creative commons

COMMONS DEED

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

dCollection

공학석사 학위논문

딥러닝 알고리즘을 활용한 자동 유방초음파 영상에서의 유방암 검출 시스템 연구

A Study on Deep-learning based
Computer-aided Detection
for Breast Cancer in Automated Breast Ultrasound

2021년 8월

서울대학교 대학원
협동과정 바이오엔지니어링전공
김 대 경

딥러닝 알고리즘을 활용한 자동 유방초음파 영상에서의 유방암 검출 연구

지도교수 Sungwan Kim

이 논문을 공학석사 학위논문으로 제출함

2021년 6월

서울대학교 대학원

협동과정 바이오엔지니어링전공

김 대 경

김대경의 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

딥러닝 알고리즘을 활용한 자동 유방초음파 영상에서의 유방암 검출 시스템 연구

김대경

서울대학교 대학원

협동과정 바이오엔지니어링 전공

본 연구의 목표는 삼차원 자동 유방초음파 영상에서 유방암을 검출하는 시스템을 개발하고 그 성능을 검증하는 것이다. 기존의 유방촬영술이나 수동 유방초음파 영상에 비하여 다양한 장점을 가지고 있어 사용이 확대되고 있는 자동 유방초음파 영상은 초음파 영상과 삼차원 영상의 특성으로 인해 영상 판독에 상대적으로 많은 시간과 노력을 요구한다. 최근 급격히 발전하고 있는 컴퓨터 비전 기술과 딥러닝(Deep-learning) 알고리즘을 활용한 컴퓨터 발견 보조 시스템(Computer-aided detection)에 대한 연구가 활발하게 이루어지고 있고 자동 유방초음파 영상에 딥러닝을 활용한 컴퓨터 발견 보조시스템이 적용될 경우 의료시스템의 효율성 및 품질 향상에 큰 기여가 가능할 것으로 보인다.

본 연구에서 개발된 유방암 검출 시스템은 크게 세 가지의 모듈로 구분할 수 있으며 각각의 모듈은 딥러닝 알고리즘 중 여러 연구를 통해 이미지 분석에 높은 효과를 보인 삼차원 합성곱 신경망(3-dimensional convolution network)을 기반으로 하여 개발되었다. 유방암 검출 시스템은 먼저 '유방 병변 후보 검출기'를 통하여 전체 자동 유방초음파 영상을 스캔하여 의심스러운 병변의 위치를 추정해낸다. 추정된 위치에 실제로 유방 병변이 있는지 높은 정확도로 판별하기 위해 '유방 병변 유무 판별기'로 실제 병변이 존재하는 영역을 분류한다. 종양으로 분류된 결과들은 종양에 대한 악성도를 판별하는 '악성 유방병변 분류기'를 통해 병변을 양성과 악성으로 분류하여 최종적으로는 악성 유방암으로 의심되는 병변을 검출하게 된다.

본 연구에서는 자동 유방초음파 영상으로부터 추가적인 라벨링 작업없이 기존의 정보를 활용하는 효율적인 방법으로 딥러닝 알고리즘을 학습하고 검증하기 위한 데이터 셋을 구축하였다. 구축된 데이터 셋을 이용해 딥러닝 알고리즘을 효과적으로 학습시키기 위하여 약지도 학습법(Weakly supervised learning)과 인공신경망 구조를 제안하였고 이를 통한 모듈의 성능향상도 확인할 수 있었다.

본연구에서는 유방암 검출 시스템을 구성하는 모듈 별 성능평가와 유방암 검출 시스템의 전체적인 성능 검증을 위해 실험을 진행하였다. 개별 모듈의 성능 평가결과, 본 연구에서 제안된 기법에 의한 성능 향상을 확인할 수 있었으며, 특히 유방 병변 유무 판별기에 사용된 템플릿 마스크와 마스크 네트워크는 성능의 큰 효과를 보였다. 본 연구에서 개발된 시스템은 검출률

87.6% (95% 신뢰구간, 81.9%-93.3%), 위양성 검출률 0.442(95% 신뢰구간 0.356-0.528)으로 기존의 다른 시스템과 비교했을 때에도 높은 성능을 나타냈다.

본 연구에서 자체적으로 구축한 데이터셋과 설계한 알고리즘을 이용하여 자동 유방 초음파영상에서 유방암 검출 시스템을 개발하였고 성능 검증결과 높은 성능을 보였다. 따라서 딥러닝을 활용한 유방암 검출 시스템은 자동 유방초음파 영상의 진단 성능을 향상시킬 수 있는 잠재력을 가지고 있다고 할 수 있다.

주요어 : 딥러닝, 자동 유방초음파, 유방암, 자동 검출 시스템
학 번 : 2019-23660

목차

국문초록 1
목차 4
그림 목차 7
표 목차 10
서론 11
1.1. 배경 11
1.1.1. 유방 촬영술의 한계 11
1.1.2. 유방 초음파 영상의 효과 11
1.1.3. 수동 유방 초음파 영상의 한계 12
1.1.4. 자동 유방 초음파 영상의 필요성 12
1.1.5. 딥러닝 기반 컴퓨터 보조 발견 시스템 13
1.2. 연구 동향 14
1.2.1. 자동 유방초음파 진단 보조시스템의 연구방향 14
1.2.2. 딥러닝을 활용한 컴퓨터 보조 발견 시스템 연구 15
1.3. 연구의 목적 16
본론 19
2.1. 전체 시스템 구조 19
2.2. 유방병변 후보 검출기 21
2.2.1. 데이터셋 구축 21
2.2.2. 알고리즘 구조 24
2.2.3. 훈련 및 예측과정 27
2.3. 유방병변 유무 판별기 29

2.3.1. 알고리즘 구조 29
2.3.2. 주 네트워크 31
2.3.3. 마스크 네트워크 31
2.3.4. 네트워크 설계 후보군 33
2.3.5. 템플릿 마스크 35
2.3.6. 유방병변의 범주 35
2.3.7 템플릿 마스크 생성과정 36
2.4. 악성 유방병변 분류기 39
2.4.1 알고리즘 구조 39
2.4.2 훈련 및 예측과정 41
2.5. 실험 방법 41
2.5.1. 실험 목적 41
2.5.2. 데이터셋 구성 42
2.5.3 알고리즘 학습 및 검증 44
2.5.3 실험 환경 45
결과 및 분석 47
3.1 유방병변 후보 검출기 성능 47
3.2 유방병변 유무 분류기 성능 54
3.2.1. 알고리즘 구조에 따른 성능 54
3.2.2. 손실 통합 기법에 따른 성능 57
3.2.3. 적용된 유방병변 특성에 따른 성능 58
3.2.4. 시각적 평가 60
3.3. 유방병변 악성도 분류기 성능 62

3.4. 유방암 검출 시스템 정확도 63
고찰 67
4.1. 실험 결과 고찰 67
4.2. 발전 방향 68
결론 69
참고 문헌 71
Abstract 75

그림 목차

그림 1 자동 유방초음파 장비와 유방을 한번에 스캔할 수 있는 넓은 구경(FOV)을 가진 reverse curve형태의 인체 공학적 탐촉자 13
그림 2 효과적인 자동유방 초음파 영상 활용을 위한 촬영 가이드. 양쪽의 유방에 위치에 따라 다수의 촬영을 권장하고 있으며 일반적으로 한쪽에 3개의 볼륨영상을 스캔하게 됨 [11]..... 14
그림 3 CNN 학습 모델에 따른 Class activation map(CAM). (1 행) 자동 유방초음파 영상에서 유방병변의 transverse, coronal, sagittal 단면. (2 행) CNN을 통해 학습한 모델의 CAM. (3 행) 본 연구에서 제안하는 기법을 적용한 CNN의 CAM. 3행의 경우에 유방병변에 더 집중하고 있음을 확인할 수 있음. 17
그림 4 전체 시스템 구성도: 유방암 검출 시스템의 전체적인 구조..... 20
그림 5 BIRADS 분류표. 0에서 6까지 분류되어 있으며 높은 값을 가질수록 위험도가 높다고 볼 수 있음..... 22
그림 6 Structured report 의 예시. 판독의들이 ABUS 판독시 기록한 위치나 크기정보들이 항목별로 저장되어 있음..... 23
그림 7 Dense layer의 densely connected CNN layer들 25
그림 8 유방병변 후보 검출기 알고리즘 구조. 연속된 6개와 12개의 Dense layer로 구성된 두개의 Dense block을 통해 유방병변의 위치를 예측함 26
그림 9 유방병변 후보 검출기의 예측 과정. 입력된 자동 유방초음파 영상을 9등분하여 입력한 후 출력된 배열을 다시 하나로 통합하여 병변의 위치를 추측함..... 28
그림 10 Mask Branch network의 구조. 유방병변의 유무를 판별하는 역할을 하는 주네트워크(Main network)와 주네트워크의 효과적인 학습을 도와주는 마스크 네트워크(Mask Branch)로 이루어져있음..... 30
그림 11 마스크 네트워크의 여러가지 구조. (1행 1열) MBN-V1. (1행 2열) MBN-V2. (2행 1열) MBN-V3. (2행 2열) MBN-V4..... 34

