest (*e.g.*, 'tiger cat' or any other type of differentiable output) as input, we forward propagate the image through the CNN part of the model and then through task-specific computations to obtain a raw score for the category. The gradients are set to zero for all classes except the desired class (tiger cat), which is set to 1. This signal is then backpropagated to the rectified convolutional feature maps of interest, which we combine to compute the coarse Grad-CAM localization (blue heatmap) which represents where the model has to look to make the particular decision. Finally, we pointwise multiply the heatmap with guided backpropagation to get Guided Grad-CAM visualizations which are both high-resolution and concept-specific.

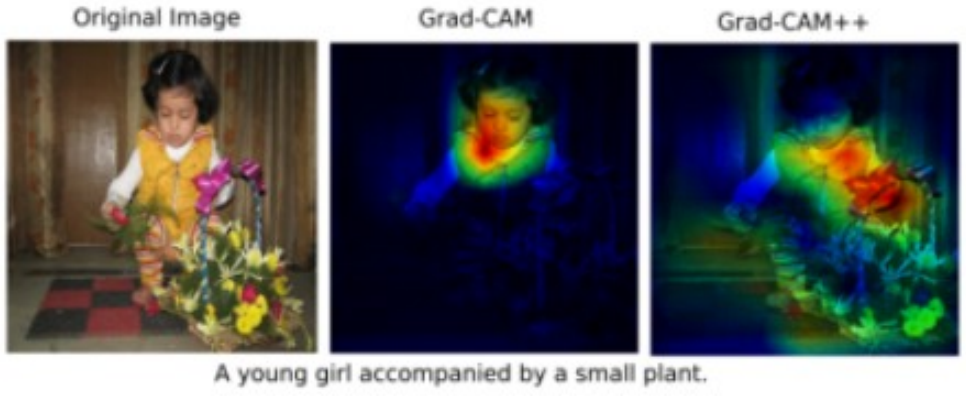
https://arxiv.org/pdf/1610.02391.pdf

- Grad-CAM lacks the ability to highlight fine grained details like pixel-space gradient visualization methods (Guided Backpropagation) by using the gradient information of the last convolutional layer of CNN
- Guided Backpropagation visualizes gradients with respect to the image where negative gradients are suppressed when backpropagating through ReLU layers.
- Guided Grad-CAM fuses Guided Backpropagation (pixel space) and Grad-CAM (class conditional property) via element-wise multiplication.

Grad-CAM ++

- Grad-CAM: fails to properly localize objects in an image if the image contains multiple occurrences of the same class.
- Grad-CAM: fails to locate the entire object, but bits and parts of it

- Grad-CAM ++: using relu to only keep the positive gradients
- Grad-CAM++: taking a weighted average of the pixel-wise gradients to store the pixel spatial information



