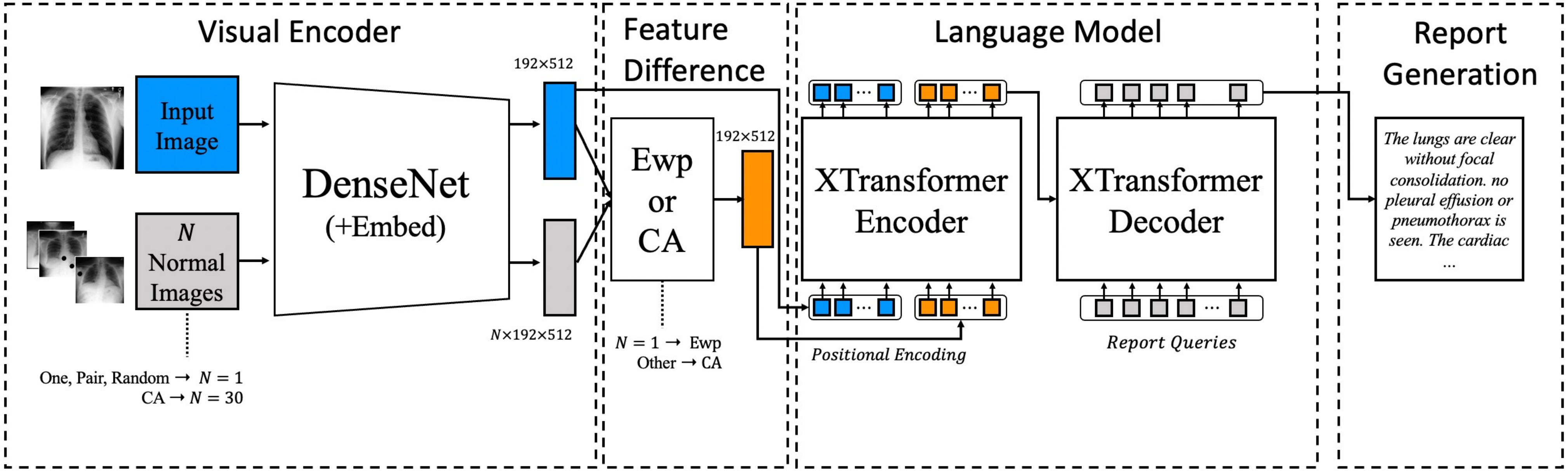
Contrastive Image Captioning Model and Future Direction of Medical AI



2022 중점연구소 여름학교 및 성과발표회

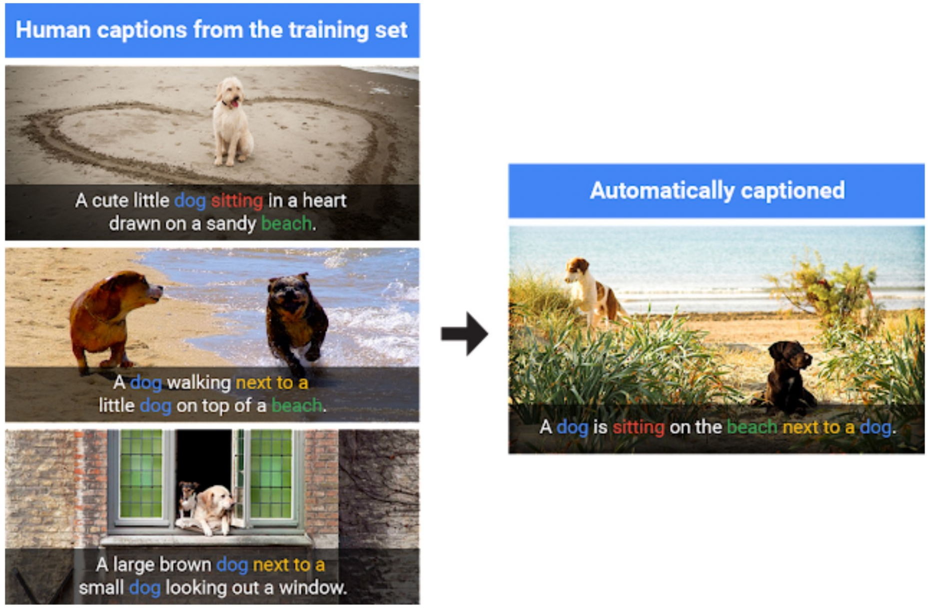
수학과 16학번 박진수

CONTENTS

Image Captioning

: Generate Sentence from Image / Auto Report Generation

General Domain



Medical Domain

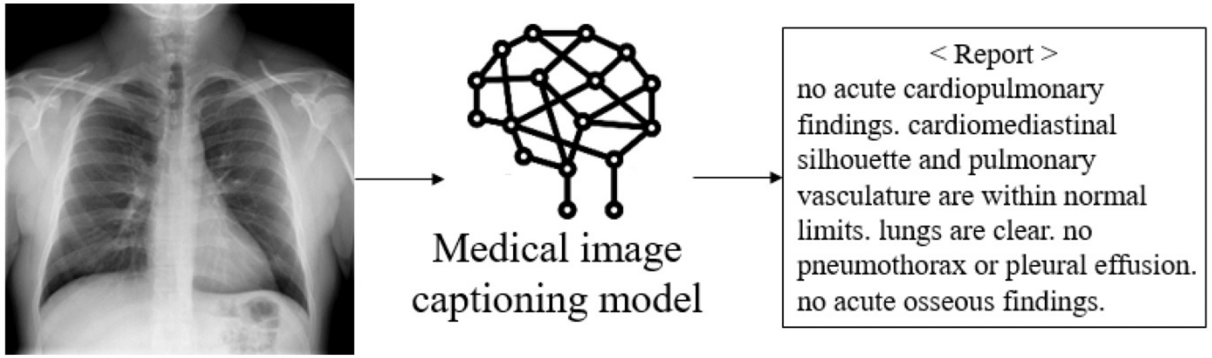


FIGURE 1. A medical image captioning model generates a draft report of the corresponding medical image.

Image Captioning

Chest X-ray Report Generation

Problem

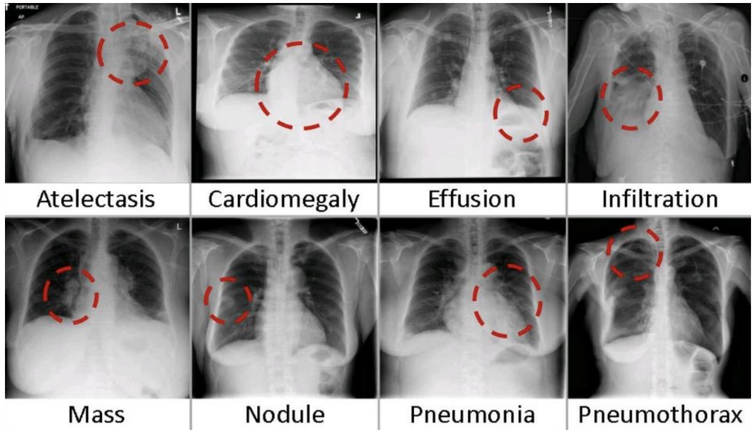
In Domain

- 신뢰도를 최우선으로하는 의학 분야: 모델 결정에 대한 판단 근거와 전향적 평가의 필요성
- But, 데이터 구축, 보안이슈, 그리고 전문의 협업 등의 문제가 산재

In Deep-Learning model

- 잘 갖춰진 (Open) Dataset의 부족
- 의학 도메인 특유의 데이터 이질성(단어, 이미지, …)
  → 전이학습 활용의 어려움

- # of normal images >>> # of abnormal images
- Abnormal image 내 대부분의 영역은 normal region이 차지
- → abnormal region을 포착하기 어려움



https://nihcc.app.box.com/v/ChestXray-NIHCC

Image Captioning

Chest X-ray Report Generation

Problem

Ch2. Interpretability

In Domain

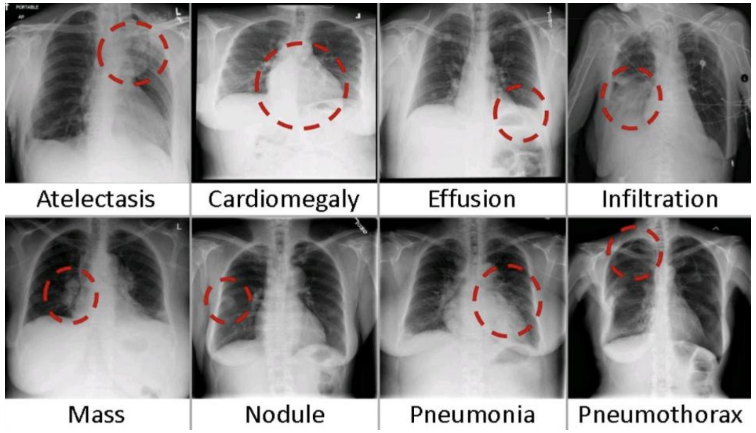
- 신뢰도를 최우선으로하는 의학 분야: 모델 결정에 대한 판단 근거와 전향적 평가의 필요성
- But, 데이터 구축, 보안이슈, 그리고 전문의 협업 등의 문제가 산재

Ch3. Federated Learning

In Deep-Learning model

- 잘 갖춰진 (Open) Dataset의 부족
- 의학 도메인 특유의 데이터 이질성(단어, 이미지, …)
  → 전이학습 활용의 어려움

- # of normal images >>> # of abnormal images
- Abnormal image 내 대부분의 영역은 normal region이 차지
- → abnormal region을 포착하기 어려움



Ch1. Contrastive Image Captioning

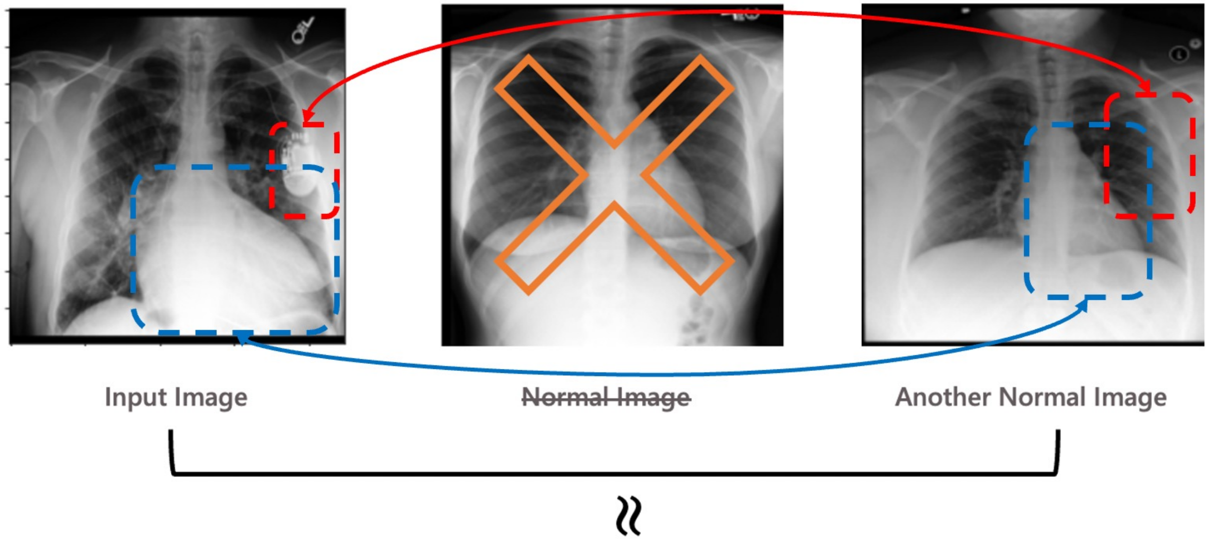
https://nihcc.app.box.com/v/ChestXray-NIHCC

Contrastive Image Captioning Model

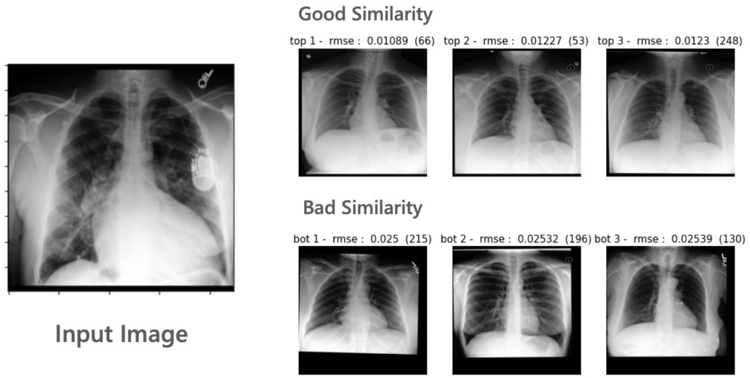
Our Hypothesis 1

**In Visual Encoder, we use input image with auxiliary normal images(like earlier research)
For capturing abnormal region!**

