



저작자표시-비영리-변경금지 2.0 대한민국

이용자는 아래의 조건을 따르는 경우에 한하여 자유롭게

- 이 저작물을 복제, 배포, 전송, 전시, 공연 및 방송할 수 있습니다.

다음과 같은 조건을 따라야 합니다:



저작자표시. 귀하는 원저작자를 표시하여야 합니다.



비영리. 귀하는 이 저작물을 영리 목적으로 이용할 수 없습니다.



변경금지. 귀하는 이 저작물을 개작, 변형 또는 가공할 수 없습니다.

- 귀하는, 이 저작물의 재이용이나 배포의 경우, 이 저작물에 적용된 이용허락조건을 명확하게 나타내어야 합니다.
- 저작권자로부터 별도의 허가를 받으면 이러한 조건들은 적용되지 않습니다.

저작권법에 따른 이용자의 권리는 위의 내용에 의하여 영향을 받지 않습니다.

이것은 [이용허락규약\(Legal Code\)](#)을 이해하기 쉽게 요약한 것입니다.

[Disclaimer](#)

공학석사 학위논문

순차적 흉부 X-선 영상에서
시간에 따른 변화 검출을 위한
인공지능 알고리즘 개발 및 검증

Development and validation of a deep
learning algorithm for longitudinal change
detection in sequential chest X-ray images

2021년 8월

서울대학교 대학원

협동과정 바이오엔지니어링전공

오 동 렬

순차적 흉부 X-선 영상에서
시간에 따른 변화 검출을 위한
알고리즘 개발 및 검증

지도교수 Sungwan Kim

이 논문을 공학석사 학위논문으로 제출함

2021년 6월

서울대학교 대학원

협동과정 바이오엔지니어링전공

오 동 렬

오동렬의 공학석사 학위논문을 인준함

2021년 6월

위 원 장 박 지 훈

부 위 원 장 Sungwan Kim

부 위 원 장 이 경 호

위 원 이 경 준

국문초록

순차적 흉부 X-선 영상에서 시간에 따른 변화 검출을 위한 인공지능 알고리즘 개발 및 검증

오동렬

서울대학교 대학원

협동과정 바이오엔지니어링 전공

최근 그래픽 처리 장치 및 빅데이터가 발전하면서, 의료 영상처리 분야에 인공지능 알고리즘을 접목시켜 주요 질병을 진단 및 검출하는 연구가 활발히 진행되고 있다. 그러나 현재까지 제안된 인공지능 기반의 분류 모델들은 주어진 단일 영상만을 독립적으로 이용하여 결과를 도출한다. 즉, 현재 촬영된 영상은 이전 기록과 잠재적으로 관련이 있음에도 불구하고, 사전에 정의된 비정상 범주만을 예측하는 횡단면적 분석을 시행하는 것이다. 분류 성능은 영상의학 임상과의 수준에 근접하였지만, 병변의 구체적인 변화에 대응하지 못한다. 이는 환자가 이전에 촬영한 영상을 분류하더라도 단순 질병의 출현 유무로는 병변의 변화를 파악하기가 어렵기 때문이다. 특히 일부 주요 질병의 경우, 동일한 질병 내에서도 그 패턴의 종류가 다양할 뿐 만 아니라, 장기간 또는 급성 변화 등

변화 양상이 환자의 임상기록에 따라 매우 다르다. 따라서 횡단면적 분석만으로는 시간에 따른 특정 변화를 검출하는 것은 불가능하기 때문에 종단면적 분석이 함께 요구된다.

본 연구는 주어진 두 영상(전,후) 간의 병변의 특정 변화를 감지하는 새로운 인공지능 알고리즘을 제안한다. 본 알고리즘의 핵심 기법은 정합되지 않은 두 영상의 기하 상관관계도를 구하여 영상 간 변화 유무에 따른 기하 상관관계도 변화 패턴을 파악하고, 변화유무를 이진 분류하는 것이다. 현재까지 종단면적 분석을 위한 기계학습용 참조표준 데이터베이스가 공개된 것이 없기 때문에, 본 연구에서는 영상 판독문을 분석하여 병변의 변화기준을 확립하고, 질환의 종류, 경과시간, 변화 형태 등에 따른 데이터 분류 체계를 구축하여 순차적 흉부 X-선 영상에 대한 참조표준 데이터베이스를 자체적으로 확보하였다.

본 연구는 알고리즘 성능을 객관적으로 분석하기 위하여 수신자조작특성(ROC)의 하의 면적(AUC)을 산출하고, 기존 개발된 알고리즘 및 유사 연구와 정량적으로 비교하였다. 본 연구에서 제안하는 기하 상관관계도를 이용한 알고리즘이 AUC=0.89 (95% 신뢰구간, 0.86-0.92) 및 Youden's index에서의 민감도=0.83, 특이도=0.82으로 가장 뛰어난 성능을 보였다. 또한 주어진 두 영상에서 특정 병변의 변화에 따른 기하 상관관계도를 정성적으로 분석함으로써 실제로 해당 변화가 발생한 위치를 역추적 및 설명할 수 있는 가능성을 제시하였다.

주요어 : 변화 검출, 흉부 X-선, 딥러닝, 기하 상관관계도

학 번 : 2018-21355

목차

국문초록.....	i
목차	iii
그림 목차.....	iv
표 목차	vi
1. 서론.....	7
1.1. 배경	7
1.2. 연구의 목적	9
2. 본론.....	12
2.1. 알고리즘 구조.....	12
2.2.1. 특징 추출	14
2.2.2. 기하 상관관계도.....	14
2.2.3. 이진 분류기.....	16
2.2. 참조표준 데이터베이스	17
3. 결과 및 분석.....	21
3.1. 판독문 가공	21
3.2. 알고리즘 성능.....	22
4. 고찰.....	28
4.1. 실험 결과 고찰	28
4.2. 알고리즘 고찰.....	28
5. 결론.....	30
참고 문헌.....	31
Abstract.....	33

그림 목차

그림 1. 약 1년이 경과한 환자의 (전,후) 영상과 각 영상의 판독 소견. 두 영상에서 관찰되는 병변은 1년동안 거의 변화가 없고 이를 단서로 비슷한 영상에 대하여 다른 진단 결과가 나오는 것을 확인할 수 있음..... 7

그림 2. 미국 ACR(AMERICAN COLLEGE OF RADIOLOGY)의 LUNG-RADS 폐 결절 분류 시스템. 각 카테고리를 구분하는 관찰 요소(FINDING)에 “NEW”, “UNCHANGED”, “SLOWLY GROWING” 등과 같은 병변(결절)의 변화를 나타내는 단어를 확인할 수 있음..... 8

그림 3. 전체 시스템 구조. (A) 각 영상에 대한 FEATURE MAP을 두 개의 CONVOLUTIONAL NEURAL NETWORK 스트림을 이용하여 병렬적으로 추출. (B) 기하 상관관계도 계산. (C) 병변의 변화에 따른 기하 상관관계도 내 특정 패턴 탐지 및 변화 판별..... 12

그림 4. SENET 블록구조: SQUEEZE MODULE과 EXCITATION MODULE로 구성된 ATTENTION BLOCK이며 일반적으로 CONVOLUTIONAL BLOCK에 추가된다. 14

그림 5. 기하 상관관계도. 4차원($H \times W \times H \times W$)의 기하 상관관계도를 3차원 공간으로 시각화($H \times W \times (H \times W)$)..... 16

그림 6. (왼쪽) 병변의 변화 유무에 따른 기하 상관관계도 내 패턴 변화에 대한 예시. (오른쪽) 변화 유무에 따른 공통된 패턴을 파악하고, 이를 이용한 이진 분류에 대한 예시 (이해를 돕기 위해 선형 분류를 예시로 활용)..... 17

그림 7. 데이터 증폭 예시. (왼쪽) 주어진 두 영상의 순서를 뒤집거나 (오른쪽) 각 영상마다 독립적인 선형 변환을 적용한다. 19

그림 8. 가공되지 않은 판독문으로부터 변화를 서술하는 키워드를 기준으로 각 판독문을 분류한 결과 21

그림 9. 변화 유무에 따른 기하 상관관계도. (1,2열) 흉부 X-선 (전,후) 영상: 빨간색으로 표기된 구역은 병변의 위치. (3열) 3차원으로 시각화 된 기하 상관관계도 (1행) 정상 흉부 및 변화 없음: “NO SIGNIFICANT INTERVAL CHANGE SINCE LAST STUDY” (2행) 비정상 흉부 및 변화 없음: “PULMONARY TB (TUBERCULOSIS), STABLE STATE” . (3행) 비정상 흉부 변화 있음: “IMPROVING STATE OF CONSOLIDATION” . (4행) 비정상 흉부 및 변화 있음: “INCREASED AMOUNT OF PLEURAL