그림 12 템플릿 마스크의 생성과정. 유방병변의 위치, 크기, 모양 특성을 고려하여 병변의 공간적 정보를 포함하는 템플릿 마스크를 생성할 수 있음..... 37
그림 13 템플릿 마스크의 한 예. 각 열은 transverse, coronal, sagittal 단면을 의미함. (행 1) 대상이 되는 유방병변. (행 2) 대상병변을 이용해 생성된 템플릿 마스크..... 38
그림 14 유방병변 악성도 분류기의 알고리즘 구조. (행 1) 기존의 DenseNet을 활용한 분류기. (행 2) 본 연구에서 제안하는 분류기 구조. 40
그림 15 훈련 및 검증용 데이터셋의 구성. BIRADS에 따라 환자군을 구분하여 각각의 모듈을 학습시키고 검증하는 데에 사용될 수 있도록 함. 43
그림 16 유방암 검출 시스템의 성능 평가에 사용된 테스트 셋의 구성. 44
그림 17 Simple 3D Unet V1의 알고리즘 구조도. 2개의 dense block의 encoder와 1개의 deconvolution layer의 decoder로 이루어져 있음. 48
그림 18 Simple 3D Unet V2의 알고리즘 구조도. 3개의 dense block의 encoder 와 2개의 deconvolution layer의 decoder로 이루어져 있음 49
그림 19 유방병변 후보 검출기의 결과 예시1. (왼쪽) 자동 유방초음파 영상의 유방병변이 존재하는 위치의 단면으로 하얀색으로 표시된 부분이 유방병변의 위치 (오른쪽) 유방병변 후보 검출기의 결과로 주황색으로 표시된 부분이 검출결과임 51
그림 20 유방병변 후보 검출기의 결과 예시2. (왼쪽) 자동 유방초음파 영상의 유방병변이 존재하는 위치의 단면으로 하얀색으로 표시된 부분이 유방병변의 위치 (오른쪽) 유방병변 후보 검출기의 결과로 주황색으로 표시된 부분이 검출결과임 52
그림 21 유방병변 후보 검출기의 결과 예시3. (왼쪽) 자동 유방초음파 영상의 유방병변이 존재하는 위치의 단면으로 하얀색으로 표시된 부

분이 유방병변의 위치 (오른쪽) 유방병변 후보 검출기의 결과로 주황색으로 표시된 부분이 검출결과 53
그림 22 마스크 네트워크를 적용한 DenseNet과 ResNet 기반의 유방병변 유무 분류기 ROC 그래프. 네트워크 구조마다 성능이 다르지만 전체적인 성능을 보았을 때에 기존 대비 높은 성능을 보임. 56
그림 23 다중 손실 가중방식에 따른 ROC 그래프 비교. 불확실성에 기반한 손실 통합 방식이 가장 높은 성능을 보임. 58
그림 24 템플릿 마스크에 활용된 유방병변의 특징에 따른 ROC 그래프 비교 60
그림 25 템플릿 마스크와 마스크 네트워크 적용 전후 CAM. (행 1) 유방병변의 단면 (행 2) 기존 모델의 CAM (행 3) 템플릿 마스크와 마스크 네트워크가 적용된 모델의 CAM 61
그림 26 유방병변 악성도 분류구에 입력되는 입력 이미지 후보군 62
그림 27 본 연구에서 개발한 유방암 검출 시스템의 결과 예시1. (왼쪽) 자동 유방초음파 영상과 하얀색으로 표시된 유방암 (오른쪽) 유방암 검출시스템에서 활성화된 정도를 나타내는 CAM과 주황색으로 표시된 최종결과 64
그림 28 본 연구에서 개발한 유방암 검출 시스템의 결과 예시2. (왼쪽) 자동 유방초음파 영상과 하얀색으로 표시된 유방암 (오른쪽) 유방암 검출시스템에서 활성화된 정도를 나타내는 CAM과 주황색으로 표시된 최종결과 65
그림 29 그림 28 본 연구에서 개발한 유방암 검출 시스템의 결과 예시3. (왼쪽) 자동 유방초음파 영상과 하얀색으로 표시된 유방암 (오른쪽) 유방암 검출시스템에서 활성화된 정도를 나타내는 CAM과 주황색으로 표시된 최종결과 66

표 목차

표 1 유방병변 후보 검출기의 성능 비교표 47
표 2 마스크 네트워크를 적용한 유방병변 유무 분류기의 성능 비교표 .. 54
표 3 다중 손실 가중 방식에 따른 성능 비교표..... 58
표 4 템플릿 마스크에 활용된 유방병변의 특징에 따른 성능비교 59
표 5 유방병변 악성도 분류기의 성능..... 63
표 6 비교군 CAD와 본 연구에서 제안한 유방암 검출 시스템의 성능비교 63

서론

1.1. 배경

1.1.1. 유방 촬영술의 한계

유방암은 전 세계 여성들의 암 사망의 주요 원인 중에 하나이며 아시아의 경우 유방암 발생률이 낮았으나 최근 발생률이 급격히 증가하고 있다. [1] 많은 연구들은 유방암 검사가 초기 단계에서 암을 발견할 수 있다는 것을 밝혀냈다. [2-4] 유방촬영술을 통한 유방암의 발견은 약 80%의 민감도가 있다고 보고되고 있으나 이는 유방 촬영술상 유방실질의 밀도가 치밀한 경우에는 50% 이하로 급격히 감소하였다. [5-7] 검진 대상 중 50세 미만의 여성들은 대개 절반 이상에서 치밀 유방 소견을 보이므로 유방촬영술 검진은 그 실효성에 문제가 있다고 할 수 있다.

1.1.2. 유방 초음파 영상의 효과

유방암 고위험군을 대상으로 한 대규모 전향적 연구에서 유방 초음파를 부가적으로 같이 시행한 결과 유방암 발견율이 유의하게 증가하였다. [8] 치밀 유방 소견을 가진 여성에서 유방촬영에 더하여 초음파 검진을 같이 시행한 다른 연구들에서도 유방암 위험도에 상관없이 1000명당 3-4개의 유방암이 초음파 검사로 더 발견되었다. 이중에서 64-90%는 검진을 통해 발견하여 조기 치료로 생존율을 늘릴 수 있는 림프절 전이가 없는 침윤암으로 밝혀졌다. [9]

1.1.3. 수동 유방 초음파 영상의 한계

수동 유방 초음파 영상(Hand-held breast ultrasound, HHUS)은 검사자가 소형 탐촉자를 통해 실시간으로 스크린을 보면서 촬영하는 방식이기 때문에 스캔 테크닉을 표준화하기 어렵고, 전체 유방을 검사하는데 상당한 시간과 노력이 필요하다. 그에 따라 전문화된 인력이 부족하며, 검사자간 편차가 클 수밖에 없다. 또한 재현성이 낮고, 위양성 병변 검출이 발생하는 문제가 있다.

1.1.4. 자동 유방 초음파 영상의 필요성

자동 유방초음파 영상(Automated breast ultrasound system, ABUS)은 고주파(high frequency)로 넓은 구경(FOV) 유방조직을 고해상도 초음파 영상으로 한 번에 빠르게 스캔하여 제공한다. 탐촉자의 모양도 유방의 곡면에 맞춘 reversed curve 형태로 고안되어 접촉 불량에 의한 인공물을 줄이고 유방의 가장자리 병변까지 충분히 포함할 수 있다. 유방을 AP, LAT, MED의 세 부분을 나누어 스캔하는 프로토콜을 적용하여 유방이 빠짐없이 포함될 수 있게 한다. 오차가 작고 재현성이 높아 유방 검사자 간의 차이를 극복할 수 있다. [10]

그림 1 자동 유방초음파 장비와 유방을 한 번에 스캔할 수 있는 넓은 구경(FOV)을 가진 reverse curve 형태의 인체 공학적 탐촉자

1.1.5. 딥러닝 기반 컴퓨터 보조 발견 시스템

영상기반의 유방암 진단은 오진이 매우 높으며 이를 낮추기 위한 다양한 연구가 현재까지도 진행되고 있다. 자동 유방 초음파 영상은 비교적 최근에 도입된 검사방법으로 영상의학과 전문의 중에서도 일부만이 경험적 데이터 판독 및 노하우를 가지고 있는 분야이며 또한 한 명의 환자를 스캔할 때마다 최소 6개 이상의 볼륨데이터로 구성되어 있어 상대적으로 영상의 분석에 많은 시간적 인적 자원을 필요로 한다.