다른 골격, 각도, 성별, 영역 → 배제 필요



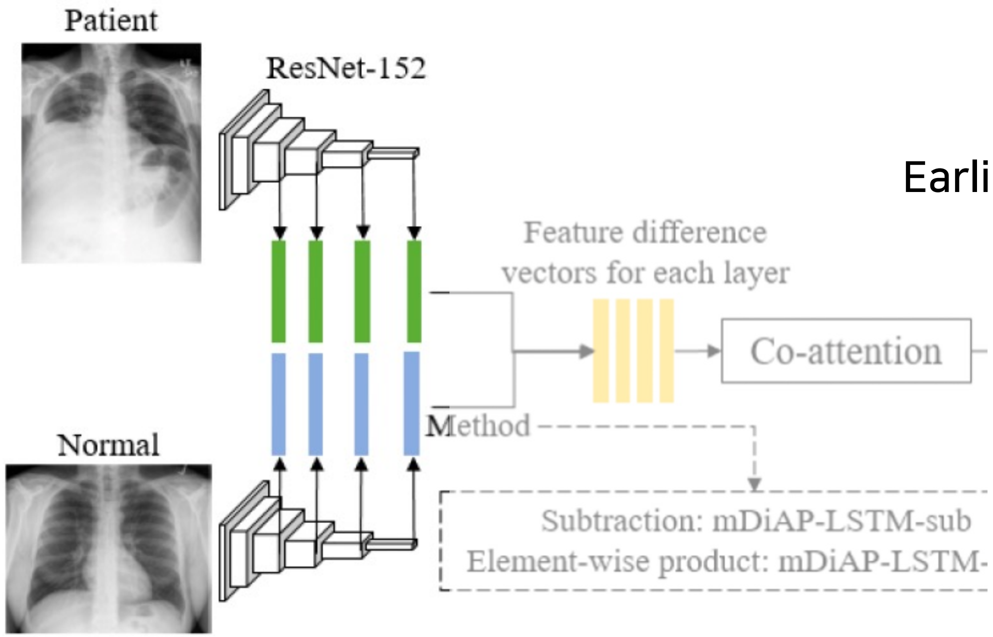
And Noise Data..



Contrastive Image Captioning Model

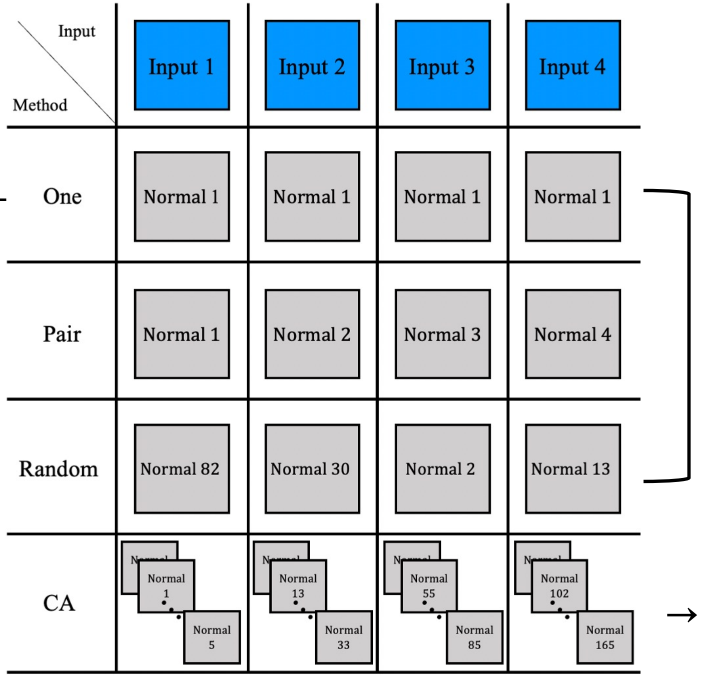
Our Hypothesis 1

In Visual Encoder, we use input image with auxiliary normal images(like earlier research)
For capturing abnormal region!



Earlier Model(Park et al., 2021)

Earlier Model ←



Our Control

→ Our Experiments

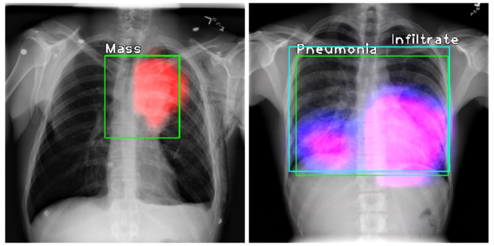
OURS

Contrastive Image Captioning Model

Our Hypothesis2

실제로 해당 분야의 많은 연구와 코드들을 보면 단순히 frozen feature를 받아 Transformer만을 학습하는 경향이 있다.

- Faster R-CNN based pre-trained feature



Fixed feat

Train !

Language Model (Transformer)

- 2-Stage Training(Visual Encoder -> Transformer)

1. Train CNN Extractor

Other Language Model

CNN Extractor

2. Freeze CNN Extractor and Secondly Train Language Model

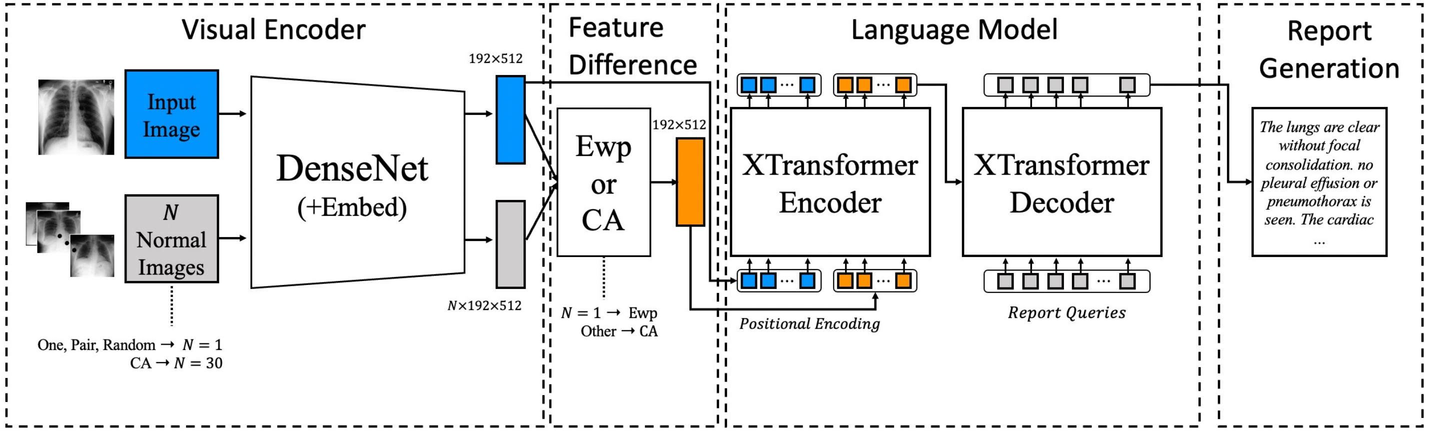
Language Model (Transformer)

Contrastive Image Captioning Model

Our Hypothesis2

실제로 해당 분야의 많은 연구와 코드들을 보면 단순히 frozen feature를 받아 Transformer만을 학습하는 경향이 있다.

OURS : Jointly Train Visual Encoder & Language Model <u>simultaneously</u>

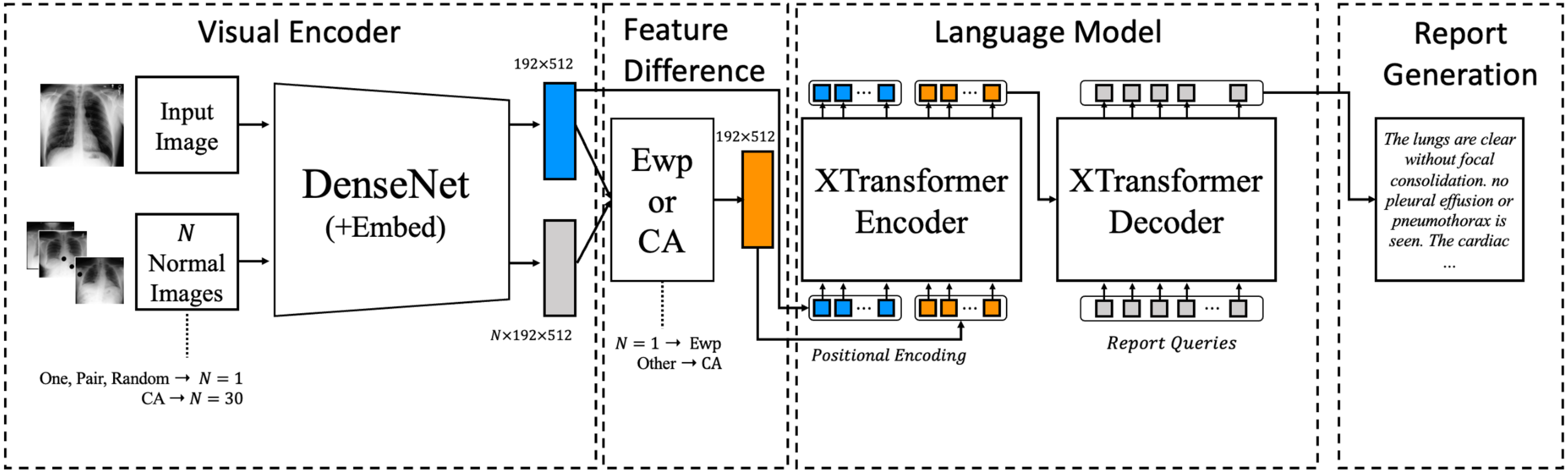


Pros(?)

- 학습 과정 자체가 간결(1-Stage)
- Captioning task로부터 추가로 학습하는 Visual Encoder의 성능 강화(Hypothesis) 가능성
- 단, Learning rate, Architecture 등 Setting을 섬세하게 다룰 필요가 있음

Contrastive Image Captioning Model

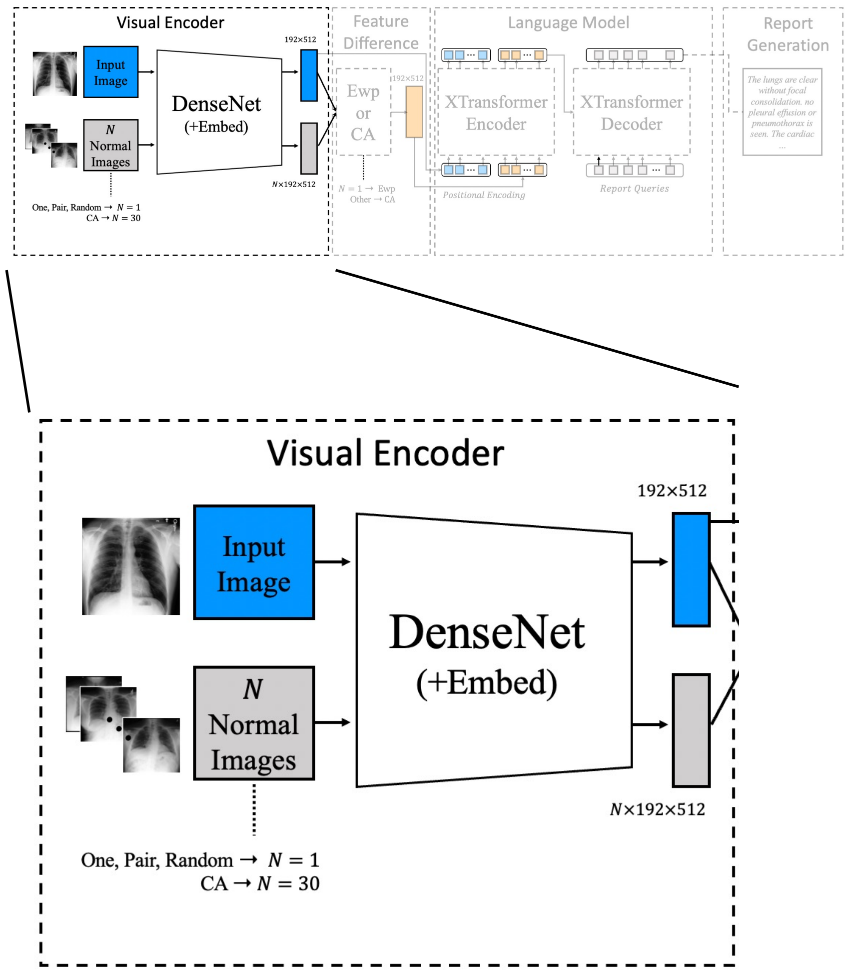
Entire Architecture



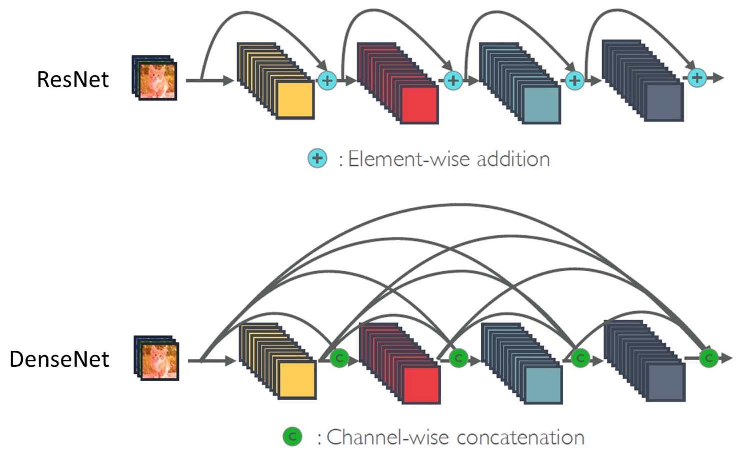
[그림 1] End-to-End Image Captioning Model(DeXTr)