EFFUSION, LEFT” 24

그림 10. 병변의 특별한 변화가 없거나 병변이 존재하지 않는 정상 흉부에 대한 기하 상관관계도, 각 영상의 판독문: “NO SIGNIFICANT INTERVAL CHANGE SINCE LAST STUDY” (1,2열) 흉부 X-선 (전,후) 영상, (3열) 3차원으로 시각화 된 기하 상관관계도 25

그림 11. 병변의 변화가 발생한 경우에 대한 기하 상관관계도. (1,2열) 흉부 X-선 (전,후) 영상, (3열) 3차원으로 시각화 된 기하 상관관계도. 각 행의 영상 판독문은 다음과 같다, (1행): “LARGE AMOUNT OF HAZINESS AND PLEURAL EFFUSION WITH PASSIVE ATELECTASIS IN RIGHT LUNG” (2행): “PROGRESSED LUNG CANCER AT LEFT UPPER LUNG WITH INCREASED PLEURAL EFFUSION SINCE LAST STUDY” (3행) “DECREASING RIGHT PULMONARY EDEMA SINCE LAST STUDY” (4행) “MILD IMPROVEMENT OF CONSOLIDATION IN BOTH LUNGS” 26

표 목차

표 1. 공개되어 있는 세계 최대 규모의 흉부 방사선 영상 데이터베이스 비교[17,18]. 종단면적 분석을 위한 정보(환자ID, 판독문, 변화 여부)를 제시하지 않으므로 본 연구가 목표하는 알고리즘의 개발이 불가능함.	18
표 2. ROC 곡선을 이용한 알고리즘 성능 비교.....	23

1. 서론

1.1. 배경

오늘날 딥러닝 기법의 급속한 발전으로 X-선 영상의 질환 분류 모델 등 컴퓨터 보조 진단 및 검출 연구가 활발히 진행되고 있다. 대부분의 연구들은 자연 이미지에서 그 성능이 철저히 검증된 알고리즘을 의료 영상에 접목시키기 위하여 간단한 수정 작업을 거친 후 주요 질병을 분류한다. 최근 제안된 기법들은 모두 준수한 성능을 보여주고 있으나 [1,2,3], 단일 시점의 영상을 입력으로 하는 횡단면적 분석(cross-sectional analysis)만으로 결과를 도출한다.

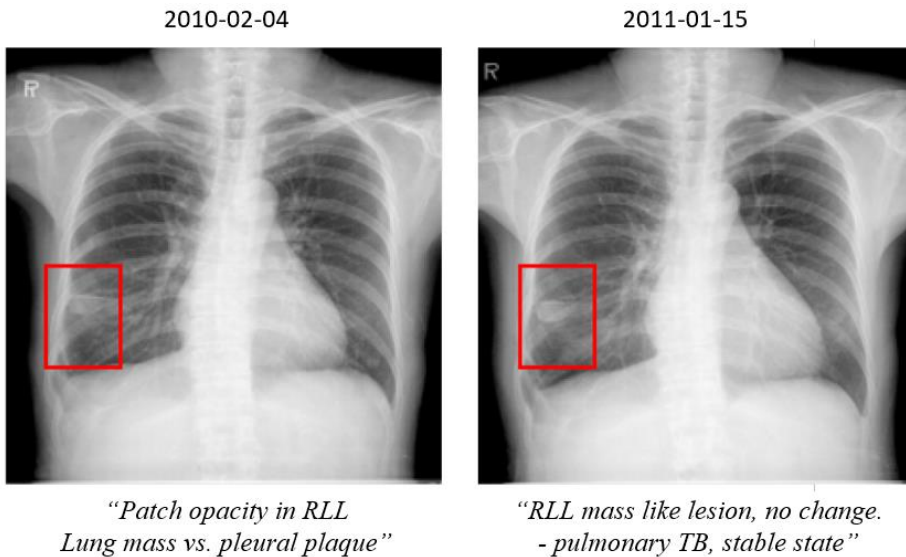


그림 1. 약 1년이 경과한 환자의 (전,후) 영상과 각 영상의 판독 소견. 두 영상에서 관찰되는 병변은 1년동안 거의 변화가 없고 이를 단서로 비슷한 영상에 대하여 다른 진단 결과가 나오는 것을 확인할 수 있음.

그러나 한 시점의 영상만으로는 해당분야 전문의도 정확한 진단을 내리는 것이 매우 어렵다. 대부분 환자의 임상 병력 및 이전

영상과의 비교를 고려한 진단을 하도록 훈련을 받는 관독의들과 달리, 현 딥러닝 알고리즘은 병변의 구체적인 변화에 대응하지 못하고 성능 또한 낮다. 특히 질환 진단 과정에서 새로운 병변의 출현을 포함하여 기존에 감지되었던 병변의 변화를 추적하는 것(그림1)은 매우 중요한 작업이다. 이는 현재 관독에 더 많은 신뢰성을 제공할 수 있기 때문에 환자의 이전 임상 기록과의 비교, 즉, 종단면적 분석(longitudinal analysis)이 항시 요구된다.

Category Descriptor	Lung-RADS Score	Findings	Management
Incomplete	0	Prior chest CT examination(s) being located for comparison Part or all of lungs cannot be evaluated	Additional lung cancer screening CT images and/or comparison to prior chest CT examinations is needed
Negative No nodules and definitely benign nodules	1	No lung nodules Nodule(s) with specific calcifications: complete, central, popcorn, concentric rings and fat containing nodules	Continue annual screening with LDCT in 12 months
Benign Appearance or Behavior Nodules with a very low likelihood of becoming a clinically active cancer due to size or lack of growth	2	Solid nodule(s): < 6 mm new < 4 mm	
		Part solid nodule(s): < 6 mm total diameter on baseline screening Non solid nodule(s) (GGN): < 30 mm OR ≥ 30 mm and unchanged or slowly growing Category 3 or 4 nodules unchanged for ≥ 3 months	
Probably Benign Probably benign finding(s) - short term follow up suggested; includes nodules with a low likelihood of becoming a clinically active cancer	3	Solid nodule(s): ≥ 6 to < 8 mm at baseline OR new 4 mm to < 6 mm Part solid nodule(s) ≥ 6 mm total diameter with solid component < 6 mm OR new < 6 mm total diameter Non solid nodule(s) (GGN) ≥ 30 mm on baseline CT or new	6 month LDCT

그림 2. 미국 ACR(American College of Radiology)의 Lung-RADS 폐 결절 분류 시스템. 각 카테고리를 구분하는 관찰 요소(Finding)에 "new", "unchanged", "slowly growing" 등과 같은 병변(결절)의 변화를 나타내는 단어를 확인할 수 있음.

또한 연간 수익 장에 달하는 방대한 양의 흉부 방사선 영상을 세밀하게 진단할 고급 인력이 매우 부족할 뿐 만 아니라,

검진 영상의 대부분은 이전 영상과 유의미한 차이를 보이지 않기 때문에 ("no significant change since last study") 병변의 변화를 정확하게 감지하여 변화 없는 영상을 제외시켜주는 것만으로도 판독의의 작업량을 획기적으로 줄일 수 있다. CT, MR와 같은 다른 모달리티도 또한 병변의 일정 변화가 해당 진단에 매우 중요한 역할을 하고 있다. 실제로 폐 결절의 위험도를 평가하는 미국의 Lung-RADS 분류 시스템(그림2)의 경우, 같은 크기의 결절이라 하더라도 변화하는 정도에 따라 등급이 상이할 수 있으며 그에 따라 치료/관리 계획이 달라질 수 있다.