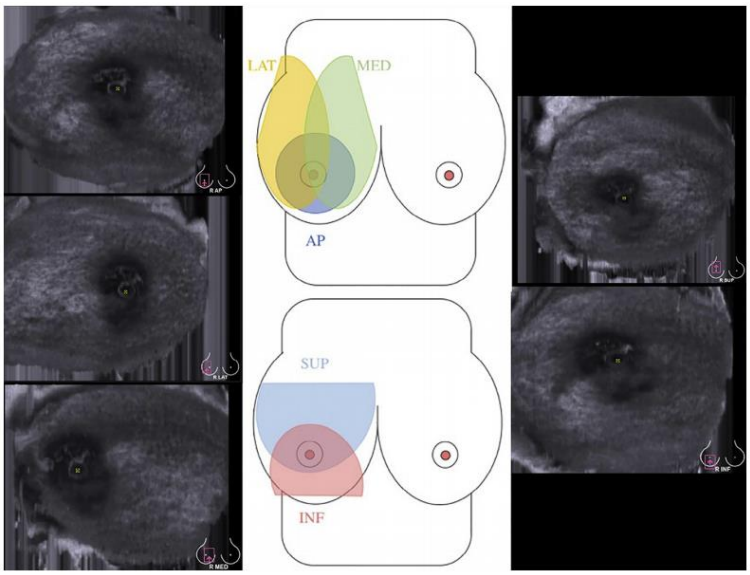


그림 2 효과적인 자동유방 초음파 영상 활용을 위한 촬영 가이드. 양쪽의 유방에 위치에 따라 다수의 촬영을 권장하고 있으며 일반적으로 한쪽에 3개의 볼륨영상을 스캔하게 됨 [11]

관측자간 차이 및 위 음성을 보완하기 위해 컴퓨터 보조 발견 시스템(Computer-aided detection system, CAD)을 접목하려는 연구들이 진행되고 있으며 실제 판독 시 진단능을 향상시키는 것으로 보고되고 있다. 딥러닝(Deep-learning) 기술은 근래 그 응용 범위를 폭발적으로 넓혀가고 있다. 컴퓨터 비전, 자연어처리, 음성인식 등의 다양한 분야에서 기존 기술 대비 월등한 성능을 보이고 있으며, 의료영상 분야에서도 역시 최근 몇 년간 딥러닝의 활용 범위가 점점 넓어지고 있는 추세이다.[12-14]

1.2. 연구 동향

1.2.1. 자동 유방초음파 진단 보조시스템의 연구방향

CAD 시스템의 주요 과제는 암으로 성장할 수 있는 유방

병변을 감지하는 것이다. 자동 유방초음파 영상의 유방병변은 대체로 어두운 영역으로 표현되는데 이는 지방 조직, 초음파 음영, 무반향 조직과 같은 다양한 원인으로 인해서도 유사하게 나타날 수 있다. 따라서 많은 CAD 시스템이 이러한 부위를 유방병변으로 잘못 감지하여 위 양성으로 나타낸다. 심지어 영상의학과 전문의들도 위 양성 영역과 유방병변을 구별하는 데 어려움을 겪고 있으며 유방 초음파 검사에서 다수의 위 양성 반응이 검출되었다[15, 16].

1.2.2. 딥러닝을 활용한 컴퓨터 보조 발견 시스템 연구

컴퓨터 보조 발견 시스템은 일반적으로 크게 두 가지 단계로 이루어져 있다. 먼저 병변의 위치를 찾은 다음 각 영역이 악성인지 양성인지를 결정한다. 첫 번째 단계에서는 악성병변의 가능성을 가진 모든 영역을 놓치지 않기 위해 상대적으로 민감도에 높은 비중을 두어 모델을 설계한다. 두 번째 단계에서는 불필요한 생체검사를 발생시킬 수 있는 위양성을 줄이기 위해 검출된 후보 영역들을 분류한다. 신뢰할 수 있는 높은 성능의 분류기를 개발하기 위해, 최근 많은 연구[17,18]에서는 이미지 분류 작업에서 좋은 성능을 보인 합성곱 신경망(Convolutional Neural Network, CNN)을 사용했다. 여러 연구에서는 기존의 이미지 처리 기술에 기반한 유방병변 분류기를 제안했지만, 여전히 위 양성의 문제는 충분히 해결되지 않았다. [19-22]

1.3. 연구의 목적

본 연구에서는 합성곱 신경망을 기반으로 한 딥러닝 기술을 사용하여 자동 유방초음파 영상에서 유방암으로 의심되는 영역을 검출하는 시스템을 개발하는 것을 목표로 한다. 이를 통해 자동 유방초음파 영상의 판독 결과 향상과 효율성을 증대시킬 수 있는 기술을 개발하는 것이 본 연구의 궁극적인 목적이다.

일반적인 이미지와 비교했을 때 상대적으로 자동 유방초음파 스캔에 표시된 유방병변은 명확한 경계를 나타내지 않는 경우가 많다. 그림 2의 첫 번째 행에서 볼 수 있듯이, 유방병변은 종종 불규칙한 모양의 덩어리의 형태로 나타나며, 각각의 유방병변은 주변부와 구별하기 어렵다. 또한, 자동 유방초음파 특성상 낮은 이미지 해상도와 높은 노이즈 수준의 자동 유방초음파는 불분명한 유방병변에 대해 CNN을 훈련하는 것을 더욱 어렵게 만든다.

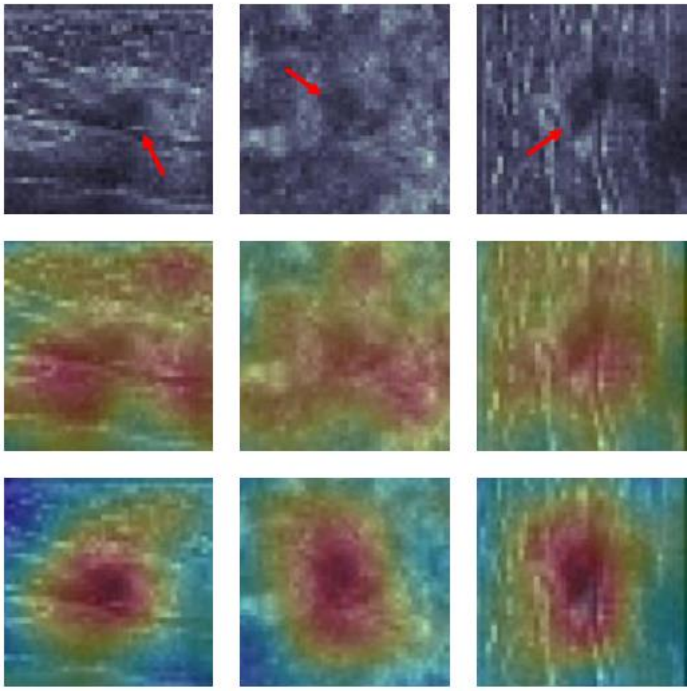


그림 3 CNN 학습 모델에 따른 Class activation map(CAM). (1 행) 자동 유방초음파 영상에서 유방병변의 transverse, coronal, sagittal 단면. (2 행) CNN을 통해 학습한 모델의 CAM. (3 행) 본 연구에서 제안하는 기법을 적용한 CNN의 CAM. 3행의 경우에 유방병변에 더 집중하고 있음을 확인할 수 있음.

이러한 CNN을 통한 유방병변의 학습의 어려움을 고려하였을 때, 본 연구에서 인공신경망의 특정 구조 및 유방병변의 공간정보를 사용한 훈련 방법이 CNN 분류기의 성능을 향상시키는데 도움이 된다는 것을 경험적으로 발견했다. 그림 3의 두 번째 행은 기존의 CNN을 사용한 질량 분류를 위한 클래스 활성화 맵(Class Activation Map, CAM)을 보여준다. CAM에서 붉은색에 가까울수록 알고리즘이 학습 시 집중하는 부분이라고 볼 수 있다. 활성화된 네트워크 가중치는 유방병변에 집중되지 않고 배경의 어두운 영역까지 광범위하게 분포한다. 대조적으로, 본 연구에서

제안하는 방법이 적용된 그림 3의 세 번째행에서는 유방병변이 존재하는 영역을 집중적으로 활성화한다.

합성곱 신경망을 훈련하려면 가급적 배경과 구별할 수 있는 대상 객체의 명확한 특징을 포함하는 많은 샘플 이미지가 필요하다. 따라서 자동 유방초음파 영상에서 인공신경망을 학습시키기에 충분할 만큼의 데이터를 수집해야 한다. 자동 유방 초음파 영상에서 각각의 병변을 찾아서 표기하는 것은 전문인력과 많은 작업시간을 요구하며 정확한 초음파영상에서 유방병변의 경계는 의료 전문가에게도 명확하지 않다. 본 연구에서는 데이터 라벨링에 소요되는 시간과 인력을 감소시키기 위하여 기존의 존재하는 정보를 최대한 활용한 딥러닝을 위한 데이터를 수집하였고 최소한의 정보로 모델을 훈련하기 위해 약지도 학습법을 적용하였다.