Contrastive Image Captioning Model

Visual Encoder - DenseNet



DenseNet

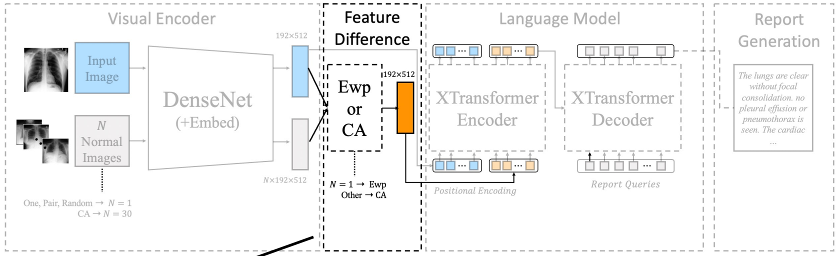


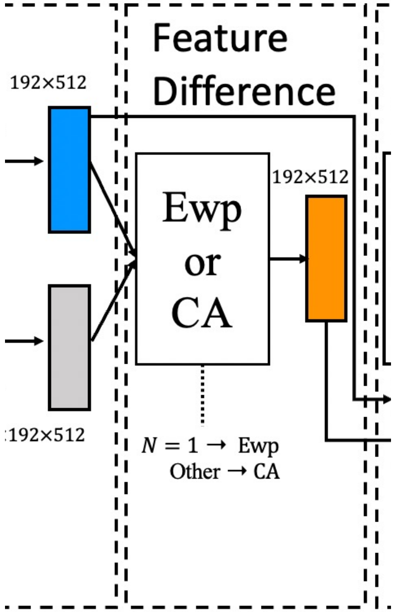
- **Lighter** than Resnet
  - So, avoid overfitting by data scarcity
- Reinforce **Hierarchical feature** for **Fine-to-Coarse information capturing**(Hypothesis)

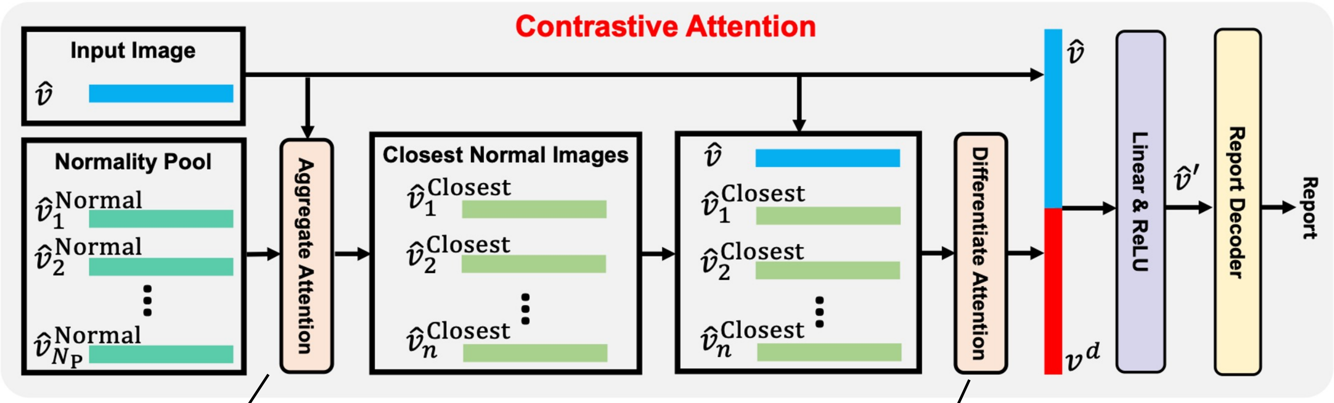
https://pdfs.semanticscholar.org/c3d9/26a85d85a83126f405ad40ff453611148c15.pdf

Contrastive Image Captioning Model

Feature Difference Module : Contrastive Attention Block







유사도가 높은 정상 이미지만 추출(Weighted Sum)

- Multi-head Attention
- $\hat{v} = \frac{1}{L}\sum_{l=1}^{L} f_i^l$, $P_t = \frac{1}{L}\sum_{l=1}^{L} f_{n,t}^l$
- $\hat{v}^{Closest} = Att(\hat{v}, P) = softmax\left(\frac{\hat{v}W^v(PW^P)^T}{\sqrt{512}}\right)P$

공통 정보 추출

- $\widetilde{v^c} = Att([\hat{v}; P], [\hat{v}; P])$
- $v^c = Average\ Pooling(\widetilde{v^c})$
- $v^d = \hat{v} - v^c$

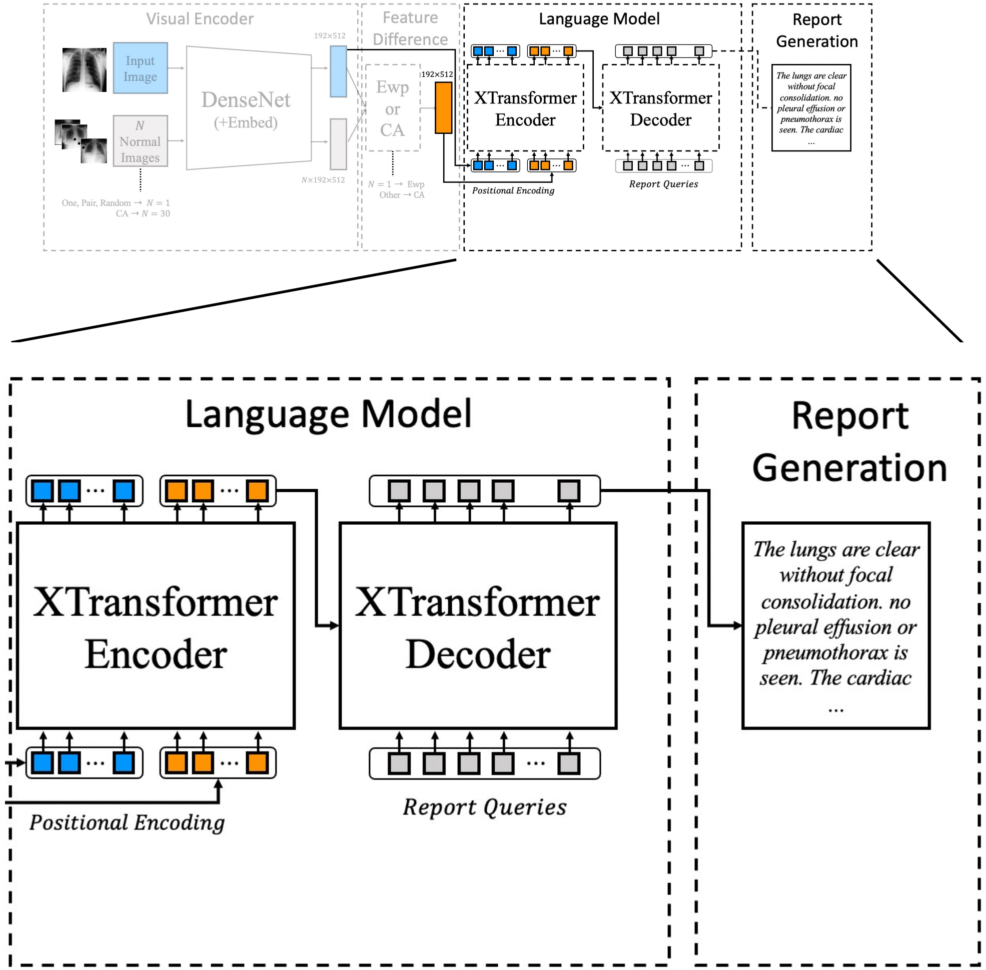
입력 이미지 정보 보존

- $f_c^l = ReLU([f_i^l; v^d]W')$
- $f_c = \{f_c^1, f_c^2, \dots, f_c^L\}$

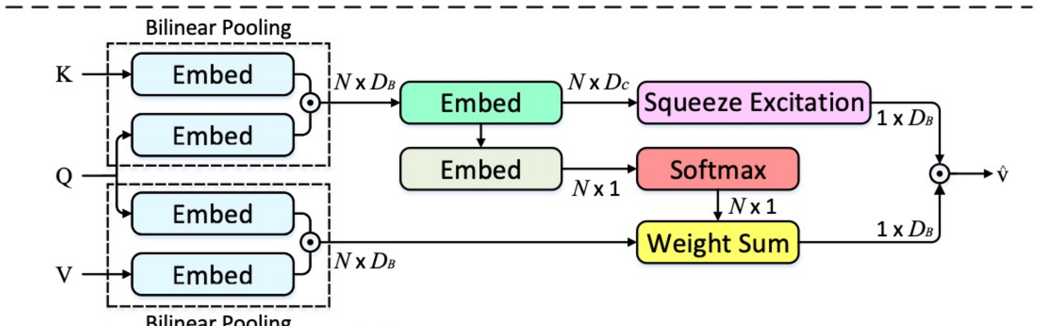
Liu et al., Contrastive Attention for Automatic Chest X-ray Report Generation(2020)

Contrastive Image Captioning Model

Feature Difference Module : Contrastive Attention Block



X-Transformer

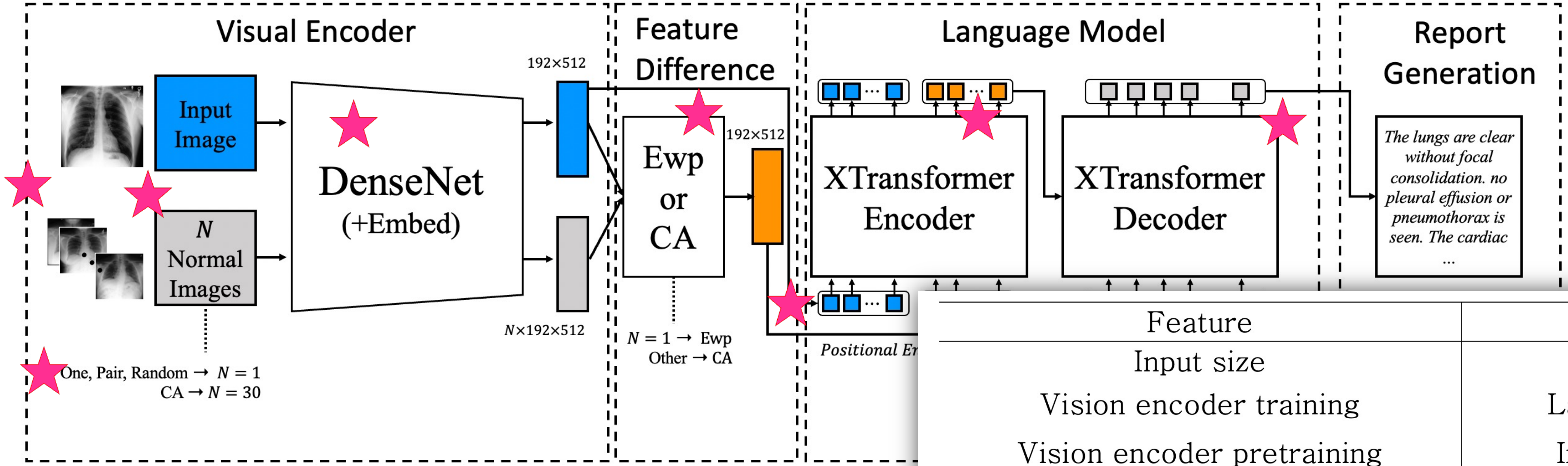


(b) X-Linear attention block

- XTransformer : X-Linear Attention Transformer(SoTA, 2020)
- 내적 → 외적
- 1^{st} order feature interaction -> 2^{nd} order feature intecraction
- Element-wise interaction → All-pairs interaction
- Fine-grained visual recognition

Contrastive Image Captioning Model

In Different Setting..