1.2. 연구의 목적

본 연구는 흉부 X-선 영상을 대상으로 특정 병변의 생성, 제거 그리고 변화를 종단면적으로 분석을 시도한 세계 최초의 연구이다. 극히 일부분의 연구 논문[4,5]에서 순차적 시점의 영상을 입력으로 하는 분석 결과를 제시하였으나 최종 질환의 분류 정확도를 높이기 위한 목적이며, 병변의 변화를 감지하지 않는다. 따라서 본 연구는 환자의 전, 후 영상을 종단면적 분석하여 병변의 변화의 감지하여 판독의의 작업량을 줄이고, 진단 효율성을 높이는 데에 기여하고자 한다.

본 연구는 현재까지 공개된 바 없는 종단면적 분석을 위한 기계학습용 참조표준 데이터베이스를 자체적으로 설계 및 최초 구축한 연구이다. 특히 일부 질병의 경우, 동일한 질병 내에서도 그 패턴의 종류가 다양하며, 장기간 또는 급성 변화 등 변화 양상이 환자마다 다르다. 즉, 상기한 참조표준 데이터베이스의 부재는 학습

데이터에 의존하는 인공지능 알고리즘의 성능 향상에 부정적인 영향을 끼치므로 횡단면적 분석의 중요성을 인식함에도 불구하고 사전 연구가 쉽지 않았을 것으로 예상된다. 본 연구에서는 가공되지 않은 영상 판독문을 분석하여 해당 병변의 변화기준을 확립하고, 질환의 종류, 경과시간, 변화 형태 등에 따른 데이터 분류 체계를 구축하여 순차적 흉부 X-선 영상에 대한 참조표준 데이터베이스를 자체적으로 확보 하고자 한다.

본 연구에서 개발한 알고리즘을 기반으로 응급환자 자동 선별시스템, 판독조건 자동 생성 등 흉부 X-선 자동 판독기술의 선도적 연구 성과에 기여하는 것이 본 연구의 궁극적인 목표이다.

1.3. 유사 연구사례

딥러닝을 이용하여 주어진 두 흉부 X-선 영상에서 변화를 추적 관찰하는 연구는 소수 진행되어 왔다. 대표적으로 횡단면적 분석 모델에 두 영상을 채널 축으로 결합한 영상을 입력으로 활용하는 것이다.

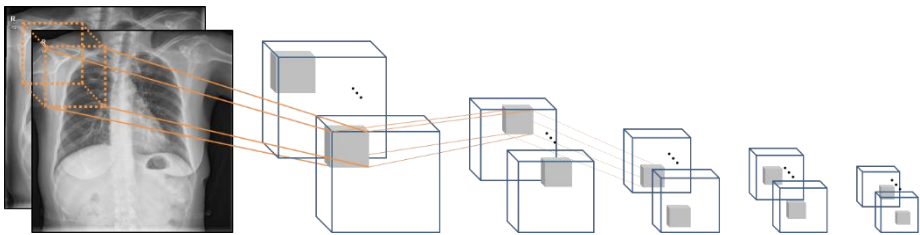


그림 3. 변화 검출을 위한 유사 연구사례1. 기존 횡단면적 분석을 위한 딥러닝모델의 입력 영상에서 기존 채널 축에 비교할 영상을 결합하여 입력으로 활용.

위 구조의 단점은, 합성곱 신경망 특성 상 동일한 위치에 대하여

동일한 필터를 적용한 특징맵을 얻는 다는 점에서 모든 위치에 대한 변화 가능성에 대처하기 어렵다는 것이다. 즉, 두 영상이 사전에 정합이 되어있지 않거나 영상의 품질이 상이한 경우 성능에 부정적인 영향을 끼친다. 합성곱 신경망의 각 레이어를 통과하면서 특징맵 크기가 작아지면서 특징맵 내의 한 픽셀들이 대표하는 입력 영상의 범위가 커짐에 따라, 비교적 전체적인 영상의 특징을 추출하는 것이 가능하다고 볼 수 있으나, 그 과정에서 크기가 작은 특징들은 무시된다. 따라서 폐 결절 또는 석회화 변화 등 비교적 크기가 작은 병변에 대한 변화 검출이 불가능에 가깝다.

단, 새로운 질병의 발생. 기존 질병의 치료 등 질병 분류에 대한 변화를 감지하는 것은 가능하다. 일반적인 질병 분류 네트워크를 이용하여 질병 클래스 변화를 관찰하고, 동일한 환자에 대해서 각 시점 영상 예측결과에 대한 확신도 비교를 통하여 일정 변화를 관찰할 수 있다. 그러나 동일한 병변의 구체적인 변화는 관찰할 수 없고, 학습 전략이 질병 분류이기 때문에 확신도의 변화가 병변의 변화를 설명할 수 없다.

2. 본론

2.1. 알고리즘

본 연구에서는 변화가 정의될 수 있는 최소한의 입력인 영상 2개(이전영상, 현재영상)를 활용한다. 주어진 한 쌍의 영상 (I_0, I_1) 으로부터 변화 감지 문제를 변화 유무에 대한 이진 분류 문제로 재정의하였고, 별도의 분리된 학습과정이 필요 없는 단-대-단 학습(end-to-end learning) 방식의 구조를 제안한다. 본 방법은 두 영상에 대해서 하나의 모델을 적용하기 때문에 두 영상의 상관관계를 학습하는 데에 더욱 효과적이다. 즉, 두 개의 입력으로부터 특정한 관계성에 의하여 딥러닝 모델의 가중치가 변화하게 된다. 또한 단일 모델을 사용하기 때문에 컴퓨터 메모리 측면에서도 유리하다는 장점이 있다.

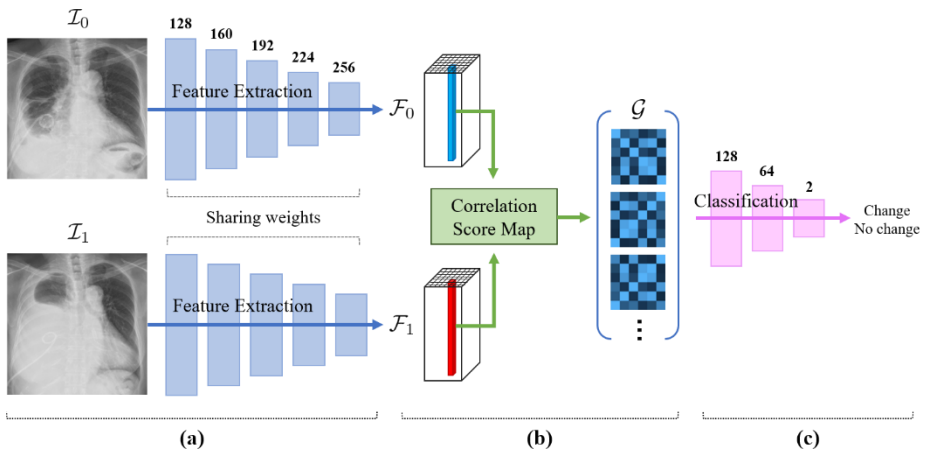


그림 4. 전체 시스템 구조. (a) 각 영상에 대한 feature map을 두 개의 convolutional neural network 스트림을 이용하여 병렬적으로 추출. (b) 기하 상관관계도 계산. (c) 병변의 변화에 따른 기하 상관관계도 내 특정 패턴 탐지 및 변화 판별.

먼저, 두 개의 합성곱 신경망(convolutional neural network)을

통해 각 입력 영상에서 한 쌍의 특징맵(feature map), $(\mathcal{F}_0, \mathcal{F}_1)$ 을 생성한다.

$$\mathcal{F}_{\{0,1\}} \in \mathbb{R}^{h \times w \times d}, \quad h, w, d \in \mathbb{N}$$

이 때 두 신경망은 가중치를 공유하여 동일한 특징맵 추출기로 사용된다. 최초 모델 훈련과정에서는 임의의 특징을 추출하지만, 목적함수(loss function)를 최소화하는 특징맵 추출기로 변화하게 된다. 각 특징맵은 $h \times w$ 해상도를 가지며, 합성곱 신경망을 통과하면서 입력의 해상도보다 작은 해상도를 갖는다. 이는 차원이 축소되면서 최종 출력에 필요한 특징만을 학습하도록 유도하는 효과를 가진다. 합성곱 신경망을 통해 추출된 특징맵은 $h \times w$ 내 각 픽셀에 대한 d 차원의 지역 설명자(local descriptor)의 집합으로 해석될 수 있다.