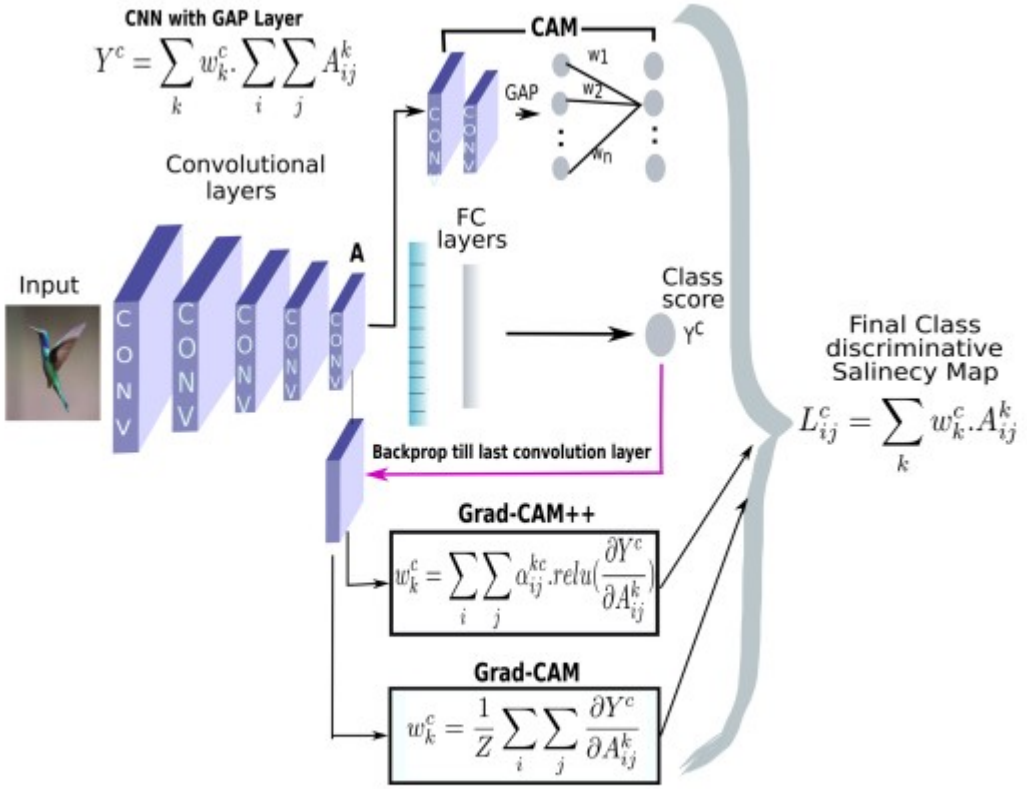


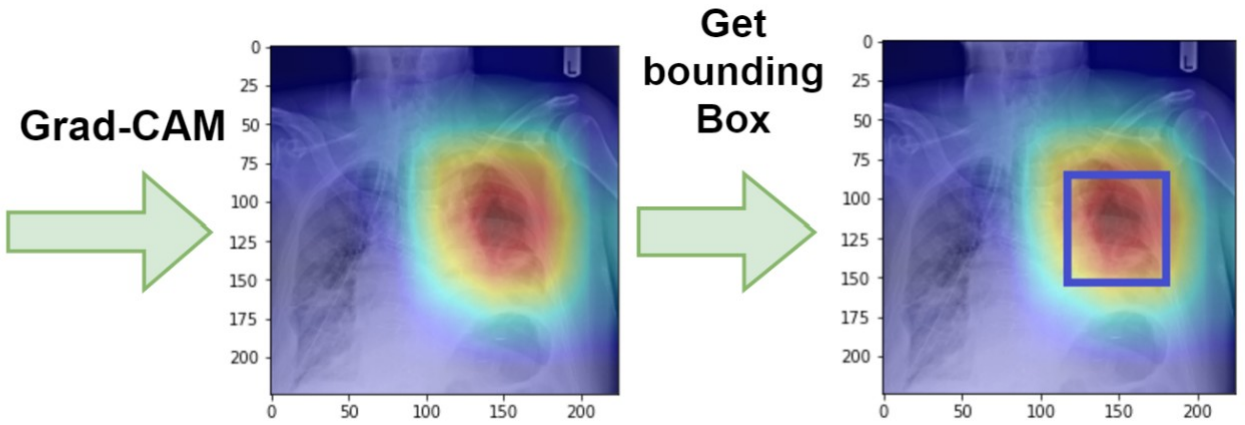
Fig. 3. An overview of all the three methods – CAM, Grad-CAM, Grad-CAM++ – with their respective computation expressions.

α_{ij}^{kc}'s are weighting co-efficients for the pixel-wise gradients for class c and convolutional feature map A^k.

https://arxiv.org/pdf/1710.11063.pdf

Generating Bounding Box for Chest X-Ray Images

DenseNet121 disease prediction	
Disease class 1	0.5
Disease class 2	0.12
Disease class 3	0.07
Disease class 4	0.83
Disease class 5	0.24
Disease class 6	0.01
Disease class 7	0.39
Disease class 8	0.42



https://github.com/thtang/CheXNet-with-localization