Feature	Candidates
Input size	**448**, 896
Vision encoder training	Last block, **All blocks**
Vision encoder pretraining	ImageNet, **CheXpert**
Leverage normal	No, One, Pair, Random, CA
Backprop normal	Backprop or **not**
Feature difference	ewp, CA
Concatenate input feature	**Concat** or not
Embed dimension(XTransformer)	256, **512**, 768
Number of normal images in CA(N)	10, **30**, 50

Contrastive Image Captioning Model

Quantitative Evaluation: Performance

Previous Model	Bleu-1	Bleu-2	Bleu-3	Bleu-4	METEOR	ROUGE_L	CIDEr	Remarks
Resnet + Transformer Decoder (End-to-End)	0.3391	0.2166	0.1490	0.1071	0.1574	0.2900	0.2251	
Resnet-Features + X-Tranfsformer (2 stage)	0.3610	0.2432	0.1745	0.1325	0.2240	0.3387	**0.4045**	
End-to-End(Ours)	**Bleu-1**	**Bleu-2**	**Bleu-3**	**Bleu-4**	**METEOR**	**ROUGE_L**	**CIDEr**	
DeXTr-512+No	0.3563	0.2433	0.1727	0.1265	0.2244	0.3390	0.2627	Backprop-Normal,
DeXTr-512+Pair	0.3577	0.2454	0.1761	0.1318	0.2230	0.3469	0.3579	
DeXTr-512+Random	0.3749	0.2573	0.1838	0.1355	0.2300	0.3430	0.3251	
DeXTr-512+ CA10+concat	0.3546	0.2480	0.1789	0.1332	0.2279	0.3475	0.2556	CA → Embed
DeXTr-512+CA30+concat	0.3683	0.2555	0.1843	0.1375	**0.2314**	**0.3523**	0.3689	CA → Embed
DeXTr-512+CA30+no-concat+Dropout	**0.3794**	**0.2620**	**0.1901**	**0.1430**	0.2301	0.3474	**0.3835**	Embed → CA
DeXTr-512+CA30+concat+Dropout	**0.3925**	**0.2712**	**0.1965**	**0.1475**	**0.2346**	**0.3500**	0.3690	Embed → CA

[표 2] 정량 평가 결과(DeXTr-512 를 중심으로)

Contrastive Image Captioning Model

Quantitative Evaluation: diversity of used words

Word in generated report	Ground Truth	정상이미지 활용	정상이미지활용 x	
Word	Target	CA30	No-normal	CA30-512
normal	57	52	**94**	81
enlarged	31	14	**4**	15
small	27	55	33	**25**
low	75	43	**3**	47
clear	30	23	**6**	56

[표 3] 입력 이미지와 정상 이미지 간의 Feature difference 를 강조한 CA30 모델의 결과와 정상 이미지를 사용하지 않은 No-normal 모델의 결과 사이의 특정 단어 빈도수 비교 (Test data 활용)(small size ~ lower volume)

"enlarged, small, low, clear"와 같이 미세한 특징을 표현하는 단어 빈도수↑

Contrastive Image Captioning Model

Quantitative Evaluation: diversity of used words

Input Image

Target

"there is prominence of the bilateral hila with increased interstitial markings consistent with mild volume overload.there are no focal airspace opacities to suggest pneumonia.there are small bilateral pleural effusions. there is no pneumothorax. the mediastinal contours are normal. **the heart size is mildly enlarged.** mild pulmonary edema with small bilateral pleural effusions. mild cardiomegaly. recommend a repeat radiograph after diuresis to reevaluate the prominent hila"

Prediction of CA30 ★

"pa and lateral views of the chest provided. midline sternotomy wires and mediastinal clips are again noted. There is a small right pleural effusion. **the heart size is mildly enlarged.** the mediastinal contour is normal. the mediastinal contour is stable. there is no pneumothorax. bony structures are intact. no free air below the right hemidiaphragm. no acute findings in the right hemidiaphragm. mild cardiomegaly."

Prediction of No-normal

"pa and lateral views of the chest provided. midline sternotomy wires and mediastinal clips are again noted. there is a small right pleural effusion. there is no pneumothorax. **the heart size is normal.** bony structures are intact. no free air below the right hemidiaphragm. mild cardiomegaly with mild pulmonary edema."

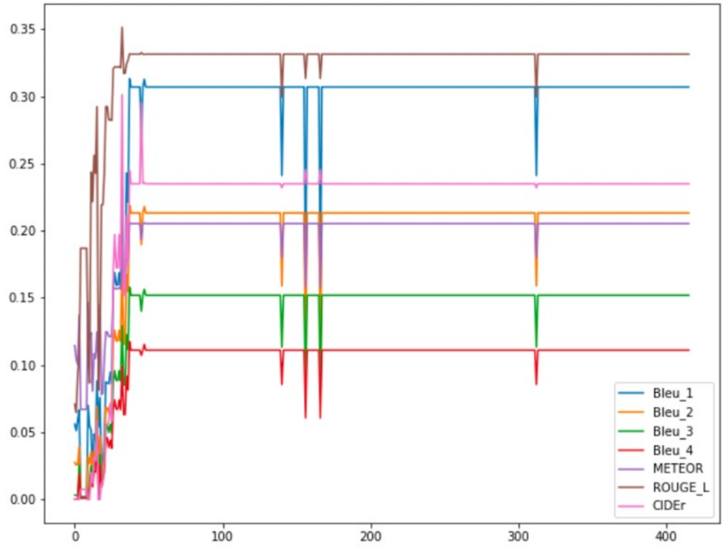
[그림 4] CA30 모델과 No-normal 모델의 일부 결과 비교

Contrastive Image Captioning Model

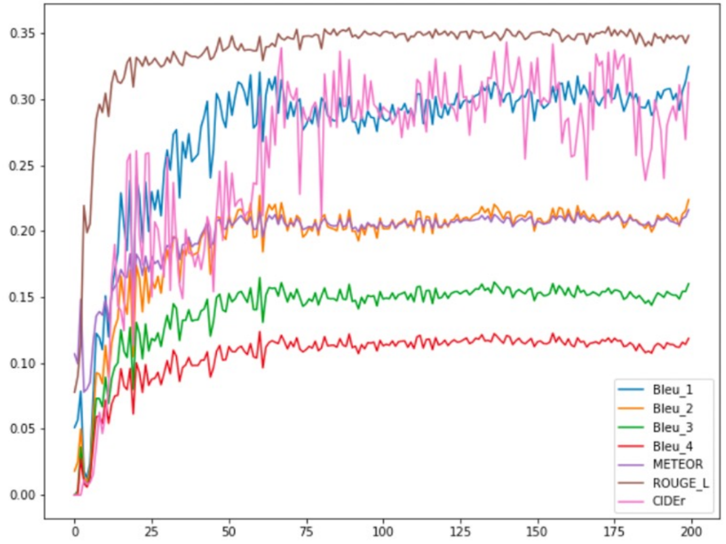
About Visual Encoder: ResNet vs DenseNet

Visual Encoder : Resnet-152(62m) vs Dense-121(7m)

ResNet-152



DenseNet-121

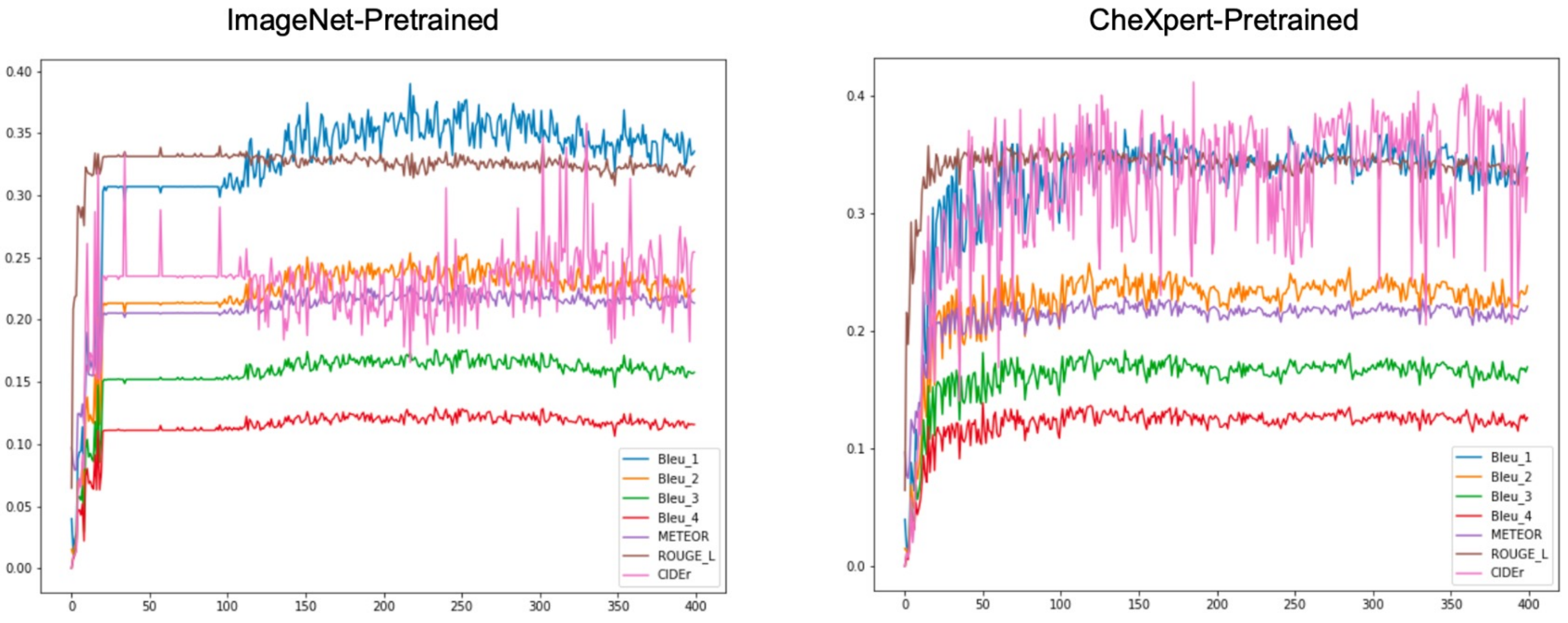


[그림 5] Visual Encoder 로 ResNet-152 를 사용했을 때와 DenseNet-121 를 사용했을 때의 비교.

Contrastive Image Captioning Model

About Visual Encoder : Pre-trained dataset

Pretrained : ImageNet vs CheXpert



[그림 6] Visual Encoder(DenseNet-121)를 ImageNet 에 사전학습 했을 때와 흉부 X-ray dataset 인 CheXpert 에 사전학습 했을 때의 결과 비교.