다음으로, 두 특징맵을 구성하는 지역설명자들 간의 내적을 통하여 기하학적 상관관계도(geometric correlation map)를 생성한다. 본 상관관계도는 입력으로 주어진 두 영상의 유사성으로 해석될 수 있고, 이진 분류기(binary classifier)의 입력이 된다. 따라서 본 연구에서 제안하는 시스템은 입력 영상에서의 특징 변화가 발생했을 때 지역 설명자의 값이 바뀌면서 기하 상관관계도 내의 특정 패턴이 발생하는지 감별한다. 특히 학습과정에서 변화 유무에 다른 기하 상관관계도 내 패턴의 차이를 감별할 수 있도록 가중치가 변화하게 된다. 다음 하위 목차는 각 단계의 세부 사항을 설명한다.

2.1.1. 특징 추출

본 연구는 각 입력 영상으로부터 특징맵을 추출하는 인공지능 알고리즘으로 자연 이미지에서 뛰어난 성능을 보여준 최상위 모델들 [6,7]을 능가한 squeeze-and-excitation network (SENet) [8]를 채택하였다.

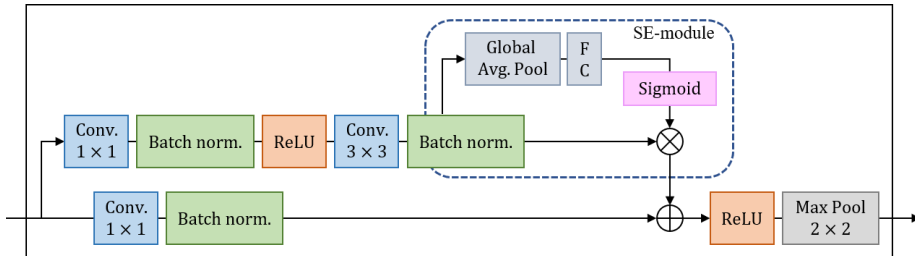


그림 5. SENet 블록구조: squeeze module과 excitation module로 구성된 attention block이며 일반적으로 convolutional block에 추가된다.

총 5개의 블록으로 구성되어 있으며, 각 블록은 입력 특징맵들에 가중치를 부여하여 최종 출력에 대한 중요도를 평가하고 이들을 증폭/감쇠하여 성능 향상을 도모한다. 또한 두 입력 영상에 동일한 인공지능 모델을 적용하면 시각적으로 유사한 영상 패치는 비교적 비슷한 지역설명자를 형성하기 때문에, 모델의 파라미터를 공유하여 각 입력 영상에 동일한 모델을 적용하였다. 특징 추출모델 내 각 블록은 128, 160, 192, 224 및 256 채널의 특징맵을 생성하고, 출력 해상도는 각 블록을 통과할 때마다 입력대비 1/2씩 감소한다.

2.1.2. 기하 상관관계도

추출된 두 특징맵 간의 가능한 모든 조합의 지역 설명자에 대해서 기하 상관 관계성을 계산한다. 매번 환자의 자세, 스캔 각도 등 촬영 조건이 동일하지 않기 때문에, 두 입력 영상은 정렬되어

있지 않다. 따라서 동일한 인덱스의 지역설명자는 실제로 동일한 영상 패치를 가리키고 있다고 볼 수 없으므로 본 연구에서는 가능한 모든 조합을 이용하였다. [9,10]. 즉, 임의의 한 특징맵 \mathcal{F}_0 내 (i,j) 위치의 지역설명자에 대해서 \mathcal{F}_1 의 모든 지역설명자와 상관관계를 계산하게 되면 각 위치당 $h \times w$ 크기의 점수도(score map), $S^{i,j} \in \mathbb{R}^{h \times w}$ 을 생성할 수 있다.

$$S^{i,j}(p,q) = \tau(\mathcal{F}_0(i,j), \mathcal{F}_1(p,q))$$

이 때, τ 는 두 개의 지역설명자 벡터를 입력 인수로 받아서 내적 값을 계산한다. 모든 위치에 대해서 반복하여 계산하면 최종적으로 $h \times w$ 개의 $h \times w$ 크기의 점수도를 얻을 수 있으며 본 구에서는 기하 상관관계도를 칭한다.

$$G = \{S^{1,1}, S^{1,2}, \dots, S^{h,w}\}, G \in \mathbb{R}^{h \times w \times h \times w}$$

내적은 벡터의 크기를 반영하기 때문에 각 크기만큼 나누어 기하 상관관계도를 정규화 하였고, 음의 상관관계는 모두 제거하였다.

$$\tau(v_1, v_2) = \max\left(0, \frac{v_1 \cdot v_2}{\|v_1\| \|v_2\|}\right)$$

두 입력 영상 간의 중 방향 변화에 따라 지역설명자 또한 값이 바뀌므로, 기하 상관관계도 내에서도 특정 패턴의 변화를 찾아볼 수 있다. 특정 변화가 없어도, 흉부 X-선 영상의 촬영 조건은 비슷하기 때문에 정렬이 크게 틀어져 있지 않아서 동일한 위치의 지역중심자는 강한 상관관계를 보이는 경향이 있다.

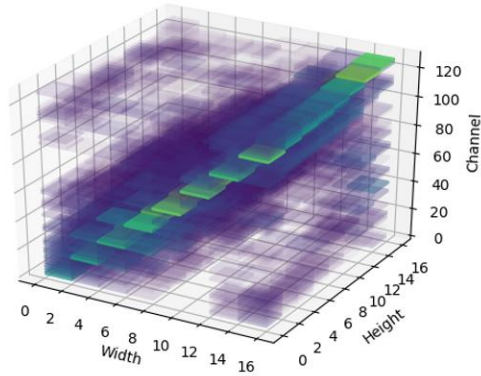


그림 6. 기하 상관관계도. 4차원($h \times w \times h \times w$)의 기하 상관관계도를 3차원 공간으로 시각화($h \times w \times (h \times w)$).

예를 들어, 그림 5의 기하 상관관계도는 $h \times w$ 크기의 점수도 $S^{i,j}$ 를 $h \times w$ 개로 쌓아 3차원으로 시각화 한 그림이다. 동일 위치 (i, j)의 밝은 색상은 강한 상관관계를 나타내며 빈 공간은 사전에 제거된 음의 상관관계를 나타내고, 짙은 색상일수록 약한 상관관계를 나타낸다.

2.1.3. 이진 분류기

입력 영상에 특정한 변화가 생길 경우, 지역 설명자 값이 변화함에 따라 기하 상관관계도 내의 특정 양의 관계가 생성 또는 일부 제거될 수 있다. 변화 위치에 따라 관계도 내 값이 바뀌는 위치 또한 달라지게 되므로 이러한 공간적 패턴을 잘 감지할 수 있도록 설계된 이진 분류기로 FCN (fully convolutional network) [11]를 채택하였다. 총 3개의 합성곱 블록으로 구성되어 있으며, 각 블록은 128, 64, 2 채널의 특징맵을 생성하고, 마지막 블록의 채널은 변화 유무를 의미한다.

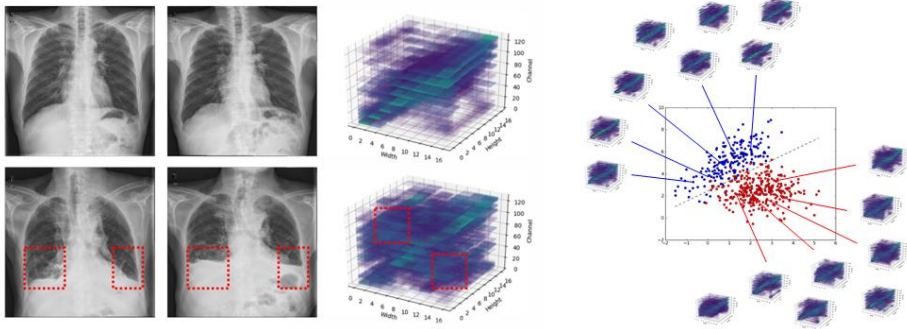


그림 7. (왼쪽) 병변의 변화 유무에 따른 기하 상관관계도 내 패턴 변화에 대한 예시. (오른쪽) 변화 유무에 따른 공통된 패턴을 파악하고, 이를 이용한 이진 분류에 대한 예시 (이해를 돕기 위해 선형분류를 예시로 활용).