유방암 검출 시스템의 전체적인 성능 측정뿐만 아니라 이를 구성하는 각각의 모듈에 대한 성능 평가도 진행한다. 개발된 유방암 검출 시스템의 성능 평가를 통해 시스템의 효과성을 확인함과 동시에 좀 더 나은 시스템으로 가기 위한 자료로써 활용한다.

본론

2.1. 전체 시스템 구조

본 연구에서 제시하는 유방암 검출 시스템은 유방병변의 가능성을 가진 후보 영역을 최대한 검출해내는 '유방병변 후보 검출기', 유방병변 후보 영역 중에서 위양성 영역을 제거하는 '유방병변 유무 판별기, 그리고 유방병변의 악성 여부를 판단하는 '악성 유방병변 분류기'로 구성되어있다.

전체적인 시스템을 도식화하여 표현하면 그림 4와 같다. 스캐너를 통해 생성된 자동 유방초음파영상을 세 축에 대하여 동일한 스케일로 만드는 전처리 과정을 통해서 시스템에 입력될 삼차원 데이터를 만든다. 입력 영상은 유방병변 후보 검출기를 통해서 유방병변의 가능성을 가진 영역을 검출해낸다. 검출된 모든 위치에 대해서 중심좌표를 기준으로 동일한 사이즈의 볼륨을 생성해낸다. 생성된 볼륨들은 유방병변 유무 판별기의 입력 데이터로 사용된다. 전체 유방병변 후보 볼륨 중 유방병변 유무 판별기의 결과가 음성인 이미지들은 제외한다. 유방병변으로 분류된 남은 볼륨들은 악성도 분류기에 입력된다. 악성도 분류기는 입력 볼륨내의 유방병변의 악성도를 예측하고 설정된 역치값 이상의 출력에 해당하는 영역들이 유방암 검출기의 최종 출력 영역이 되고, 이 영역이 되고 원본 삼차원 영상에 그 위치를 표시하여 보이게 된다.

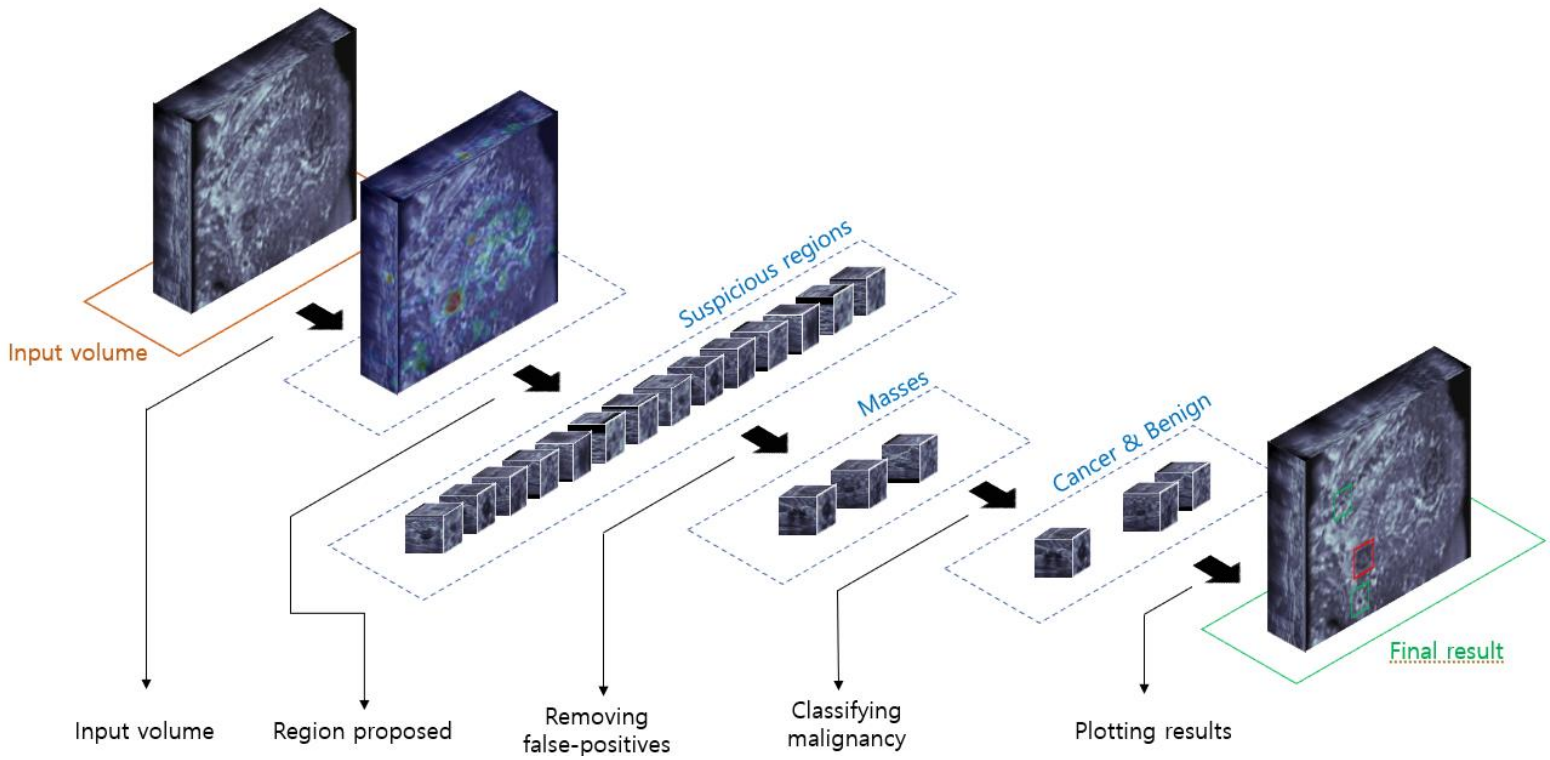


그림 4 전체 시스템 구성도: 유방암 검출 시스템의 전체적인 구조

2.2. 유방병변 후보 검출기

2.2.1. 데이터셋 구축

삼차원 의료데이터의 특성상 데이터를 라벨링하거나 기준 데이터를 만드는 데에 일정 이상의 경험을 가진 전문의의 인력이 필요하며 삼차원 영상은 이차원 영상과 비교하였을 때에 더 많은 시간과 노력을 필요로 한다. 본 연구에 사용될 데이터를 생성하는 데 필요한 시간과 비용을 최소화하기 위하여 기존의 판독과정에서 생성되는 출력물을 최대한 활용하여 데이터를 만드는 방법을 사용하였다.

본 연구에서 딥러닝 모델의 트레이닝 및 평가에 사용되는 데이터를 만들기 위해 자동 유방초음파를 판독하는 과정에서 기본 절차상 생성되는 출력물인 판독문과 Structured report를 활용한다. 판독문은 해당 병원은 PACS에 저장되며, Structured report는 판독의가 유방병변에 대해서 마킹한 정보를 의료영상 데이터 표준 프로토콜인 DICOM(Digital Imaging and Communications in Medicine)의 형태로 영상과 함께 서버에 저장된 파일을 말한다.

Final Assessment Categories			
Category		Management	Likelihood of cancer
0	Need additional imaging or prior examinations	Recall for additional imaging and/or await prior examinations	n/a
1	Negative	Routine screening	Essentially 0%
2	Benign	Routine screening	Essentially 0%
3	Probably Benign	Short interval-follow-up (6 month) or continued	>0 % but ≤ 2%
4	Suspicious	Tissue diagnosis	4a. low suspicion for malignancy (>2% to ≤ 10%) 4b. moderate suspicion for malignancy (>10% to ≤ 50%) 4c. high suspicion for malignancy (>50% to <95%)
5	Highly suggestive of malignancy	Tissue diagnosis	≥95%
6	Known biopsy-proven	Surgical excision when clinical appropriate	n/a

그림 5 BIRADS 분류표. 0에서 6까지 분류되어 있으며 높은 값을 가질수록 위험도가 높다고 볼 수 있음.

유방병변에 해당하는 데이터를 수집하기 위해 우선 판독문을 통해서 Study의 BIRADS Category 기준으로 study를 선별한다. BI-RADS는 유방영상 판독 프로토콜의 하나로, 전 세계 유방의들에게 사용이 권고되고 있고 국내 의료시스템에도 적용되어 있다.[23, 24] BIRADS category 가 4이상인 경우 유방병변이 악성일 가능성을 가지고 있다고 보고 조직검사의 대상이 되므로 유방병변 데이터를 모으기 위해서 BIRADS category 4이상인 study를 선별하여 수집하였다.