Contrastive Image Captioning Model

About Visual Encoder : Diversity of Generated Sentence

Frequency of overlapped sentences

ImageNet-Pretrained(47/110)


```
h stable cardiomegaly  2 of 47
[{"image_id": 973, "caption": "pa and lateral views of the chest were provided <eos> midline sternotomy wires and mediastinal clips
are noted compatible with prior cabg <eos> the heart remains moderately enlarged <eos> there is mild pulmonary interstitial edema and
engorged hila <eos> there is no effusion or pneumothorax <eos> no definite sign of pneumonia <eos> bony structures are intact <eos>
no free air below the right hemidiaphragm <eos> mild pulmonary edema with stable cardiomegaly <eod>"}, {"image_id": 974, "caption":
"pa and lateral views of the chest were provided <eos> midline sternotomy wires and mediastinal clips are noted compatible with prior
cabg <eos> there is mild pulmonary interstitial edema and engorged hila <eos> there is no effusion or pneumothorax <eos> no definite
sign of pneumonia <eos> bony structures are intact <eos> no free air below the right hemidiaphragm <eos> mild pulmonary edema with
stable cardiomegaly <eod>"}, {"image_id": 975, "caption": "pa and lateral views of the chest were provided <eos> midline sternotomy
wires and mediastinal clips are noted compatible with prior cabg <eos> the heart remains moderately enlarged <eos> there is mild
pulmonary interstitial edema and engorged hila <eos> there is no effusion or pneumothorax <eos> no definite sign of pneumonia <eos>
bony structures are intact <eos> no free air below the right hemidiaphragm <eos> mild pulmonary edema with stable cardiomegaly <eod>"}
, {"image_id": 976, "caption": "pa and lateral views of the chest provided <eos> there is no focal consolidation effusion or
pneumothorax <eos> the cardiomediastinal silhouette is normal <eos> imaged osseous structures are intact <eos> no free air below the
right hemidiaphragm is seen <eos> no acute intrathoracic process <eod>"}, {"image_id": 977, "caption": "pa and lateral views of the
chest provided <eos> there is no focal consolidation effusion or pneumothorax <eos> the cardiomediastinal silhouette is normal <eos>
imaged osseous structures are intact <eos> no free air below the right hemidiaphragm is seen <eos> no acute intrathoracic process
<eod>"}, {"image_id": 978, "caption": "the lungs are clear without focal consolidation <eos> no pleural effusion or pneumothorax is
seen <eos> the cardiac and mediastinal silhouettes are unremarkable <eos> no acute cardiopulmonary process <eod>"}, {"image_id": 979,
"caption": "pa and lateral views of the chest provided <eos> there is no focal consolidation effusion or pneumothorax <eos> the
cardiomediastinal silhouette is normal <eos> imaged osseous structures are intact <eos> no free air below the right hemidiaphragm is
seen <eos> no acute intrathoracic process <eod>"}, {"image_id": 980, "caption": "pa and lateral views of the chest were provided
<eos> midline sternotomy wires and mediastinal clips are noted compatible with prior cabg <eos> there is mild pulmonary interstitial
edema and engorged hila <eos> there is no effusion or pneumothorax <eos> no definite sign of pneumonia <eos> bony structures are
intact <eos> no free air below the right hemidiaphragm <eos> mild pulmonary edema with stable cardiomegaly <eod>"}, {"image_id": 981,
"caption": "pa and lateral views of the chest provided <eos> there is no focal consolidation effusion or pneumothorax <eos> the
cardiomediastinal silhouette is normal <eos> imaged osseous structures are intact <eos> no free air below the right hemidiaphragm is
seen <eos> no acute intrathoracic process <eod>"}, {"image_id": 982, "caption": "pa and lateral views of the chest were provided
<eos> midline sternotomy wires and mediastinal clips are noted compatible with prior cabg <eos> the heart remains moderately enlarged
<eos> there is mild pulmonary interstitial edema and engorged hila <eos> there is no effusion or pneumothorax <eos> no definite sign
of pneumonia <eos> bony structures are intact <eos> no free air below the right hemidiaphragm <eos> mild pulmonary edema with stable
cardiomegaly <eod>"}, {"image_id": 983, "caption": "pa and lateral views of the chest provided <eos> there is no focal consolidation
effusion or pneumothorax <eos> the cardiomediastinal silhouette is normal <eos> imaged osseous structures are intact <eos> no free
```


CheXpert-Pretrained(2/110)

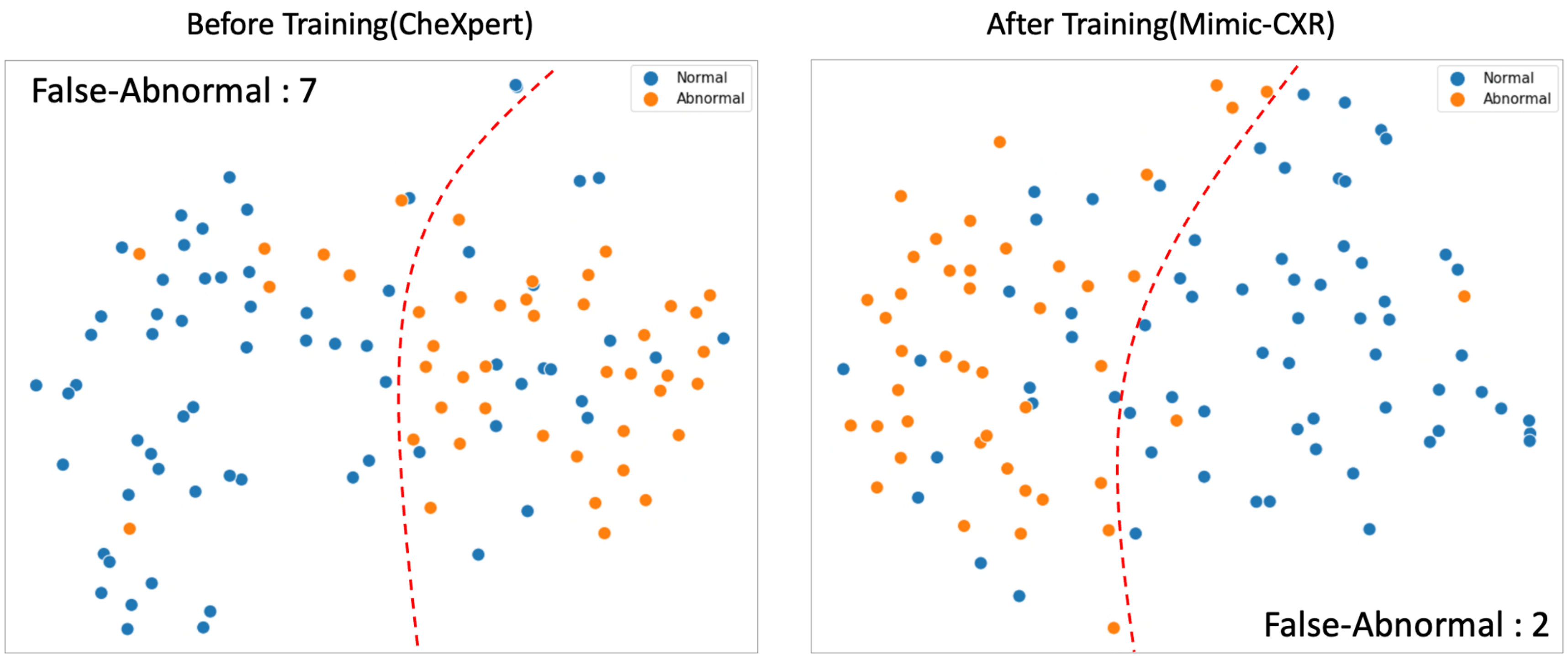

```
intrathoracic process  2 of 2
<eod>"}, {"image_id": 978, "caption": "the lungs are clear without focal co
seen <eos> the cardiac and mediastinal silhouettes are unremarkable <eos> no
"caption": "pa and lateral views of the chest provided <eos> there is no focal consolidation effusion or pneumothorax <eos> the
cardiomediastinal silhouette is normal <eos> imaged osseous structures are intact <eos> no free air below the right hemidiaphragm is
seen <eos> no acute intrathoracic process <eod>"}, {"image_id": 980, "caption": "pa and lateral views of the chest provided <eos>
midline sternotomy wires and mediastinal clips are again noted <eos> there is a small right pleural effusion <eos> there is no
pneumothorax <eos> the heart is mildly enlarged <eos> the mediastinal contour is stable <eos> the bony structures are intact <eos> no
free air below the right hemidiaphragm <eos> mild pulmonary edema with stable cardiomegaly with hilar congestion and mild pulmonary
edema <eos> mild pulmonary edema <eod>"}, {"image_id": 981, "caption": "pa and lateral views of the chest provided <eos> midline
sternotomy wires and mediastinal clips are noted compatible with prior cabg <eos> there is mild pulmonary interstitial edema and
engorged hila <eos> there is no effusion or pneumothorax <eos> the heart appears normal <eos> bony structures are intact <eos> no
free air below the right hemidiaphragm <eos> no acute intrathoracic process <eod>"}, {"image_id": 982, "caption": "pa and lateral
views of the chest provided <eos> midline sternotomy wires and mediastinal clips are again noted <eos> there is a small right pleural
effusion <eos> there is no pneumothorax <eos> the heart is mildly enlarged <eos> the mediastinal contour is stable <eos> the bony
structures are intact <eos> no free air below the right hemidiaphragm <eos> mild pulmonary edema with trace right hemidiaphragm <eos>
mild cardiomegaly with trace bilateral pleural effusions <eod>"}, {"image_id": 983, "caption": "pa and lateral views of the chest
provided <eos> there is no focal consolidation effusion or pneumothorax <eos> the cardiomediastinal silhouette is normal <eos> imaged
osseous structures are intact <eos> no free air below the right hemidiaphragm is seen <eos> no acute intrathoracic process <eod>"},
{"image_id": 984, "caption": "pa and lateral views of the chest provided <eos> there is no focal consolidation effusion or
pneumothorax <eos> the cardiomediastinal silhouette is normal <eos> imaged osseous structures are intact <eos> no free air below the
right hemidiaphragm is seen <eos> no acute intrathoracic process <eod>"}, {"image_id": 985, "caption": "pa and lateral views of the
chest provided <eos> midline sternotomy wires and mediastinal clips are again noted <eos> there is a small right pleural effusion
<eos> there is no pneumothorax <eos> the heart is mildly enlarged <eos> the mediastinal contour appears normal <eos> the bony
structures are intact <eos> no free air below the right hemidiaphragm <eos> mild cardiomegaly with mild pulmonary edema and tiny left
effusion with trace right pleural effusion <eod>"}, {"image_id": 986, "caption": "pa and lateral views of the chest provided <eos>
there is no focal consolidation effusion or pneumothorax <eos> the cardiomediastinal silhouette is normal <eos> imaged osseous
structures are intact <eos> no free air below the right hemidiaphragm is seen <eos> no acute intrathoracic process <eod>"},
{"image_id": 987, "caption": "pa and lateral views of the chest provided <eos> there is no focal consolidation effusion or
pneumothorax <eos> the cardiomediastinal silhouette is normal <eos> imaged osseous structures are intact <eos> no free air below the
right hemidiaphragm is seen <eos> no acute intrathoracic process <eod>"}, {"image_id": 988, "caption": "pa and lateral views of the
chest provided <eos> midline sternotomy wires and mediastinal clips are again noted <eos> there is a small right pleural effusion
<eos> there is no pneumothorax <eos> the heart is mildly enlarged <eos> the mediastinal contour is stable <eos> the bony structures
are intact <eos> mild pulmonary edema <eos> mild pulmonary edema with small right pleural effusion with small right pleural effusion
<eod>"}, {"image_id": 989, "caption": "pa and lateral views of the chest provided <eos> there is no focal consolidation effusion or
```


[그림 7] ImageNet 에 사전학습시킨 Visual Encoder 와 CheXpert 에 사전학습시킨 Visual Encoder 를 사용한 DeXTr 모델들의 결과 비교(110 장의 Test image 에 대해 결과 문장들이 겹치는 모습)

Contrastive Image Captioning Model

About Visual Encoder : Benefit from captioning task(vs freezing)

Visual Feature Embedding : Initial Encoder vs Captioning-trained Encoder

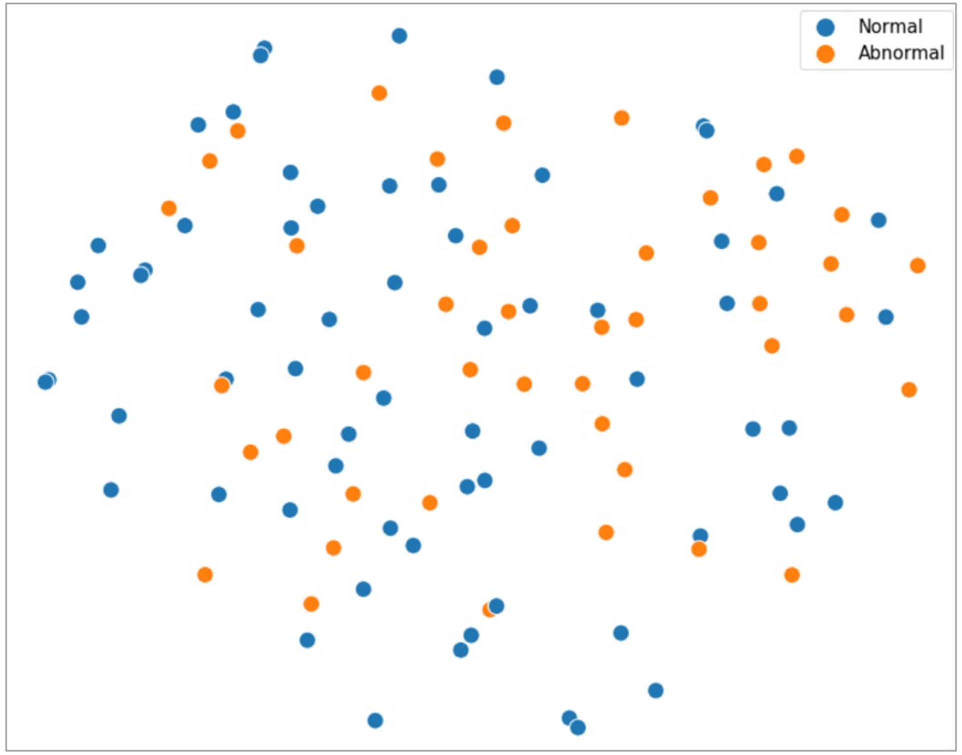


[그림 8] End-to-End Captioning Training 에 따른 Visual Feature Embedding(1024***d***) 비교(T-SNE 사용).