변화가 없는 두 흉부 X-선 영상의 기하 상관관계는 동일한 인덱스의 지역 설명자를 중심으로 양의 상관계수가 일정한 양상을 보여야 하는 반면, 임의의 변화가 있는 경우에 변화가 발생한 위치에서 상관계수의 분포가 변한다 (그림6). 따라서 이진 분류기는 상관관계도가 변화하는 공간적 패턴을 감지하여 변화유무를 잘 예측하도록 학습된다.

2.2. 참조표준 데이터베이스

현재 공개된 세계 최대 규모의 흉부 X-선 영상 데이터셋 [12,13]들은 모두 종 방향 변화를 검출하는 인공지능 알고리즘을 개발하는 데에 충분한 정보를 제공하지 않는다 (그림6). 즉, 제공되는 단일 영상의 개수와 해당 영상의 판독정보가 많다고 하더라도 한 환자에 대한 2개 이상의 순차적 영상과 시간에 따른 질병의 변화 등을 설명하는 판독정보가 존재하지 않는다. 가공되지 않은 영상 판독문으로부터 질병을 분류하는 작업은 상대적으로 많은 노동력을 요구하며, 동일 환자에 대한 X-선 촬영 기록에서

일정 변화를 의미하는 시퀀스를 추출하는 것이 어렵기 때문이다.

표 1. 공개되어 있는 세계 최대 규모의 흉부 방사선 영상 데이터베이스 비교. 중단면적 분석을 위한 정보(환자ID, 판독문, 변화 여부)를 제시하지 않으므로 본 연구가 목표하는 알고리즘의 개발이 불가능함.

Dataset	ChestXray14 [13]	CheXpert [14]
Publisher	NIH	Standford Univ
# of images	112,120	224,316
# of diseases	14	14
Study date	Available	Available
Patient ID	N/A	N/A
Report	N/A	N/A
Change info.	N/A	N/A

대부분 영상의학 전공 의사들의 판독문 기록은 그 형식이 어느 정도 갖추고 있으나, 몇몇 단어들은 편의를 위해 제외, 추가, 수정되었고, 오타자가 있거나 단어배열이 일정하지 않는 등 형식이 통일되지 않는 부분이 상당 존재한다. 또한 X-선 영상의 시퀀스를 생성하기 위해서 일반적인 조합을 이용하여 쌍을 구성하게 되는데, 이 때 질병 명칭만 사용하는 경우 [14], 해당 질병이 악화되었는지, 완화되었는지 등의 “변화”와 “변화없음”을 구분하기가 어렵다.

본 연구에서는 3차 병원에 저장된 흉부 X-선 영상을 기반으로 참조표준 데이터베이스를 자체 구축하였다. 순차적 영상을 구축하기 위하여 X-선 촬영 기록이 최소 2회 이상인 환자 영상과 판독문을 추출하였다. 추출 과정에서 판독문을 일부 반복되는 고유문장을 중심으로 분해하고, 공백, 구두점, 개행 등의 사소한 변형을 제거하고 각 문장의 빈도수를 계산하였다. 고유 문장에서 시간에 따른 변화를 서술하는 키워드를 탐색하여 해당 키워드의

존재 유무에 따라 판독문을 분류하였고, 분류된 영상을 정성적으로 검증하여 구분이 명확하고 일관된 데이터셋을 구축하였다.

2.3. 학습 전략

본 연구에서 사용되는 (전,후) 영상은 시간 차이가 존재하고, 동일한 환경에서 촬영된다는 보장이 없다. 예를 들어, 같은 환자로부터 얻은 영상이라고 하더라도, 당시 촬영 환경에 따라 영상의 품질과 속성이 매우 상이하다. 특히 의료 영상 품질은 프로토콜의 변화에 민감하기 때문에, 한정된 데이터셋만으로는 이를 대처하기 어렵다. 따라서 본 연구에서는 다음 두 가지 학습 전략을 제시한다.

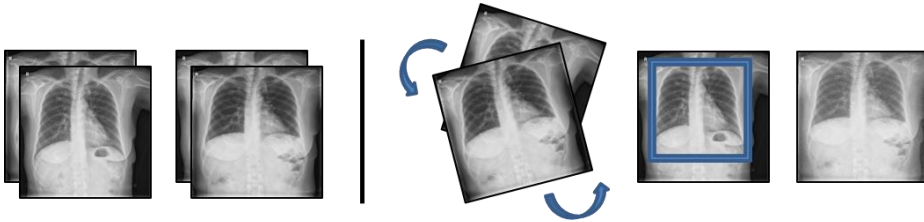


그림 8. 데이터 증폭 예시. (왼쪽) 주어진 두 영상의 순서를 뒤집거나 (오른쪽) 각 영상마다 독립적인 선형 변환을 적용한다.

첫 번째, 영상 촬영기기 및 방사선량 변경 등의 촬영 프로토콜의 변화로 인하여 영상 품질이 달라지는 경우를 대비하여 히스토그램 매칭(Histogram matching)을 적용한다. 히스토그램 매칭을 위한 참조 영상은 주어진 두 영상을 포함한 학습 데이터 중 임의의 한 영상으로 선택한다. 이는 기존보다 다양한 히스토그램에 대해 학습할 수 있을뿐 만 아니라, 다양한 히스토그램 조합에서 병변의 변화를 감지할 수 있도록 유도한다.

두 번째, 환자의 자세, 호흡 또는 촬영 시야 (Field of View; FOV)에 따라 영상의 회전 및 원근감이 달라지는 경우를 대비하여 데이터 증폭(Data augmentation) 기법을 적용한다. 각 영상마다 독립적으로 단순 확대 및 축소, 회전, 평행이동 등을 포함한 아핀 변환(Affine transformation)을 적용한다. 또한 랜덤으로 밝기를 조정하거나, (전,후) 영상을 뒤집어서, 보다 다양한 환경의 영상을 학습할 수 있도록 유도한다.

3. 결과 및 분석

3.1. 판독문 가공

판독문으로부터 반복되는 고유 문장 수로 분해한 결과, 총 252,209개이며, 이 중에서 590개 (0.2%)만이 50회 이상 판독문에 제시되었다. 해당 문장을 포함하는 판독문의 수는 약 140만개로 전체 보고서의 77%를 초과하였다. 가장 자주 제시된 문장은 “no active lesion in the lung”로 약 47만개의 판독문에 출현하였다.

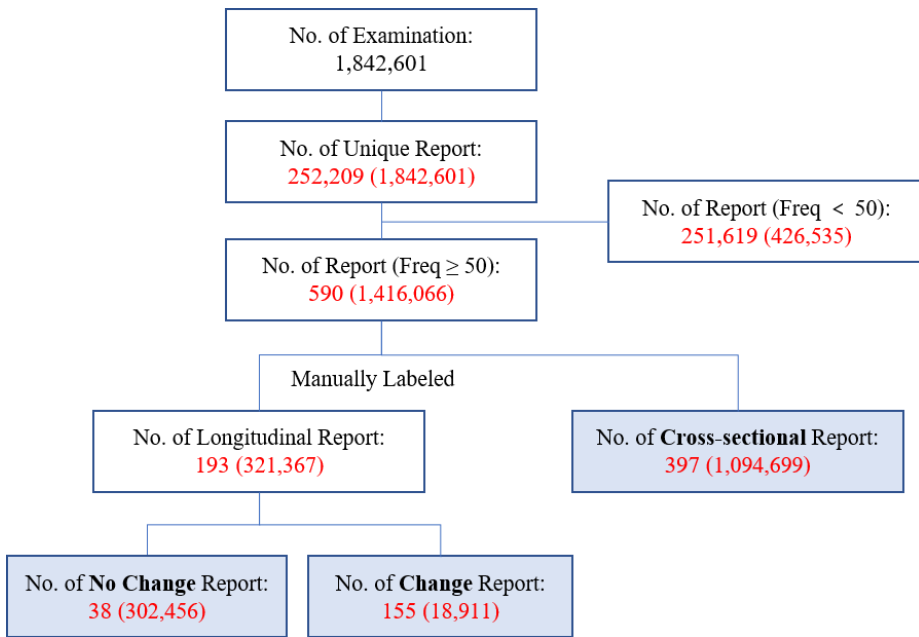


그림 9. 가공되지 않은 판독문으로부터 변화를 서술하는 키워드를 기준으로 각 판독문을 분류한 결과

590개 문장을 전체 조사하여 시간 관련 키워드(예: aggravation, increment, disappear, decrease, stable, change 등)가 포함된 문장 193개를 채택하였다. 해당 193개 문장을 “변화 있음”에 대한 155개 문장, “변화 없음”에 대한 38문장으로 총 2가지 클래스로

구분하였다. 영상에 변화가 있는 경우의 출현빈도가 높은 문장은 “decreased amount of bilateral pleural effusion”, “improving pulmonary edema”, “mild improvement of consolidation in both lungs” 등이 존재하였고, 변화가 없는 경우에는 “no interval change since last study”, “no change of stable tuberculosis”, “emphysema, no interval change” 등이 존재하였다. 이 때, 각 영상 촬영간격이 최소 30일 이상인 경우는 변화가 있는 클래스 1,751개쌍, 변화가 없는 클래스 3,271개쌍으로, 총 5,472개쌍의 영상이 본 연구의 실험 데이터셋으로 활용되었다. 데이터셋 내의 질병은 pleural effusion, pulmonary edema, pneumothorax, pleural thickening, haziness 등이 포함되었다.