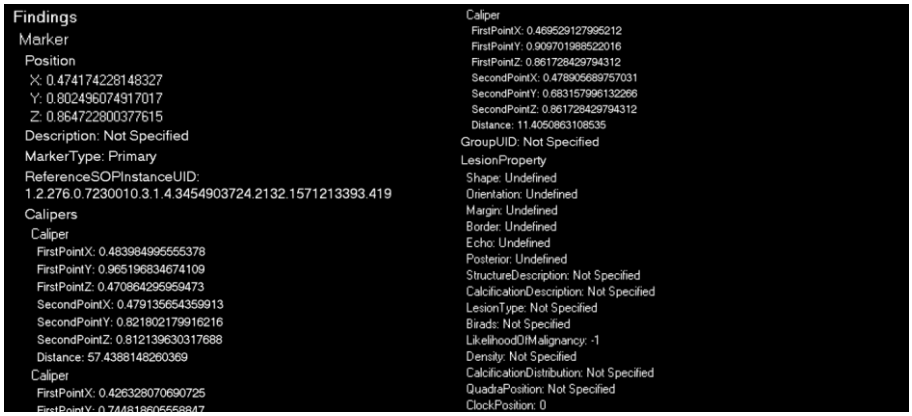


그림 6 Structured report 의 예시. 판독의들이 ABUS 판독시 기록한 위치나 크기정보들이 항목별로 저장되어 있음.

BIRADS를 통해 Study를 선별한 후 해당 study내의 유방병변의 정보를 각각의 Study에 해당하는 Structured report에서 추출해낸다. Structured report에 포함된 정보들은 그림 7에서 예시로 볼 수 있다. Structured report는 자동 유방초음파 영상의 각각의 study에 함께 생성되는 파일로 scan 이후 기록되는 여러가지 정보들을 저장하는 역할을 한다. 스캔측정후 방사선사에 의해 기록되는 Nipple의 위치나 영상의학과 전문의 판독 후 유방병변의 위치, 크기, 특성 등이 포함될 수 있다. 이때의 유방병변의 위치 및 크기 정보를 선별한 study내에서 추출한 후 해당 영역을 Cropping 하여 array로 저장하여 딥러닝 모델에 사용될 데이터로 활용한다.

일반적인 이미지의 라벨링 과정은 픽셀단위로 마킹하거나 바운딩 박스로 대상의 위치를 표시하는 것이 일반적이나 유방병변 후보 검출기의 목적은 오직 위치를 검출하면 되기 때문에 유방병변의 중심 좌표와 평균 지름을 이용하여 구(sphere)의 형태로 label array를 생성한다. 이때 label array를 Ground truth로

하여 딥러닝 모델을 훈련해 유방병변의 중심을 기준으로 활성화된 출력 배열을 생성하도록 하였다.

2.2.2. 알고리즘 구조

유방병변 후보 검출기의 구조는 Dense layer[25]를 활용한 encoder의 구조로 이루어져 있다. 유방병변 후보 검출기의 훈련을 위한 데이터는 입력값과 출력값이 모두 배열의 형태를 가지고 있기 때문에 기본적으로는 영역 분할 모델(Segmentation)의 구조로 인공신경망의 구조가 이루어져야 한다. 여러 연구에서 삼차원 의료영상 영역분할에 3-Dimensional Unet(3D Unet)을 적용하여 좋은 성능을 보였다.[25, 26] 3D Unet에서는 특징을 추출하는 encoder와 추출된 특징 맵에서 분할영역을 예측하는 decoder로 이루어 있다.

본 연구에서 제안하는 유방병변 후보 검출기는 일반적인 영역 분할 모델와 다르게 Ground truth 데이터가 명확하지 않고 대상영역의 모양을 예측하는 것이 아니라 대상의 중심으로 활성화만 되는 것이 목적이기 때문에 차이점이 있다고 할 수 있다. 따라서 유방병변 후보 검출기에서는 decoder 부분은 제외하고 encoder를 통해 입력 이미지에서 특징을 추출하고 활성화 영역을 추측하는 데에 그 기능을 집중시켰다.

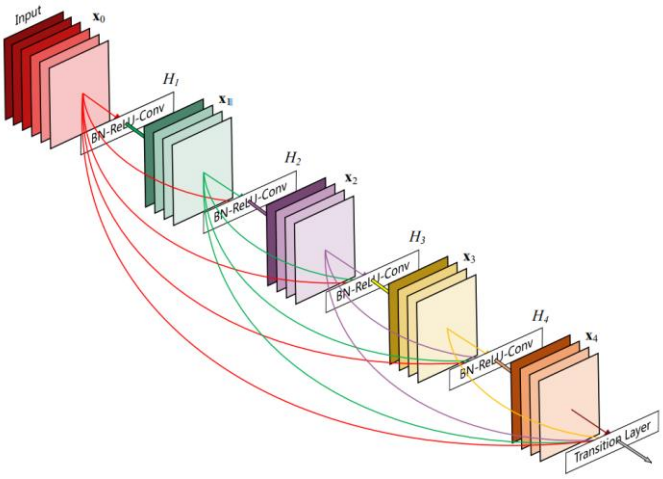


그림 7 Dense layer의 densely connected CNN layer들

유방병변 후보 검출기의 encoder는 DenseNet에서 사용되는 Dense layer(그림 8)를 활용하였다. DenseNet에서는 Densely connected 된 CNN 구조를 제안하였다. Densely connected 된 layer란 이전 CNN layer들의 특징맵을 계속해서 다음 layer의 입력과 더하여 축적하는 방식을 말한다.[25] 유방병변 후보 검출기에서는 Dense layer를 기존의 2차원에서 3차원으로 확장시키고, 각각 연속된 6개와 12개의 Dense layer로 구성된 두 개의 Dense block으로 이루어진 모델을 디자인하여 적용하였다. 유방병변 후보 검출기의 알고리즘 구조는 그림 9와 같다.

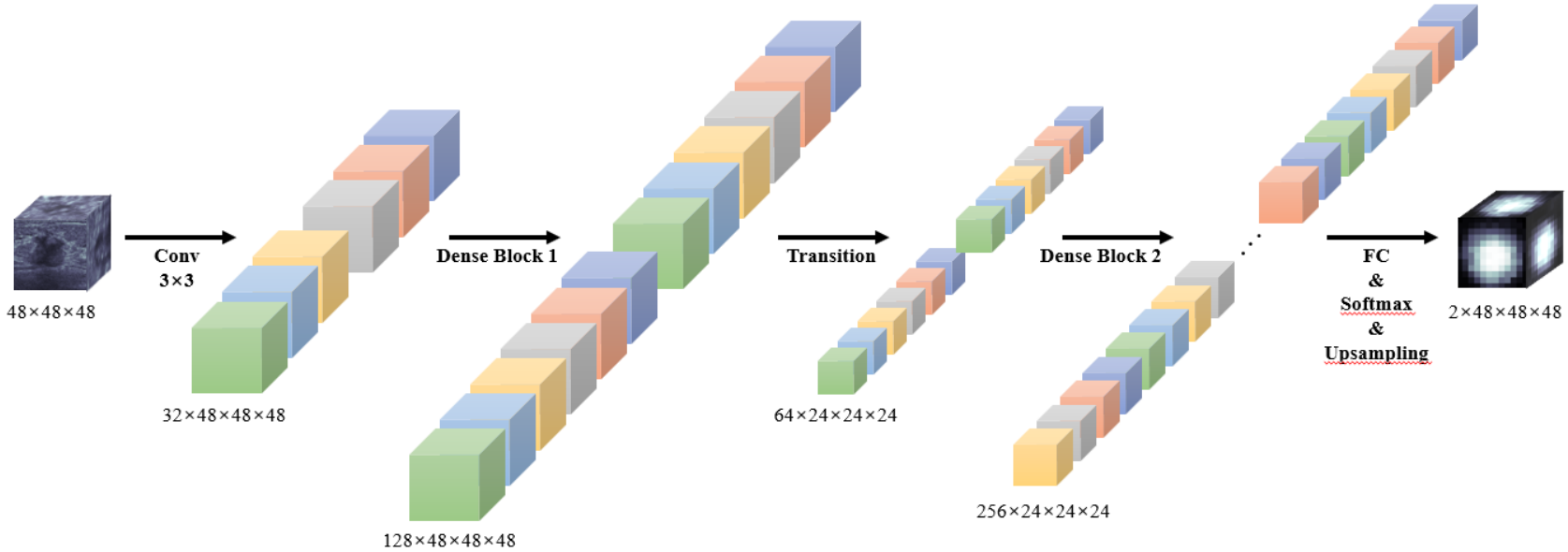


그림 8 유방병변 후보 검출기 알고리즘 구조. 연속된 6개와 12개의 Dense layer로 구성된 두개의 Dense block을 통해 유방병변의 위치를 예측함