Contrastive Image Captioning Model

About Visual Encoder : Benefit from captioning task(vs freezing)

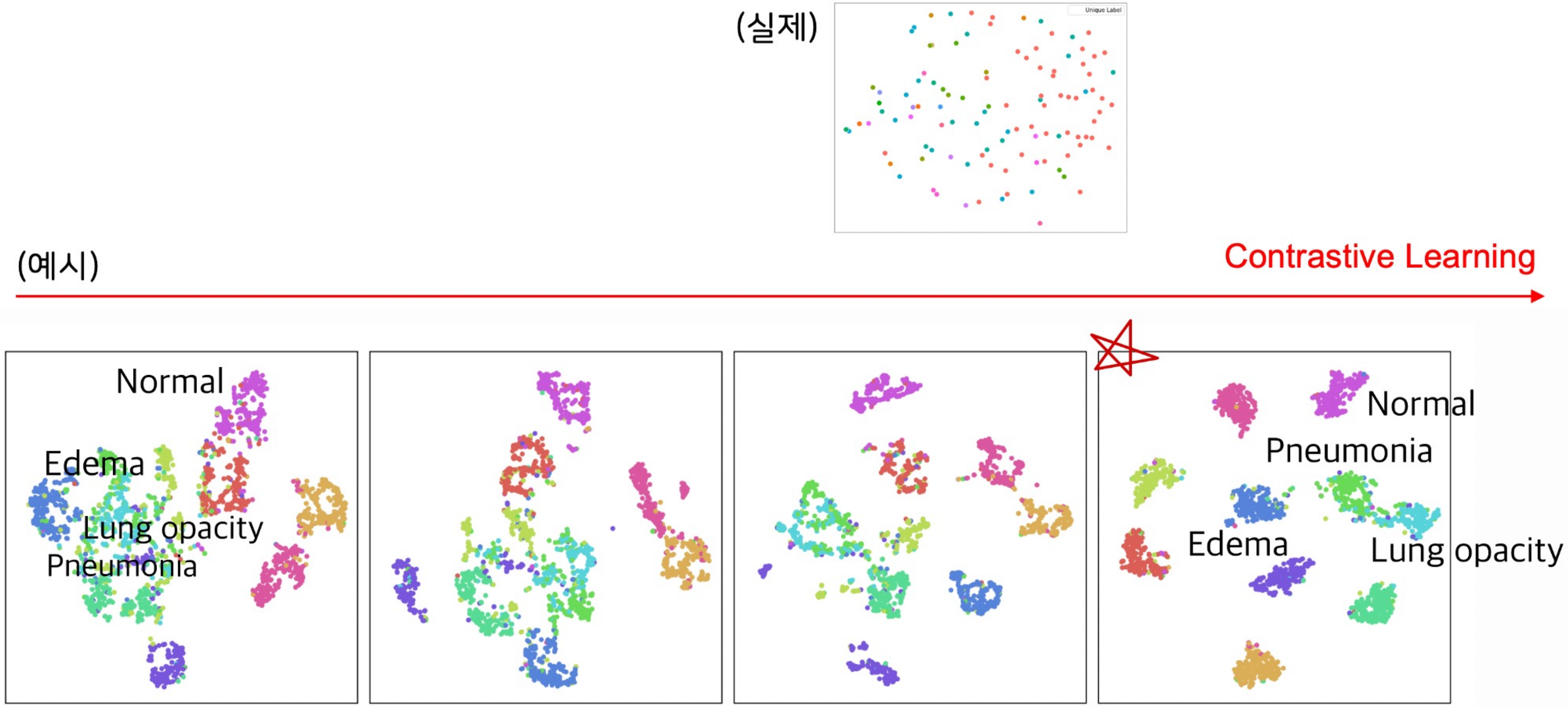
Before Training(ImageNet)



[그림 8] ImageNet 에 사전학습된 Visual Encoder 를 이용할 경우 모델은 X-ray 이미지를 제대로 판단하지 못하는 것을 알 수 있다.

Contrastive Image Captioning Model

About Visual Encoder : Future Direction?



[그림 9] 실제 Class 별 Visual 분포와 이상적인 Visual 분포(참고: Chuang et al(2020))

Contrastive Image Captioning Model

Several point about hyper parameter

A.2. DeXTR-512

End-to-End(Ours)	Bleu-1	Bleu-2	Bleu-3	Bleu-4	METEOR	ROUGE_L	CIDEr	
DeXTr-512+Random (a)	**0.3749**	**0.2573**	**0.1838**	**0.1355**	**0.2300**	0.3430	0.3251	
DeXTr-512+Random+ImageNet	0.3720	0.2492	0.1758	0.1296	0.2251	0.3279	0.2085	Local Minima
DeXTr-512+Random (b)	0.3547	0.2431	0.1759	0.1338	0.2220	0.3524	**0.4050**	Backprop-N,
DeXTr-512+No	0.3563	0.2433	0.1727	0.1265	0.2244	0.3390	0.2627	Backprop-N,
DeXTr-512+Pair	0.3577	0.2454	0.1761	0.1318	0.2230	**0.3469**	0.3579	

[표 6] 정량평가 결과(DeXTr-512)

일반적으로 No, Pair, Random 순으로 성능이 좋았으며, 정상 이미지에 대해 역전파를 수행하지 않으며 임의로 정상 이미지를 샘플링하는 DeXTr-512+Random (a) 모델의 성능이 가장 좋았다.

Contrastive Image Captioning Model

Several point about hyper parameter

A.3. DeXTR−512(Contrastive Attention)

End-to-End(Ours)	Bleu-1	Bleu-2	Bleu-3	Bleu-4	METEOR	ROUGE_L	CIDEr	
DeXTr-512+ CA10+concat	0.3546	0.2480	0.1789	0.1332	0.2279	0.3475	0.2556	CA → Embed
DeXTr-512+CA10+no-concat (x2)	0.3676	0.2501	0.1791	0.1343	0.2248	0.3436	0.2876	CA → Embed
DeXTr-512+CA30+concat (a)	0.3683	0.2555	0.1843	0.1375	0.2314	0.3523	0.3689	CA → Embed
DeXTr-512+CA30+concat (b)	0.3531	0.2417	0.1771	0.1355	0.2182	0.3539	**0.4036**	Embed → CA
DeXTr-512+CA30+no-concat	0.3630	0.2504	0.1815	0.1363	0.2217	0.3469	0.3431	Embed → CA
DeXTr-512+CA30(BiP)+no-concat + Dropout	0.3304	0.2347	0.1759	0.1353	0.2180	0.3518	0.3308	Embed → CA
DeXTr-512+CA30+no-concat+Dropout	0.3794	0.2620	0.1901	0.1430	0.2301	0.3474	0.3835	Embed → CA
DeXTr-512+CA30+concat+Dropout	**0.3925**	**0.2712**	**0.1965**	**0.1475**	**0.2346**	0.3500	0.3690	Embed → CA
DeXTr-512+CA50+no-concat+Dropout	0.3531	0.2515	0.1844	0.1389	0.2252	**0.3544**	0.2574	Embed → CA

- CA30 > CA10 > CA50
- Concat input token > No concat
- Dot-Product Attention > Bi-Linear Pooling Attention

Contrastive Image Captioning Model

Several point about hyper parameter

A.4. DeXTR-768

End-to-End(Ours)	Bleu-1	Bleu-2	Bleu-3	Bleu-4	METEOR	ROUGE_L	CIDEr	
DeXTr-768+CA10+no-concat+Dropout	0.2986	0.2084	0.1493	0.1099	0.2033	0.3323	0.2642	Embed → CA
DeXTr-768+CA10(BiP)+no-concat+Dropout	0.3460	0.2422	0.1780	0.1364	0.2109	0.3456	0.3330	Embed → CA
DeXTr-768+Random+no-concat+Dropout	0.3707	0.2541	0.1803	0.1310	0.2260	0.3355	0.2555	Embed → CA

[표 8] 정량평가 결과(DeXTr-768)

DeXTr 모델에서 XTransformer 의 Embedding 차원을 768 로 늘린 DeXTr-768 모델의 경우 DeXTr-512 모델보다 전반적으로 좋지 않은 결과를 반환했다. DeXTr-512 모델에 비해 파라미터 개수가 2 배 가량(152m) 되는 반면, 학습 데이터는 1,000 개 미만으로 굉장히 부족하기 때문에 과적합이 발생했다고 해석할 수 있다.

CONTENTS

Problem

In Medical Domain

- 신뢰도를 최우선으로하는 의학 분야: 모델 결정에 대한 판단 근거와 전향적 평가의 필요성

Interpretability of Medical Deep Neural Network

We need any form of evidence…

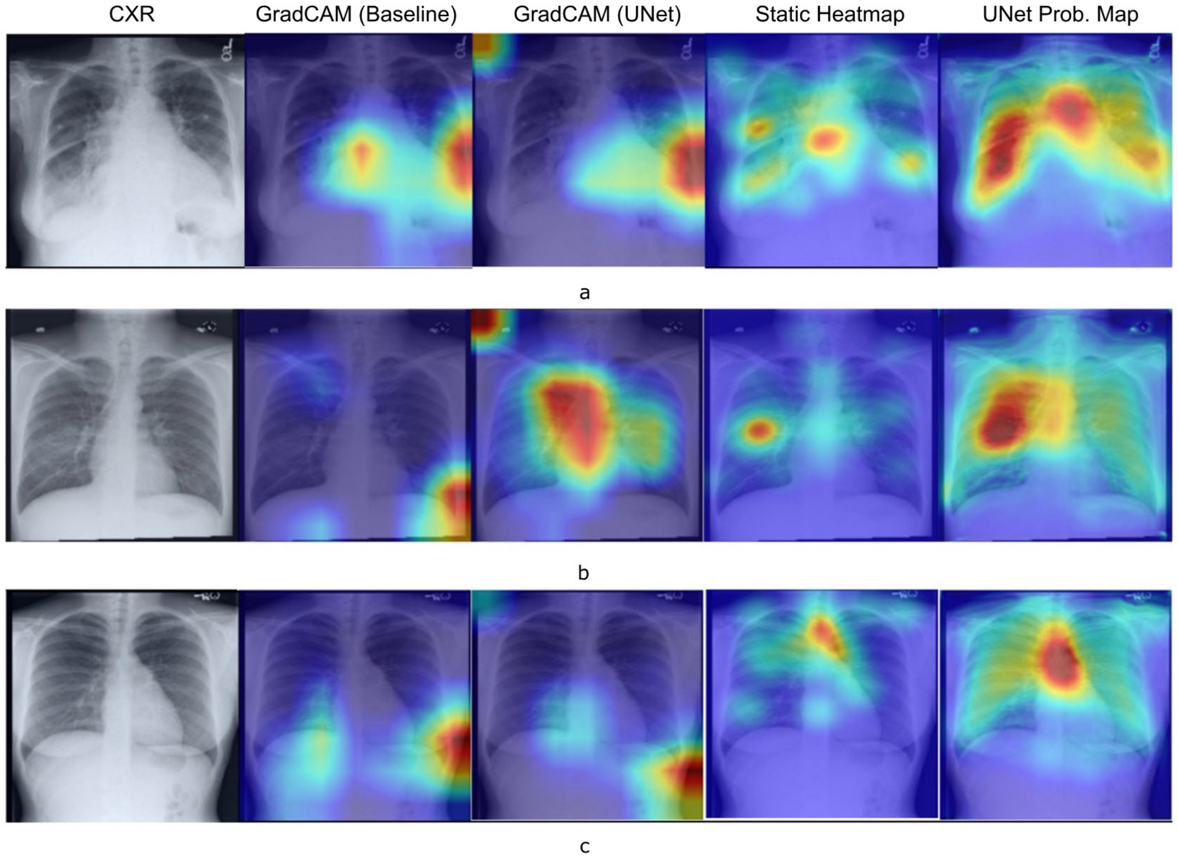


Fig. 17 Qualitative comparison of the interpretability of U-Net based probability maps in comparison with GradCAM for a few example use cases. (**a**) CHF. The physician's eye gaze tends to fall on the enlarged heart