3.2. 알고리즘 성능

본 연구에 활용된 데이터셋 총 5,472개 쌍의 영상을 training set (4,370개), validation (551개) 및 test set (551개)로 무작위 분배하였다. 본 연구에서 제안하는 알고리즘의 성능 평가는 test set에서만 진행하였고, 모든 영상은 256×256 크기로 조정 및 정규화 하였다. 원활한 기계학습을 위하여 고성능 그래픽장치가 장착된 서버를 이용하였다. Validation set을 이용하여 중간 결과를 평가한 뒤, 평가 결과에 따라 파라미터를 수정하는 등의 개선과정을 반복하였다. 최종 진단 정확도 평가를 위해서, test set에 대해서 예민도, 특이도를 수신자판단 곡선(ROC)을 이용하여 분석하였다.

횡단면적 분석을 시행하는 네트워크는 본 연구의 목표를 달성할 수 없기 때문에 알고리즘 비교 대상 선정에서 제외되었고,

공정한 평가를 위하여 모두 동일한 데이터셋을 활용하였다. 비교 알고리즘은 단일 스트림과, 이중 스트림 구조에서 두 특징맵의 기하 상관관계를 산출하는 알고리즘을 t-LSTM [4], channel-wise concatenation [15]으로 대체 한 것, 그리고 본 연구에서 제안하는 알고리즘까지 총 4가지로 모델로 구성하였다. 단일 스트림 구조의 경우는 baseline 모델로서, 두 영상의 입력을 2채널로 구성된 단일 입력으로 인식하여 하나의 특징맵을 추출하고 별도의 비교연산 없이 변화 유무를 판별하는 알고리즘이다.

표 2. ROC 곡선을 이용한 알고리즘 성능 비교

Model	AUC (95% CI)
Single stream (baseline)	0.77 [0.72, 0.81]
Two-stream + t-LSTM	0.78 [0.72, 0.81]
Two-stream + channel-wise concatenation	0.82 [0.78, 0.86]
Two-stream + geometric correlation map	0.89 [0.86, 0.92]

표 1은 95% 신뢰 구간으로 수신자조작특성(ROC) 곡선의 하의 면적(AUC)으로 측정된 성능 비교표이다. 단일 스트림 모델은 영상 비교에 대한 직접적인 디자인이 없었기에 가장 낮은 성능을 보였다. 많은 연구들이 시퀀스 데이터 분석에 채택하는 LSTM을 사용한 모델이 그 다음으로 성능이 낮았다. 이는 시간 해상도가 중요한 LSTM의 입력 영상이 2개로 고정되어 있기 때문이다. 오히려 간단한 channel-wise concatenation 전략으로도 실질적인 성능 향상이 있었다는 점을 비추어 볼 때, 2개 이상의 독립된 특징맵을 추출하는 이중 스트림의 구조가 비교연산을 내재하고 있으며, 비교를 효율적으로 할 수 있도록 학습된다는 것을 관찰할

수 있다. AUC=0.89 (95% 신뢰구간, 0.86-0.92) 및 Youden's index에서의 민감도=0.83, 특이도=0.82으로 가장 뛰어난 성능을 보였다.

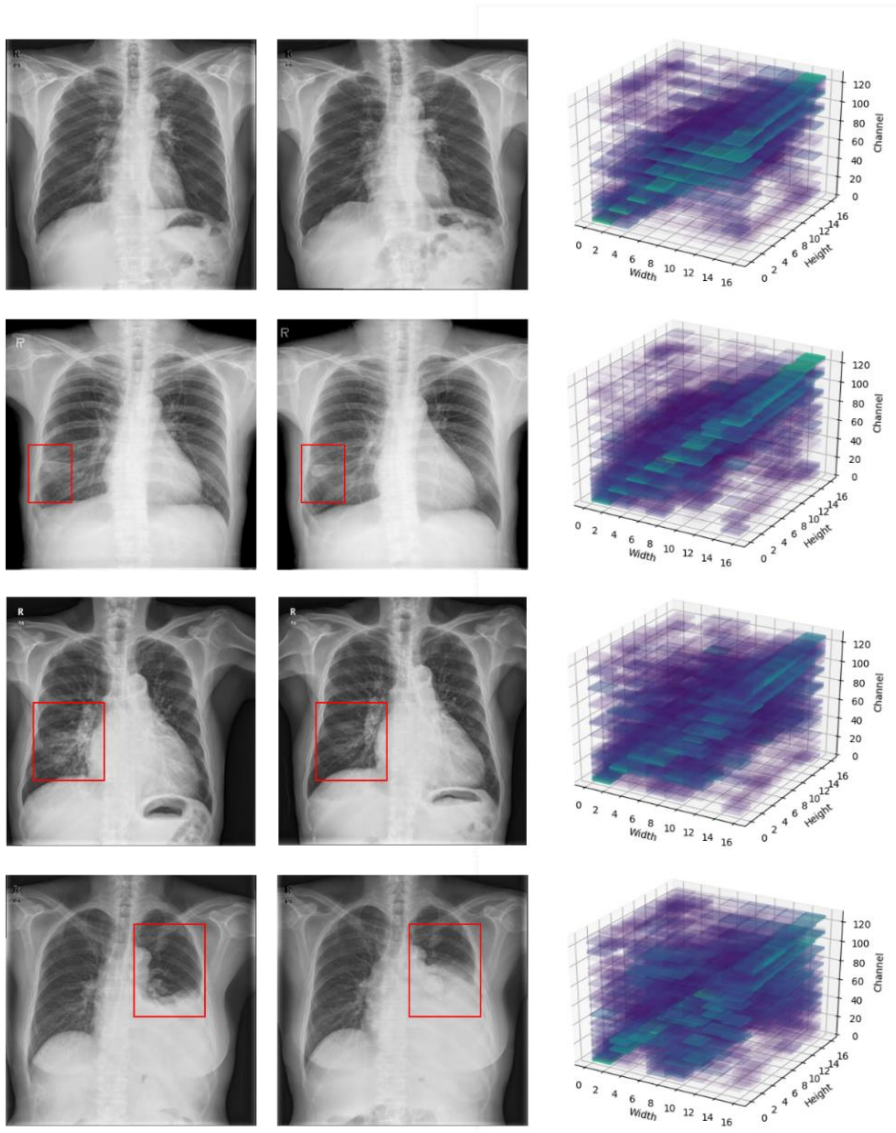


그림 10. 변화 유무에 따른 기하 상관관계도. (1,2열) 흉부 X-선 (전,후) 영상: 빨간색으로 표기된 구역은 병변의 위치. (3열) 3차원으로 시각화 된 기하 상관관계도 (1행) 정상 흉부 및 변화 없음: “no significant interval change since last study” (2행) 비정상 흉부 및 변화 없음: “pulmonary TB (tuberculosis), stable state”. (3행) 비정상 흉부 변화 있음: “improving state of consolidation”. (4행) 비정상 흉부 및 변화 있음: “increased amount of pleural effusion, left”.