2.2.3. 훈련 및 예측과정

유방병변 후보 검출기의 훈련과정에서는 2.2.1절에서 설명한 방법으로 구축된 데이터셋을 활용한다. 48×48×48 크기의 고정된 크기를 가진 자동초음파 영상을 알고리즘에 입력하여 출력된 예측 배열과 그에 대응되는 라벨 배열 간의 손실 값을 Cross entropy loss(식 1)로 계산하여 모델의 가중치를 업데이트한다.

$$L_{mask} = \frac{1}{NM}\sum_{j}^{N}\sum_{i}^{M}\left[\left(1 - z_{i,j}\right)\log\left(1 - q_{i,j}\right) + z_{i,j}\log q_{i,j}\right] \quad (1)$$

여기서 j 는 입력 배치 샘플의 인덱스를 의미하고, i는 $\mathrm{H} \times \mathrm{W} \times \mathrm{D}$ 공간에 있는 복셀의 인덱스, $\mathrm{M} = \mathrm{HWD}$는 알고리즘 출력의 총 복셀 수, $z_{i,j} \in \{0,1\}$는 label array $T \in R^{H \times \mathrm{W} \times \mathrm{D}}$의 픽셀 값이다

그림 10은 유방병변 후보 검출기의 예측과정을 나타낸다. 전체 자동 유방초음파 영상을 알고리즘에 입력하게 되면 100GB 이상의 메모리가 필요하므로 과도한 하드웨어 성능을 요구하게 된다. 이를 극복하기 위하여 유방병변 후보 검출기의 예측과정에서는 자동 유방초음파 영상을 Coronal view 기준으로 9등분하여 입력한 후 각각의 예측배열을 다시 하나로 통합하여 입력된 영상에 대한 예측 배열로 사용한다. 예측배열에서 설정한 역치값보다 높은 영역의 무게중심(Centroid) 좌표를 계산하여 이를 유방병변의 후보로 간주하고 다음 단계인 유방병변 분류기의 입력볼륨을 생성하는 데에 사용된다.

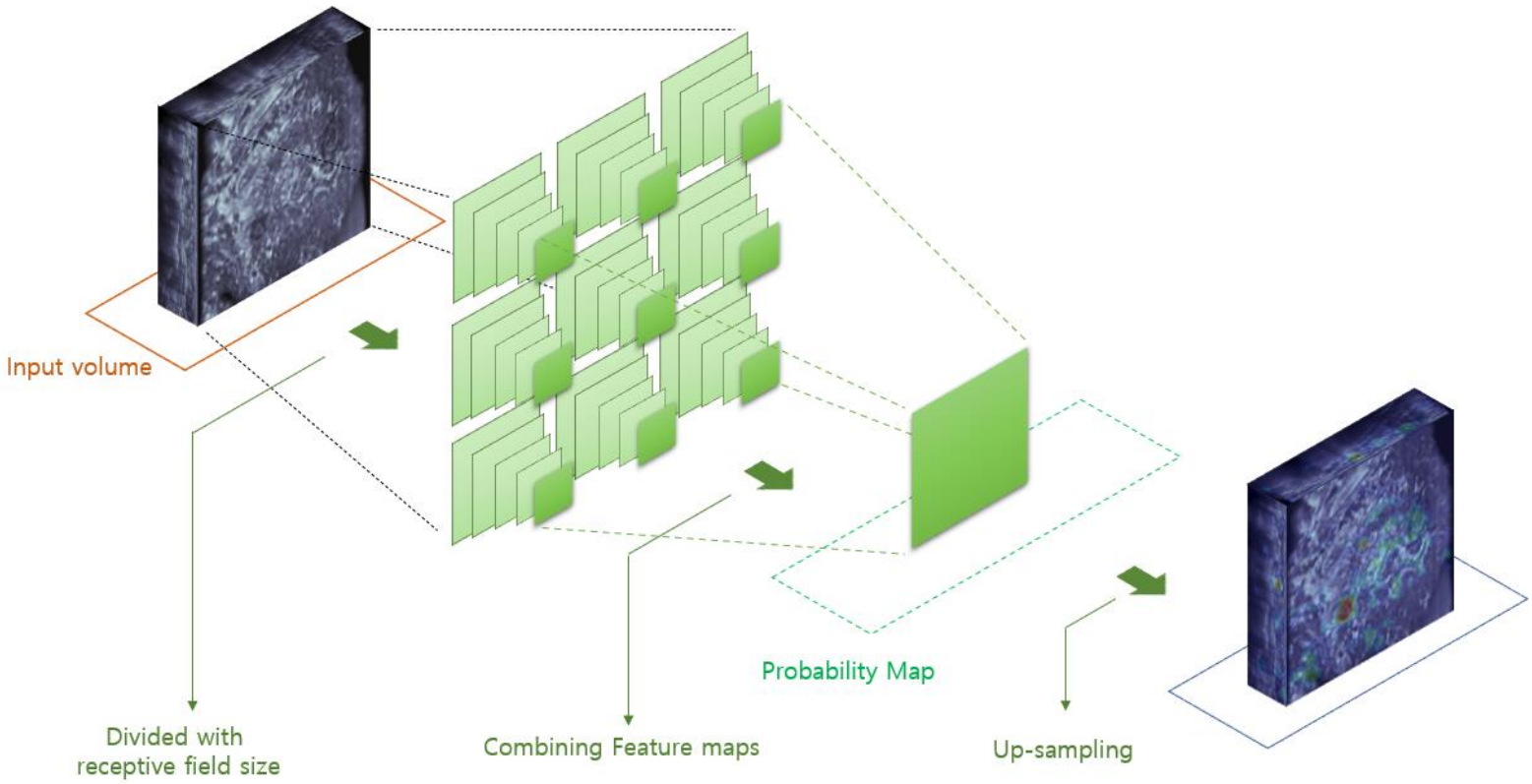


그림 9 유방병변 후보 검출기의 예측 과정. 입력된 자동 유방초음파 영상을 9등분하여 입력한 후 출력된 배열을 다시 하나로 통합하여 병변의 위치를 추측함.

2.3. 유방병변 유무 판별기

2.3.1. 알고리즘 구조

유방병변 유무 분류기의 알고리즘에는 Mask Branch Network 라는 새로운 구조를 제안한다. 제안된 방법의 전체적인 구조는 그림 8에 설명되어 있다. 제안된 알고리즘의 구조는 주 네트워크(Main Network)와 마스크 네트워크(Mask branch Network)라는 두 개의 병렬 네트워크로 구성된다. 주 네트워크는 CNN을 기반으로 한 Encoder 및 분류기로 작동한다. 모든 고성능 CNN 구조는 주 네트워크가 될 수 있으며, 본 연구에서는 DenseNet[25] 구조를 기반으로 하는 삼차원 이미지 분류기를 사용한다. 주 네트워크 중간에서 마스크 분기 네트워크는 연결되어 판독문에서 생성된 템플릿 마스크와 비교하여 손실을 계산한다. 템플릿 마스크에 대한 자세한 내용은 2.4 절에서 다루어질 것이다. 각 네트워크의 손실은 하나의 손실로 통합되어 모델을 훈련하는 과정에 사용된다.