https://www.nature.com/articles/s41597-021-00863-5

Interpretability of Medical Deep Neural Network

We need any form of evidence…

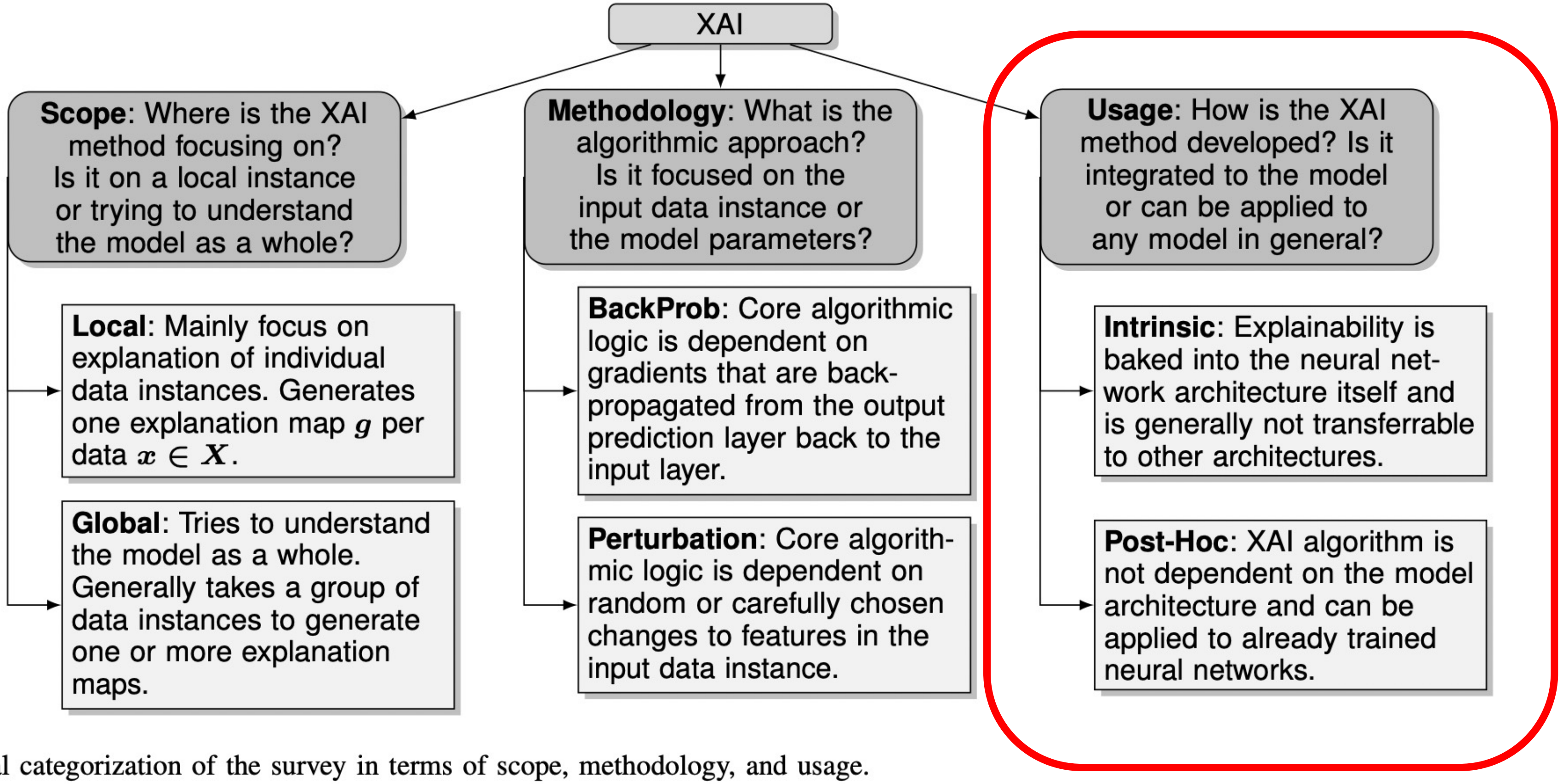
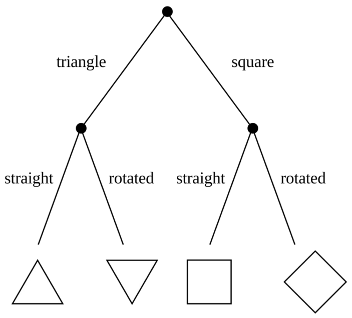


Fig. 1. General categorization of the survey in terms of scope, methodology, and usage.

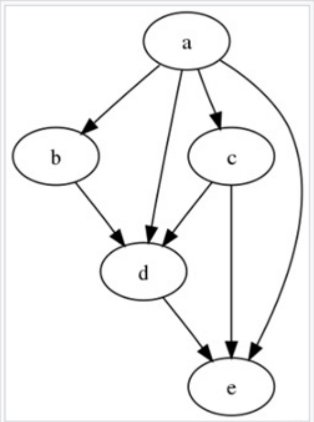
Interpretability of Medical Deep Neural Network

Usage: Intrinsic vs Post-Hoc

Intrinsic



Decision Tree



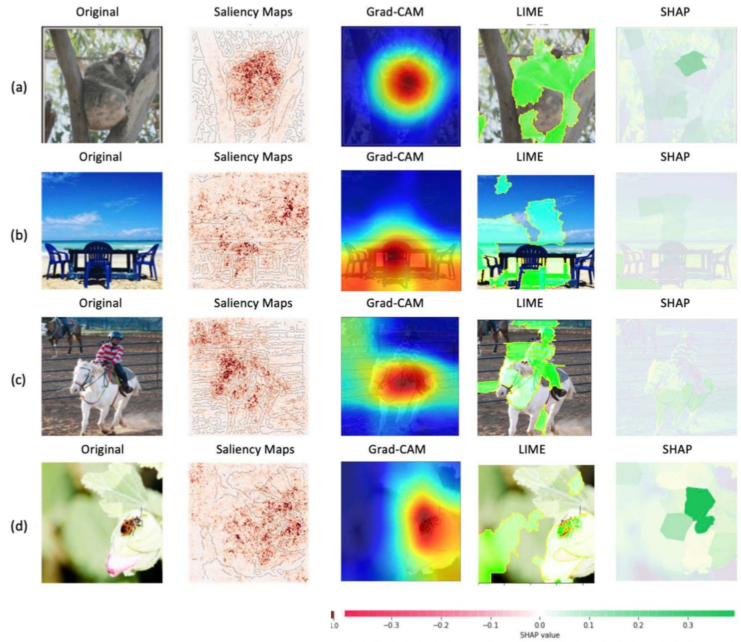
Example of a directed acyclic graph.

Bayesian Model

Adversarial Explanation (후술)

- 설명 능력은 좋지만, 일반적으로 성능이 떨어짐
- 기존의 학습된 모델들을 활용할 수 없음

Post-Hoc



- 성능 저하 없이 설명 가능하나, 설명은 떨어질 수 있음
- 기존에 학습된 모델들을 모두 활용할 수 있음

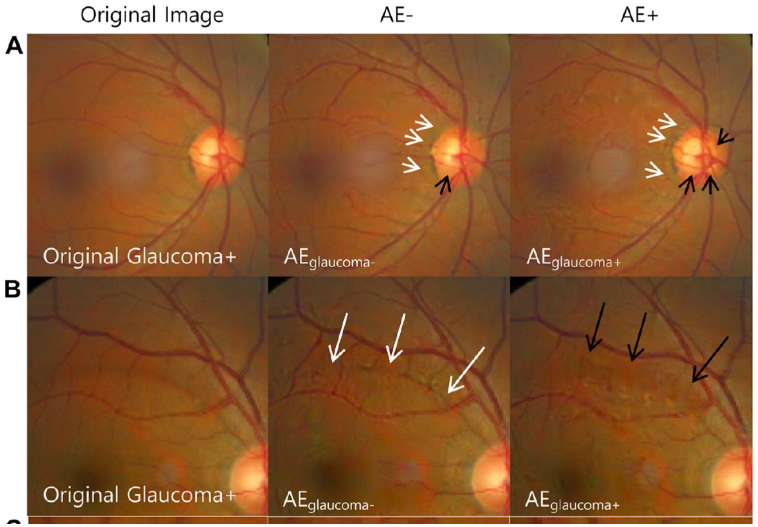
Interpretability of Medical Deep Neural Network

Intrinsic Method : Adversarial Explanation

AMERICAN ACADEMY OF OPHTHALMOLOGY®

Explaining the Rationale of Deep Learning Glaucoma Decisions with Adversarial Examples

Jooyoung Chang, MD,[1] Jinho Lee, MD,[2,3] Ahnul Ha, MD,[2,4] Young Soo Han, MD,[2,4] Eunoo Bak, MD,[2,4] Seulggie Choi, MD,[1] Jae Moon Yun, MD,[5] Uk Kang, PhD,[6,7] Il Hyung Shin, PhD,[6] Joo Young Shin, MD,[8] Taehoon Ko, PhD,[9] Ye Seul Bae, MD,[5,9] Baek-Lok Oh, MD,[2,4] Ki Ho Park, MD, PhD,[2,4] Sang Min Park, MD, PhD[1,5]



Interpretability of Medical Deep Neural Network

Intrinsic Method : Adversarial Explanation

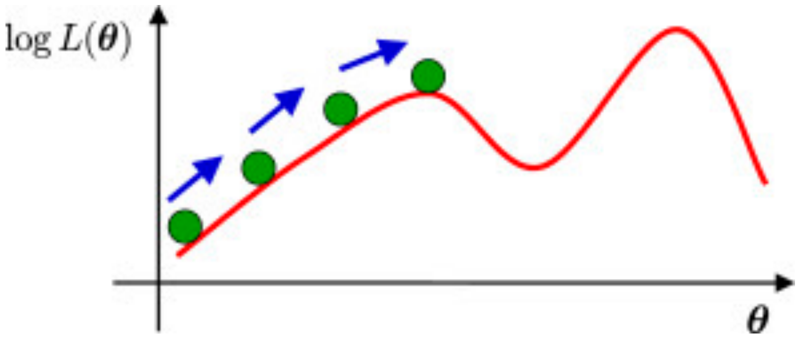
노이즈 Generator를 학습 → 이 노이즈를 활용해 Discriminator 학습 → 노이즈에 강한 Discriminator 탄생

Algorithm 1: Process used to generate adversarial examples.

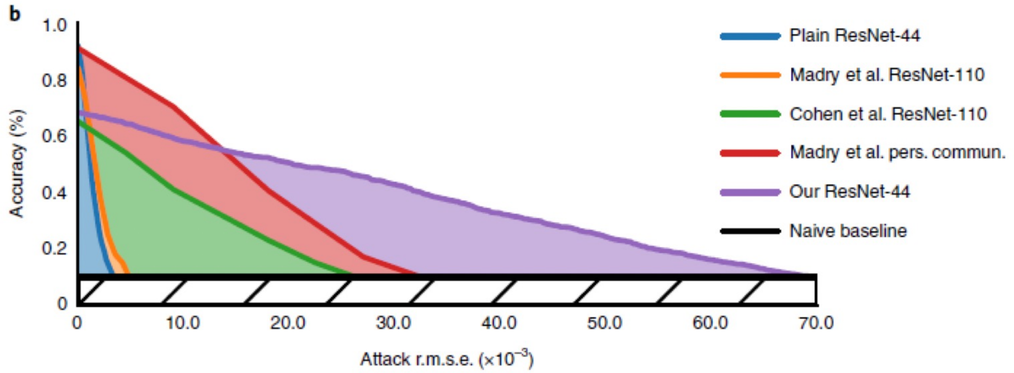
Input: N, the number of attack-optimizing steps; $f(\cdot)$, the NN; x, the network input; t, the true class of the input; $o(\cdot)$, an optimizing method such as SGD with momentum; $g(s,t)$, a goal function returning true if the network outputs from the attack are suitably different from the true class t; η, a balancing term between categorical loss and MSE loss.