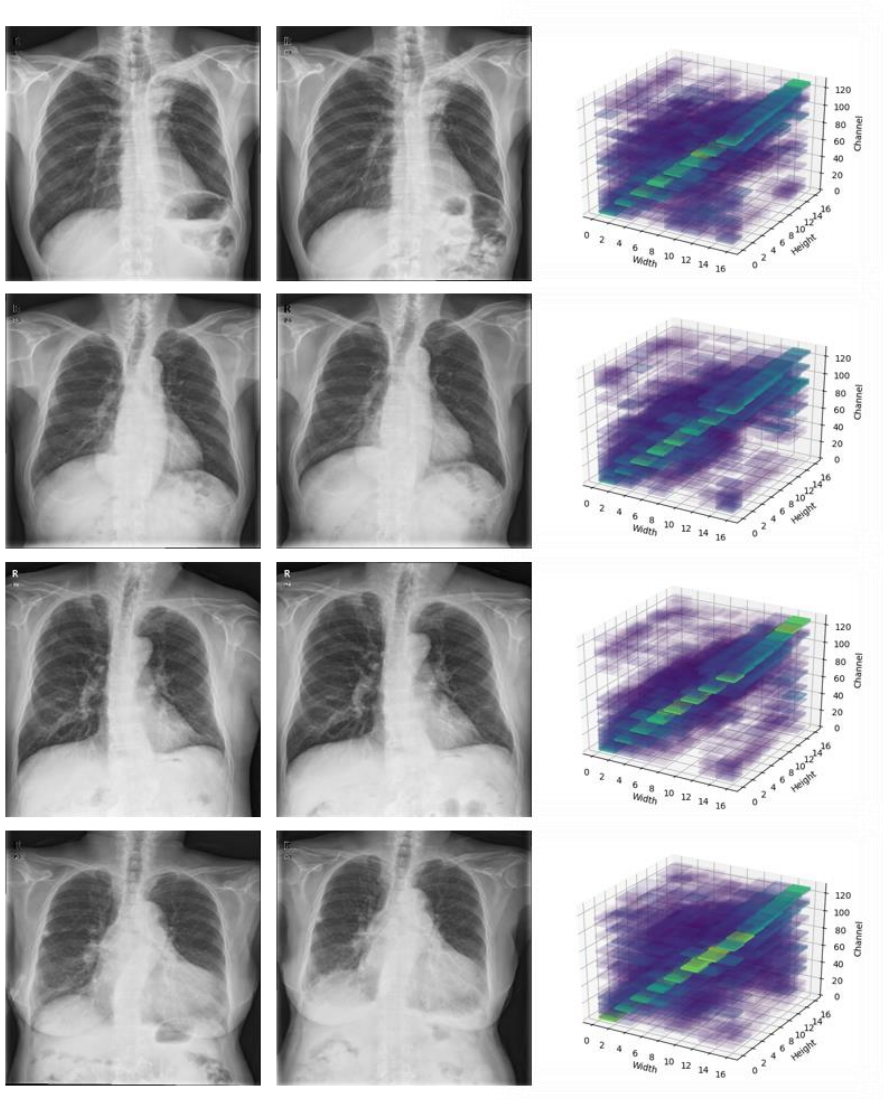


그림 11. 병변의 특별한 변화가 없거나 병변이 존재하지 않는 정상 흉부에 대한 기하 상관관계도, 각 영상의 판독문: “no significant interval change since last study” (1,2열) 흉부 X-선 (전,후) 영상, (3열) 3차원으로 시각화 된 기하 상관관계도

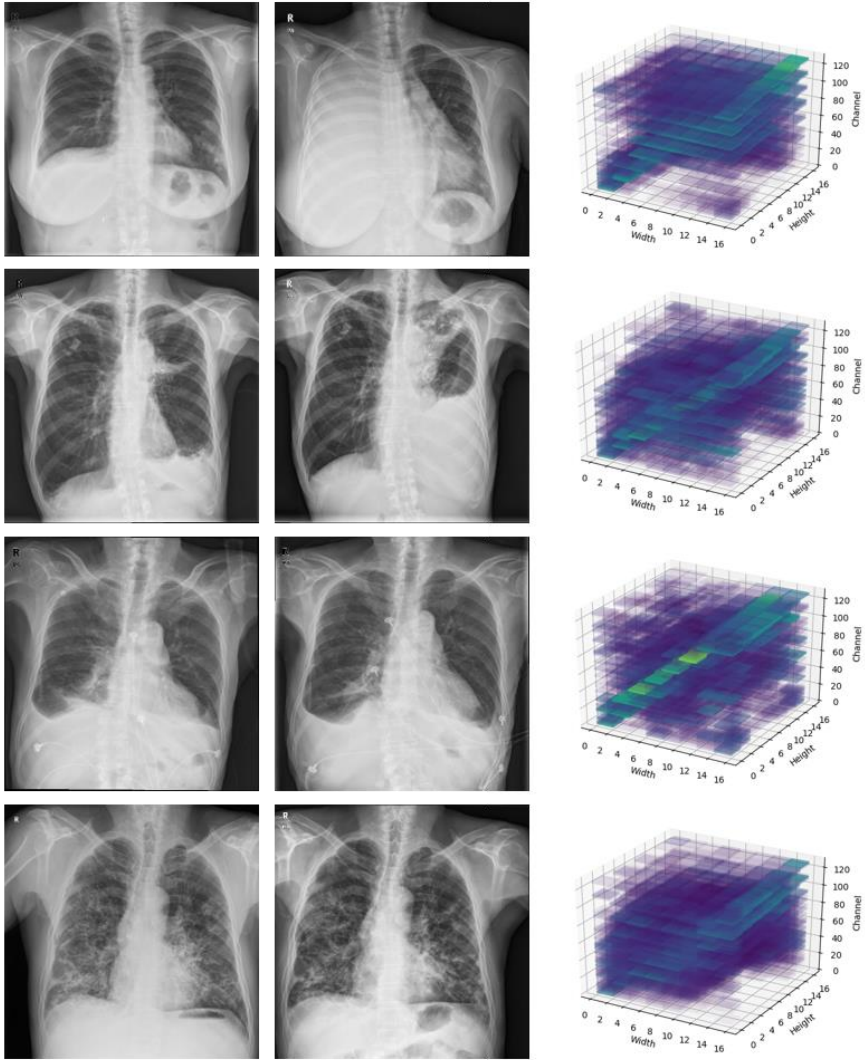


그림 12. 병변의 변화가 발생한 경우에 대한 기하 상관관계도. (1,2열) 흉부 X-선 (전,후) 영상, (3열) 3차원으로 시각화 된 기하 상관관계도. 각 행의 영상 판독문은 다음과 같다, (1행): “large amount of haziness and pleural effusion with passive atelectasis in right lung” (2행): “progressed lung cancer at left upper lung with increased pleural effusion since last study” (3행) “decreasing right pulmonary edema since last study” (4행) “mild improvement of consolidation in both lungs”

그림8-10은 본 연구에서 제안하는 알고리즘의 기하 상관관계도 예를 보여준다. 1,2열은 각각 이전 및 현재 영상의 쌍을 나타내고, 각 영상에서 병변의 빨간색 상자로 표기되어 있다. 3열은 1,2열 영상에 대한 기하 상관관계도를 3차원으로 시각화한 것이고, 병변의 변화에 따라 특정 패턴을 보여준다. 1행은 정상 흉부의 변화가 없는 경우, 2행은 비정상 흉부의 변화가 없는 경우를 나타낸다. 각 영상의 동일한 위치에 대해서 비교적 강한 상관관계를 띄며 이외 위치에 대해서는 음의 상관관계를 띄거나 약한 양의 상관관계를 띤다. 반면에 3,4행은 비정상 흉부의 변화가 있는 경우를 나타내는데 폐의 경계선이 바뀌어 동일한 위치의 상관관계가 약해지는 등 1,2행의 기하 상관관계도와는 다른 패턴을 보여준다. 기하 상관관계도 내에서 변화하는 패턴은 입력 영상에서 임의의 변화가 발생한 위치와 밀접한 관계가 있다. 지역설명자의 인덱스를 참고하여 영상 패치를 추출하였을 때, 실제 변화가 있는 영상 패치와 동일하였다.