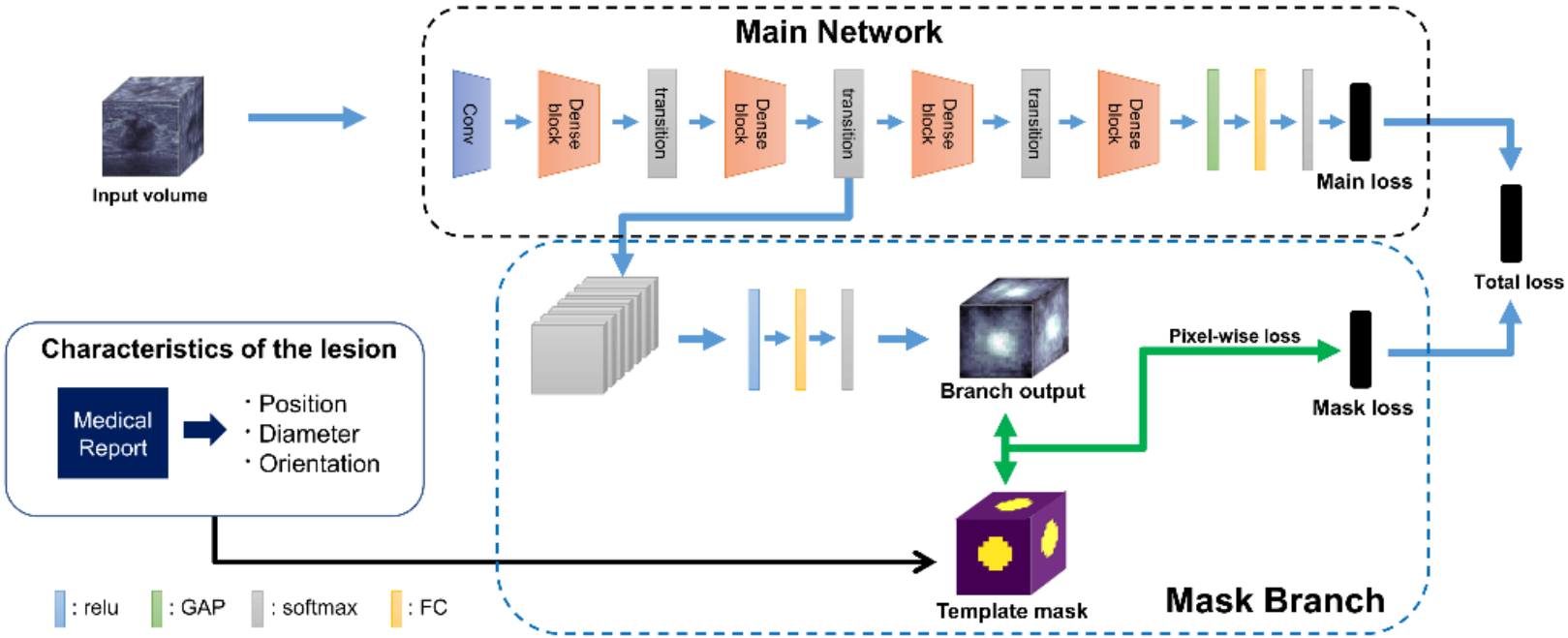


그림 10 Mask Branch network의 구조. 유방병변의 유무를 판별하는 역할을 하는 주네트워크(Main network)와 주네트워크의 효과적인 학습을 도와주는 마스크 네트워크(Mask Branch)로 이루어져있음

2.3.2. 주 네트워크

주 네트워크는 유방병변의 존재를 결정하는 유방병변 분류기를 구현한다. DenseNet-BC-121[25]을 기반으로 한 Transition layer가 있는 4개의 Dense Block으로 구성되어 있다. Transition layer는 Batch Normalization, 1×1×1 합성곱 신경망, Average Pooling을 수행하기 위해 두 개의 Dense Block 사이에 배치된다. Global Average Pooling(GAP) 계층은 마지막 Dense Block의 끝에 연결되고, 그다음에는 Fully Connected layer(FC) 와 Softmax activation layer가 순차적으로 연결된다. 우리는 주 손실을 수식 (2)과 같이 입력 샘플의 Cross entropy loss로 정의한다.

$$L_{main} = \frac{1}{N}\sum_{j}^{N}[(1-y_j)\log(1-p_j) + y_j \log\, p_j] \qquad (2)$$

여기서 j는 입력 샘플의 인덱스를 의미하며, N은 샘플 수이고, $y_j \in \{0,1\}$은 Ground Truth 라벨이며, p_j는 입력 샘플이 유방병변을 포함하는지 여부를 추정하는 확률값이다.

2.3.3. 마스크 네트워크

분기 아키텍처의 아이디어는 서로 다른 네트워크 간의 상호작용을 활용하여 딥러닝 모델을 훈련하는 것이다. [26,27] 마스크 네트워크는 유방병변의 공간적인 정보를 주 네트워크에 통합하는 것을 목표로 하며, 이는 주 네트워크가 유방병변 주위의 영역에서 의미 있는 특징을 추출하도록 돕는다. 마스크 네트워크는 유방병변의 위치와 모양 정보를 포함한 마스크 배열을 예측하는 Branch Output을 생성한다. 마스크 네트워크는 주 네트워크의

중간에 연결된 다음, 연결된 지점 이전의 계층들의 특징을 마스크 네트워크 및 주 네트워크에 의해 공유된다.

Branch output과 마스크의 손실 최적화를 통해 마스크 네트워크와 주 네트워크가 공유하는 부분의 네트워크는 대상의 공간 정보를 학습할 수 있다. 이 과정에서 네트워크는 유방병변이 존재하는 영역에 집중하여 학습하게 되고, 공간적인 강조 효과와 함께 분류기가 더 잘 훈련된다.

본 연구에서는 마스크 네트워크가 1×1×1 합성곱 계층과 두 번째 transition layer의 pooling layer 사이의 위치에 있는 것을 제안하지만, 마스크 네트워크는 주 네트워크의 어떤 위치에도 연결될 수 있다. 입력 영상 $I \in R^{H \times \mathrm{W} \times \mathrm{D}}$, $F \in R^{H' \times W' \times D'}$ 는 메인 네트워크의 중간에서 추출된 특징 맵(feature map)으로 마스크 네트워크의 입력으로 사용되며, 여기서 H', W', D'는 각 축의 높이, 폭, 깊이, H, W, D의 절반을 의미한다. Branch output B 는 수식 3와 같이 표현할 수 있다.

$$B(\mathcal{I}, w) = softmax\left(f\left(\sigma\left(\mathcal{F}(\mathcal{I})\right), w\right)\right) \qquad (3)$$

여기서 $f(\cdot,\cdot)$은 3D 합성곱 함수를 나타내며, σ는 활성화 함수를 나타내고, $B \in R^{H' \times W' \times D' \times 2}$ s는 생성된 Branch output을 의미하며, $w \in R^{2 \times \mathrm{C} \times 1 \times 1 \times 1}$ 는 합성곱 함수 내에서 사용되는 가중치 배열을 나타낸다. 결과적으로 Branch output은 유방병변의 유무에 따른 복셀 수준 이진 확률 배열이다.

Branch output과 마스크 간의 비교를 통해 마스크 손실, 즉 픽셀 단위의 Cross entropy loss가 발생하며, 이를 통해 모델은 픽셀 단위 클래스 확률을 학습한다. 대상 병변의 공간 정보가

포함된 마스크를 사용하면 branch output과 마스크 사이의 픽셀 단위 방식으로 마스크 손실이 계산됩니다. 마스크 손실의 정의는 수식 (4)와 같다.

$$L_{mask} = \frac{1}{NM}\sum_{j}^{N}\sum_{i}^{M}\left[\left(1 - z_{i,j}\right)\log\left(1 - q_{i,j}\right) + z_{i,j}\log q_{i,j}\right] \quad (4)$$

여기서 j 는 입력 배치 샘플의 인덱스를 의미하고, i는 $\mathrm{H} \times \mathrm{W} \times \mathrm{D}$ 공간에 있는 복셀의 인덱스, $\mathrm{M} = \mathrm{HWD}$ 는 분기 출력의 총 복셀 수, $z_{i,j} \in \{0,1\}$는 마스크 $T \in R^{H \times \mathrm{W} \times \mathrm{D}}$의 픽셀 값이다.

최적화에 사용되는 전체 모델의 총 손실은 주 손실과 마스크 손실의 조합이다. 두가지 손실값을 통합하는 과정에서 효과적으로 결합하기 위해 다중 작업 손실에 일반적으로 사용되는 Uncertainty loss weight method[28]을 수식 (5)와같이 손실을 결합하기 위해 채택한다.

$$L_{total} = \frac{1}{\sigma_1^2}L_{main} + \frac{1}{\sigma_2^2}L_{mask} + \log\sigma_1 + \log\sigma_2 \quad (5)$$

여기서 σ_1 과 σ_2 는 L_{main} 과 L_{mask} 의 상대적 무게를 유동적으로 학습하고 각 손실간의 균형을 조절하는 훈련 가능한 변수이다.

2.3.4. 네트워크 설계 후보군

본 연구에서 그림 11과 같이 MBN-V1에서 MBN-V4까지 4가지 유형을 가진 마스크 네트워크를 설계했다. 우리는 보조 네트워크의 수, 네트워크의 연결 위치, 복수의 보조 네트워크를 사용할 경우 출력을 결합하는 방법 등 여러 가지 요소를 가진 마스크 네트워크의 아키텍처를 탐구했다. 마스크 네트워크는 2.3.3

절에서 설명한 MBN-V1에과 동일하지만 마스크 네트워크의 적절한 구조는 메인 네트워크와 적용되는 작업에 의해 다양해질 수 있다. 우리는 메인 네트워크와 병렬로 단일 또는 다중 하위 네트워크를 설계한다.