Output: δ_{best}, the adversarial noise which satisfies the goal $g(\cdot)$ and has minimal vector length.

begin
- $\delta \leftarrow \vec{0}$
- $M_{best} \leftarrow \inf$
- **for** $n \in [0, ..., N-1]$ **do**
  - $\hat{x} \leftarrow c(x+\delta)$ `// c(·) clips elements of its argument to a valid input range, e.g.` $[0,1]$
  - $y \leftarrow f(\hat{x})$
  - $s \leftarrow softmax(y)$
  - **if** $g(s,t)$ **then**
    - $\Delta\delta \leftarrow 2\delta$ `//` L_2 `loss for magnitude`
    - **if** $||\delta||_2 < M_{best}$ **then**
      - $M_{best} \leftarrow ||\delta||_2$
      - $\delta_{best} \leftarrow \delta$
  - **else**
    - $\Delta\delta \leftarrow \partial s_t / \partial(x+\delta)$
    - $\Delta\delta \leftarrow \eta \frac{\Delta\delta}{||\Delta\delta||_2}$ `// Fixed gradient magnitude`
  - $\delta \leftarrow o(\delta, \Delta\delta)$ `// Apply optimizer step`



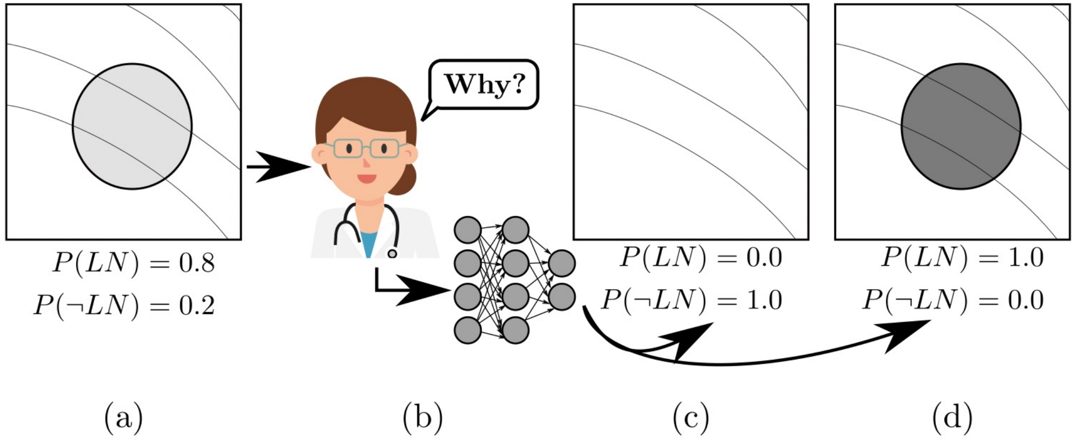
Gradient Ascent를 이용해 모델의 판단을 바꾸는 방향으로 노이즈 학습

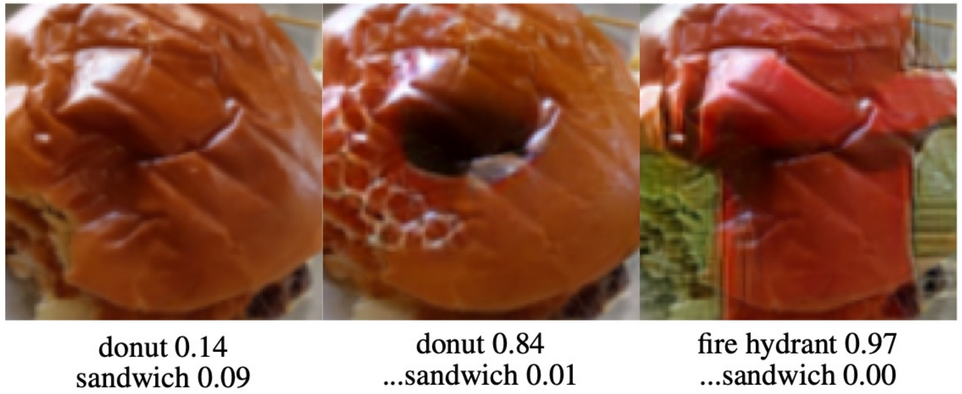


Interpretability of Medical Deep Neural Network

Intrinsic Method : Adversarial Explanation

노이즈 Generator를 학습 → 이 노이즈를 활용해 Discriminator 학습 → 노이즈에 강한 Discriminator 탄생





노이즈에 강하다 → 이미지에 큰 변화를 줄 때만 모델이 판단을 바꾼다

이 성질을 이용해 노이즈 이미지를 일종의 High-level Explanation으로 사용가능

Interpretability of Medical Deep Neural Network

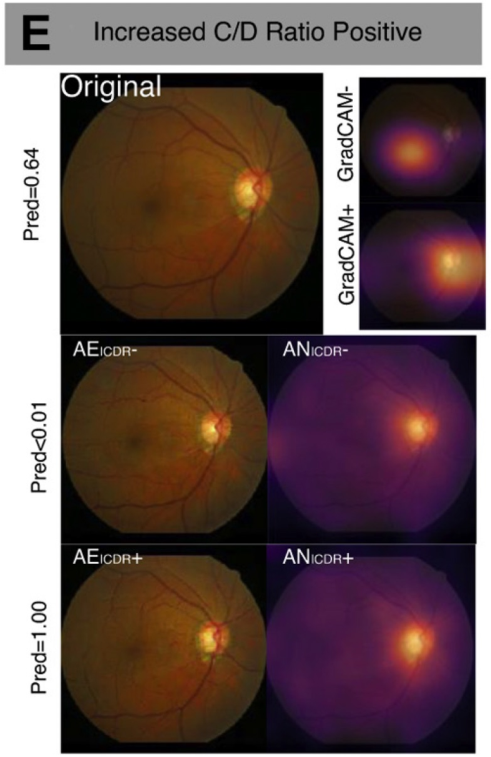
Adversarial Explanation: Good in human study

AE	Grad-CAM
Score for Adversarial Explanation (Mean ± Standard Deviation)	Score for Gradient-weighted Class Activation Mapping (Mean ± Standard Deviation)
3.94 ± 1.33	2.55 ± 1.24
4.14 ± 1.25	3.14 ± 1.20
3.83 ± 1.36	2.24 ± 1.14
3.99 ± 0.93	2.86 ± 1.17
4.17 ± 0.88	2.91 ± 1.22
3.75 ± 0.95	2.79 ± 1.11
4.76 ± 0.56	2.85 ± 1.34
4.87 ± 0.41	3.77 ± 1.05
4.70 ± 0.61	2.37 ± 1.21
4.87 ± 0.47	2.68 ± 1.09
4.84 ± 0.59	3.41 ± 0.87
4.88 ± 0.42	2.44 ± 1.04
2.12 ± 0.92	1.80 ± 1.03
1.93 ± 1.16	2.37 ± 1.15
2.17 ± 0.84	1.66 ± 0.95

> > >

- AE is Good Explanation… But lose classification accuracy(trade-off)
- GradCAM is post-hoc method

→ Can used for all model and don't lose accuracy



In our implementation..

CONTENTS

Problem

- But, 데이터 구축, 보안이슈, 그리고 전문의 협업 등의 문제가 산재

- 잘 갖춰진 (Open) Dataset의 부족
- 의학 도메인 특유의 데이터 이질성(단어, 이미지, …)
  → 전이학습 활용의 어려움

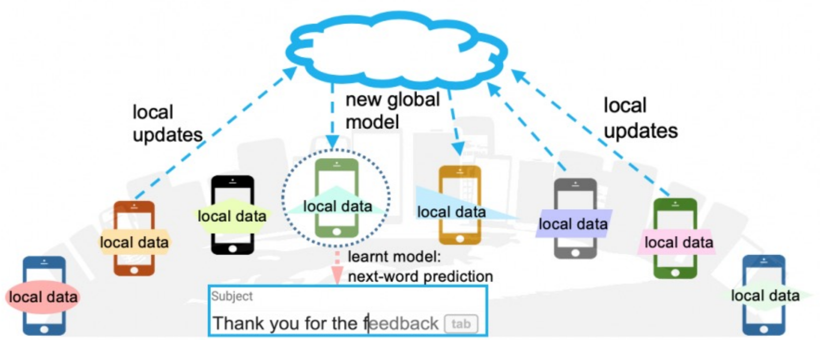
Federated Learning(연합학습)

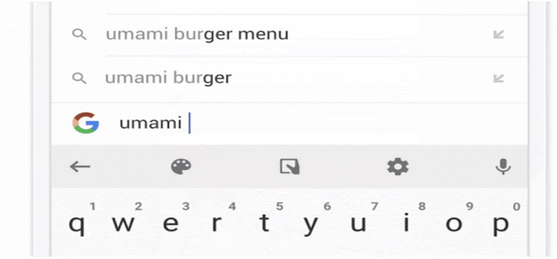
Cross-Device Federated Learning

Google AI Blog
The latest from Google Research

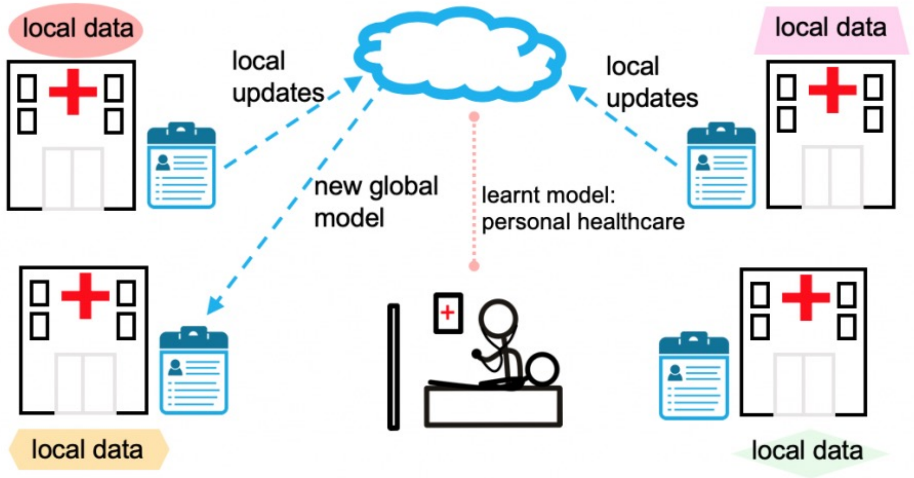
Update local model with local data, Then send local model's information to global model.

Cross-Device





Cross-Silo



Don't send local data to server.
But leverage all data indirectly through global model.
→ 민감한 데이터들의 교환 없이 좋은 모델을 사용가능

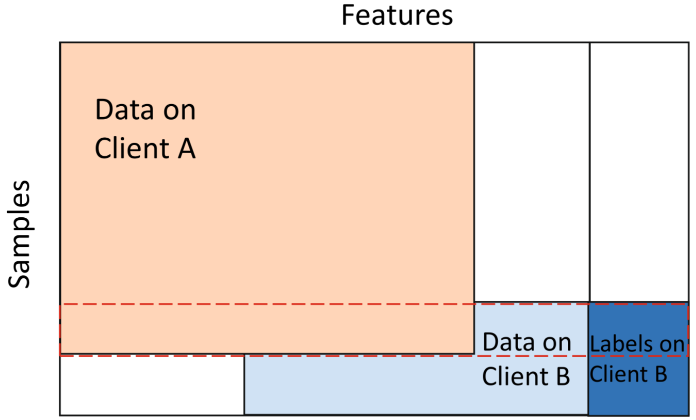
https://blog.ml.cmu.edu/2019/11/12/federated-learning-challenges-methods-and-future-directions/

Federated Learning(연합학습)

Limitation : Non-IID and Heterogenous data distribution

Limitation

- Local Client는 데이터 개수, 도메인, 라벨, feature, Class 개수, 컴퓨팅환경 등이 상이(heterogeneous)함.
- But, optimization algorithm like Stochastic Gradient Descent (SGD) runs on a large dataset partitioned **homogeneously** across servers in the cloud.



(b) Asymmetric federated learning

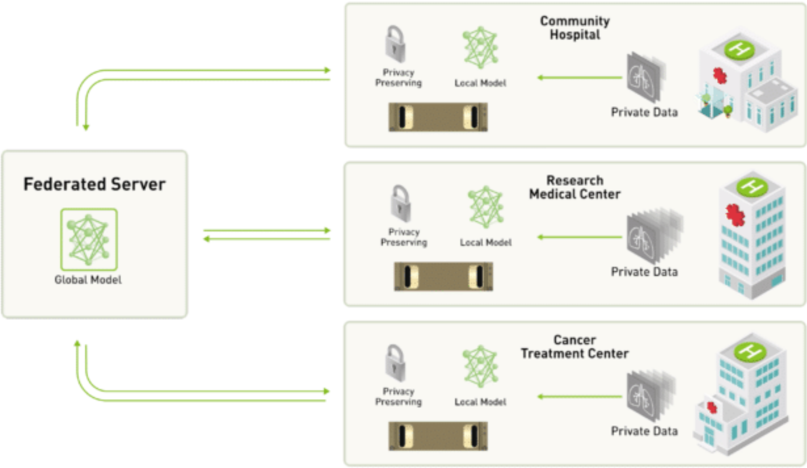
Saeed, et al.(2022), **Federated self-supervised** learning of multi sensor representations for embedded intelligence.

Zhang, Zhe, et al(2021). "**Semi-supervised federated** learning with **non-IID data**: Algorithm and system design."

Huang et al.(2020), "Learn From Others and Be Yourself in **Heterogeneous Federated Learning**."

Federated Learning(연합학습)

Future Direction for data scarcity and privacy



- 한국 경북대병원
- 워싱턴 DC 국립아동병원(Children’s National Hospital)
- NIHR 캠브리지 생물의학연구센터
- 도쿄 자위대 중앙병원(Self-Defense Forces Central Hospital)
- 국립대만대(National Taiwan University) MeDA 연구소
- MAHC 및 대만 국민건강보험청
- 토론토대학(University of Toronto)
- 메릴랜드 베데스다 국립보건원
- 위스콘신대 매디슨 의과 및 공중 보건 대학
- 뉴욕 메모리얼 슬로언 케터링 암센터
- 뉴욕 마운트 시나이 헬스 시스템(Mount Sinai Health System)

https://blogs.nvidia.co.kr/2020/10/08/federated-learning-covid-oxygen-needs/