4. 고찰

4.1. 실험 결과 고찰

주어진 (전,후) 영상에서 변화 감지 알고리즘은 딥러닝 알고리즘을 기반이므로, 알고리즘 학습에 사용하는 데이터베이스 구축에 매우 민감하다. 따라서 변화 검출 알고리즘뿐 만 아니라 데이터베이스 구축 과정에서도 새롭고 창의적인 시도가 필요하다. 예를 들어 본 연구에서는 시간에 따른 변화를 서술하는 키워드를 포함하는 문장을 선별하여 판독문을 구분하였지만, 보다 정교한 자연어 처리 알고리즘[16]을 도입하거나, 본 연구에서 선별되지 않은 판독문을 가공할 수 있는 최신의 기술이 도입되면 더 준수한 성능의 알고리즘을 획득할 수 있을 것이다.

영상을 그대로 사용하는 것은 픽셀 단위로 해석하는 것과 동일하며 이는 매우 고차원이기 때문에 해석이 불가능에 가깝고, 과적합의 원인이 된다. 따라서 본 연구에서는 적절한 모델을 사용하여 각 영상에서의 고유 특징 (지역설명자)를 추출하였다. 이 특징 추출 네트워크의 구조를 개선하고, 원본 입력 영상의 해상도를 해치지 않는다면 더욱 성능 개선을 도모할 수 있을 것이다.

4.2. 알고리즘 고찰

변화 감지를 위한 기법은 기술적으로 매우 도전적이다. 유사한 영상에서 변화를 감지하는 기술은 주로 전통적인 영상처리 기술의 하나인 영상 정합에 의존하여 제한된 영역에서 발표되며 선형 정합 등 단순한 영상 정합 기법은 영상내 비선형적 움직임에 의한 차이를 정확하게 인식하기 어려워 위양성(false positive)의

변화 감지 예측을 도출하기 쉽다. 그에 반해 세밀한 정합이 가능한 비강체 정합(non-rigid registration)은 영상간 차이를 주변의 구조물의 변형을 통하여 극복하면서 위음성(false negative)의 변화 감지 예측을 도출하기 쉽다.

이러한 점을 비추어 볼 때 일반적인 정합 기법들은 일반적 영상을 대상으로 개발된 경우가 많아 방사선 영상과 같은 투시적인 영상에서 구조물의 변화가 촬영 각도에 따라서도 심하게 변화하는 경우에 대처하기 어렵다고 볼 수 있다. 즉, 본 연구에서 제안하는 알고리즘은 영상 정합을 직접 수행하지 않고, 추출된 지역설명자의 상관 관계도의 분포를 파악하여 그 패턴을 이용하기 때문에 상기한 기법보다 준수한 성능과 프로세싱 소요시간 단축을 기대할 수 있다.

5. 결론

본 연구는 흉부 X-선 영상 및 판독문을 활용한 인공지능 시스템을 구축했다는 점에서 보편적이거나, 다음과 같은 차별성을 제공한다. 첫째, 선행연구는 횡단면적 영상을 입력으로 판독문을 생성해내는 알고리즘 개발에 중점이 있는 반면 본 연구는 영상 시퀀스에서 변화를 탐지하는 알고리즘을 개발하였다. 둘째, 본 연구에 필요한 기계학습용 참조표준 데이터베이스를 자체적으로 확보했으며, 질환의 종류, 경과시간, 변화 형태 등에 따라 데이터를 분류할 수 있는 체계를 구축하였다. 특히 영상 내의 변화정보 (병변의 크기, 위치정보 등) 없이 영상내 픽셀 정보만을 변화를 판별할 수 있도록 단-대-단 학습 시스템으로 설계하였다.

본 연구는 5,000여건의 비교적 적은 수의 영상 데이터로 본 시스템의 개발 가능성을 타진하였다. 또한 판독문 분석 과정에서 고유 문장, 키워드 중심의 추출 등 비교적 간단한 전처리 기법을 사용하여 데이터가 완전히 정제되지 않은 가혹한 조건이었음에도 불구하고 기존 및 비교 기법 대비 우수한 검출 성능을 확인하였다.

본 연구에서 제안하는 알고리즘은 급성 변화를 검출하여 응급상황 선별에 활용될 수 있고, 전이 유무를 알아낼 수 있다. 또한 변화를 설명하는 자동 판독소견 생성 기법에 활용할 수 있는 등, 본 연구에서 자체적으로 구축한 참조표준 데이터셋을 이용하여 실제 임상에서의 판독과정을 시뮬레이션 하여 기존 개발되었던 진단 기술의 고도화에 기여할 수 있을 것으로 예상된다.

참고 문헌

- [1] Lakhani, Paras, and Baskaran Sundaram. "Deep learning at chest radiography: automated classification of pulmonary tuberculosis by using convolutional neural networks." *Radiology* 284.2 (2017): 574–582.
- [2] Nam, Ju Gang, et al. "Development and validation of deep learning-based automatic detection algorithm for malignant pulmonary nodules on chest radiographs." *Radiology* 290.1 (2019): 218–228.
- [3] Rajpurkar, Pranav, et al. "Chexnet: Radiologist-level pneumonia detection on chest x-rays with deep learning." *arXiv preprint arXiv:1711.05225* (2017).
- [4] Santeramo, Ruggiero, Samuel Withey, and Giovanni Montana. "Longitudinal detection of radiological abnormalities with time-modulated LSTM." *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support*. Springer, Cham, 2018. 326–333.
- [5] Singh, Ramandeep, et al. "Deep learning in chest radiography: detection of findings and presence of change." *PloS one* 13.10 (2018).
- [6] Wang, Fei, et al. "Residual attention network for image classification." *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2017.
- [7] Wang, Xiaolong, et al. "Non-local neural networks." *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2018.
- [8] Hu, Jie, Li Shen, and Gang Sun. "Squeeze-and-excitation networks." *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2018.
- [9] Rocco, Ignacio, Relja Arandjelovic, and Josef Sivic. "Convolutional neural network architecture for geometric matching." *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2017.
- [10] Rocco, Ignacio, Relja Arandjelović, and Josef Sivic. "End-to-end weakly-supervised semantic alignment." *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2018.
- [11] Long, Jonathan, Evan Shelhamer, and Trevor Darrell. "Fully convolutional networks for semantic segmentation." *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2015.
- [12] Lakhani, Paras, and Baskaran Sundaram. "Deep learning at chest radiography: automated classification of pulmonary tuberculosis by using convolutional neural networks." *Radiology* 284.2

- (2017): 574–582.
- [13] Jaeger, Stefan, et al. "Two public chest X-ray datasets for computer-aided screening of pulmonary diseases." *Quantitative imaging in medicine and surgery* 4.6 (2014): 475.
 - [14] Irvin, Jeremy, et al. "Chexpert: A large chest radiograph dataset with uncertainty labels and expert comparison." *Proceedings of the AAAI Conference on Artificial Intelligence*. Vol. 33. 2019.
 - [15] Setio, Arnaud Arindra Adiyoso, et al. "Pulmonary nodule detection in CT images: false positive reduction using multi-view convolutional networks." *IEEE transactions on medical imaging* 35.5 (2016): 1160–1169.
 - [16] Joulin, Armand, et al. "Bag of tricks for efficient text classification." *arXiv preprint arXiv:1607.01759* (2016).
 - [17] Wang, Xiaosong, et al. "Chestx-ray8: Hospital-scale chest x-ray database and benchmarks on weakly-supervised classification and localization of common thorax diseases." *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2017.

Abstract

Development and validation of a deep learning algorithm for longitudinal change detection in sequential chest X-ray images

Dong Yul Oh

Interdisciplinary Program in Bioengineering

The Graduate School

Seoul National University

The diagnostic decision for chest X-ray image generally considers a probable change in a lesion, compared to the previous examination. We propose a novel algorithm to detect the change in longitudinal chest X-ray images. We extract feature maps from a pair of input images through two streams of convolutional neural networks. Next, we generate the geometric correlation map computing matching scores for every possible match of local descriptors in two feature maps. This correlation map is fed into a binary classifier to detect specific patterns of the map representing the change in the

lesion. Since no public dataset offers proper information to train the proposed network, we also build our own dataset by analyzing reports in examinations at a tertiary hospital. Experimental results show our approach outperforms previous methods in quantitative comparison. We also provide various case examples visualizing the effect of the proposed geometric correlation map.

Keywords: Change detection, chest X-ray, neural network, longitudinal analysis, geometric correlation map

Student Number: 2083-21355