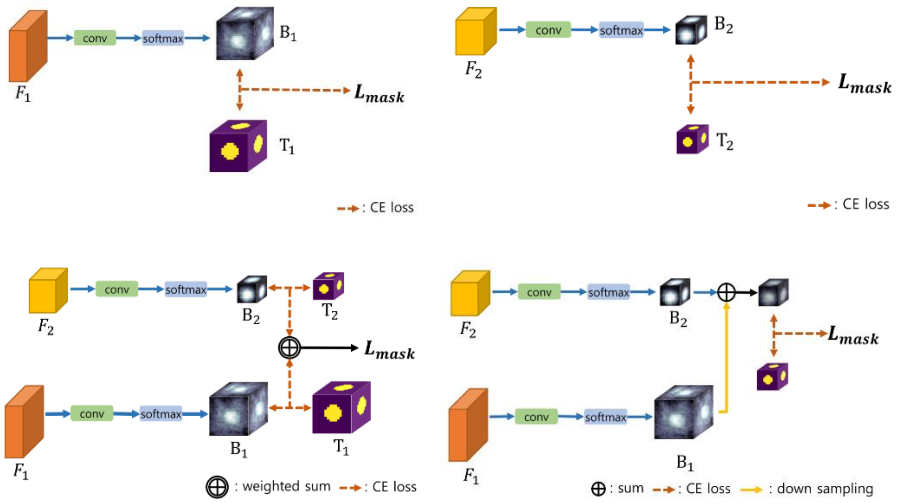


그림 11 마스크 네트워크의 여러가지 구조. (1행 1열) MBN-V1. (1행 2열) MBN-V2. (2행 1열) MBN-V3. (2행 2열) MBN-V4

그림 11의 좌측상단에서 볼 수 있듯이 MBN-V1은 두 번째 transition layer에 연결되어 특징 맵 F_1을 추출하고 결과적으로 B_1이 된다. 마스크 손실 L_{mask}는 B_1과 T_1의 차이를 줄이는 데 사용된다. MBN-V2는 그림 11에서 우측상단의 F_2, B_2, T_2와 같이 주 네트워크에서 세 번째 Transition layer에 통합된다. MBN-V1 및 MBN-V2와 달리 마스크 분기 네트워크의 다른 버전은 MBN-V1 및 MBN-V2에 사용된 마스크 네트워크들의 조합으로 구성된다. MBN-V3에서는 두 분기가 분리되어 그림 11의 좌측하단과 같이 서로 크기가 다른 B_1과 B_2를 출력한다. Branch output은 T_1 및 T_2와

대응하여 두 개의 마스크 손실을 계산한다. 마지막으로, 각각의 손실은 uncertainty loss weights method[12]를 통해 결합되고 마스크 네트워크는 각각 다른 손실에 의해 최적화된다. 이와 다르게 MBN-V4는 그림 11의 우측 하단에 표현되어 있는 것처럼, 출력 맵을 더하기 연산을 통해 하나의 단일 배열로 통합했다. MBN-V4에서, B_1이 3D 평균 풀링(Average pooling)을 통해 B_2과 동일한 모양으로 축소된 후, B_1 과 B_2 는 요소별 합계에 의해 통합된다. 마스크 네트워크는 모델을 훈련하는 과정에서 관심 영역에 집중할 수 있도록 하는 네트워크이므로 추론 단계에서 계산 복잡성을 증가시키지 않는다.

2.3.5. 템플릿 마스크

자동 유방초음파 영상의 병변이 아닌 영역의 일부 음영은 종종 유방병변의 음영과 구별하기 어렵다. 분할 마스크의 경우 임상 전문가가 완전히 표기한 픽셀 단위 분할 배열이 질량에 이상적인 데이터가 될 수 있지만, 3차원 영역의 경계를 수동으로 표기하려면 전문가의 많은 시간과 노력이 필요할 것이다. 이 문제를 해결하기 위해, 수동으로 대상 영역을 레이블링하는 대신 판독과정에서 기록되어 이미 존재하는 정보를 활용할 것을 제안한다.

2.3.6. 유방병변의 범주

의심스러운 병변의 위치, 크기 및 범주는 진단 중에 일반적으로 판독문에 기록된다. 유방병변의 세방향에 따른 지름의 길이를 측정함에 따라 유방병변의 위치와 크기가 중심좌표와 지름의 형태로 기록된다. 판독과정에서 적용되는 BIRADS의

요인중에 'shape', 'orientation', 'margins', 'posterior acoustic feature' 등의 일부 시각적 특성은 유방 영상 진단 중에 판독문에 기록된다. 그러한 특성 중, 우리는 유방과 상대적인 긴 축의 방향을 나타내는 orientation 특징을 사용한다. 유방병변은 병변의 장축이 피부표면에 평행한지의 여부에 따라 'parallel' 과 'not parallel'로 구분된다. 이러한 속성을 활용하여 공간 정보를 포함하고 분류기에 특징을 추출할 위치를 안내하는 단순화된 레이블 마스크인 '템플릿 마스크'를 자동으로 생성할 수 있다. 다른 다양한 유방병변의 특징들을 사용할 수 있지만 'orientation'을 사용하는 것만으로도 효과가 있었고 그 비교 결과도 결과 및 분석 절에서 설명할 것이다. 사용 가능한 정보가 제한되어 있기 때문에, 이 간단한 마스크는 주석 프로세스에 의해 얻어진 정확한 분할 레이블 맵으로 간주할 수는 없으나, 네트워크가 유방병변이 위치한 곳으로부터 유의미한 특징을 추출하도록 유도하기에는 충분한 정보를 포함하고 있다.

2.3.7 템플릿 마스크 생성과정

일반적인 분할 맵과 마찬가지로 템플릿 마스크는 양성 및 음성의 영역으로 구성된 이진 레이블 배열이다. 양성 영역은 유방병변의 위치, 직경 및 방향을 사용하는 규칙에 의해 생성된다. 템플릿 마스크의 모양은 구 또는 타원체의 두 가지 유형 중 하나이다. 유방병변의 orientation 특성이 'parallel'인 경우 피부에 수직인 직경이 절반으로 줄어들어 양성 영역의 모양이 구에서 타원체로 바뀐다. 유방병변의 중심 좌표 (x_t, y_t, z_t), 직경 그리고 방향을 사용하여 단순화된 분할 레이블인 템플릿 마스크

$T(x_t, y_t, z_t)$를 식 (6)와 같이 정의할 수 있습니다.

$$T(x_t, y_t, z_t) = \begin{cases} 1 & if \ \frac{(x-x_t)^2}{r_t} + \frac{(y-y_t)^2}{\alpha r_t} + \frac{(z-z_t)^2}{r_t} \leq 1 \\ 0 & else \end{cases} \quad (6)$$

여기서 α 는 orientation 특성이 parallel일 경우 0.5 이고 아닐 경우에는 1이된다.

그림 12는 각 단계별로 템플릿 마스크를 만드는 프로세스를 보여준다. 먼저, 템플릿 마스크는 Structured report에 기록된 병변의 (x_t, y_t, z_t) 중심 위치에 중심이 있는 구(球)로 시작한다. 구면 반지름은 유방병변의 세 축의 방향에 따른 반지름의 평균으로 정의된다. 마지막으로, 유방병변이 'parallel' 방향의 특성을 가지고 있는 경우, 피부에 수직인 직경은 절반으로 줄어들면서 양의 영역의 모양은 구에서 타원체로 바뀐다.

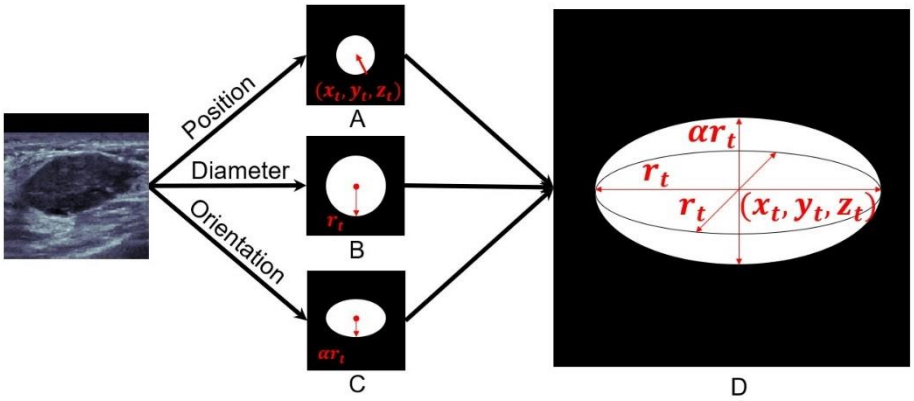


그림 12 템플릿 마스크의 생성과정. 유방병변의 위치, 크기, 모양 특성을 고려하여 병변의 공간적 정보를 포함하는 템플릿 마스크를 생성할 수 있음.

그림 13은 템플릿 마스크의 한 예시이다. 그림 15-(a), (b), (c) 는 유방병변의 중앙의 위치에서의 Transverse, coronal, sagittal 단면이 표시되어 있다. 각 축에 유방병변의 직경이 10mm, 18mm,

15mm인 경우 병변의 평균 직경은 14.3mm이다. 템플릿 마스크의 중심은 병변의 중심과 동일한 위치에 위치한다. 예시의 유방병변은 'parallel' 특징을 가지기 때문에, 피부표면에 수직인 반직경이 절반이 되어 단축과 장축의 반직경이 각각 7.15mm, 14.3mm인 타원체의 모양을 띄며 그 단면은 그림 15-(d), (e), (f)에 나타낸 것과 같다.

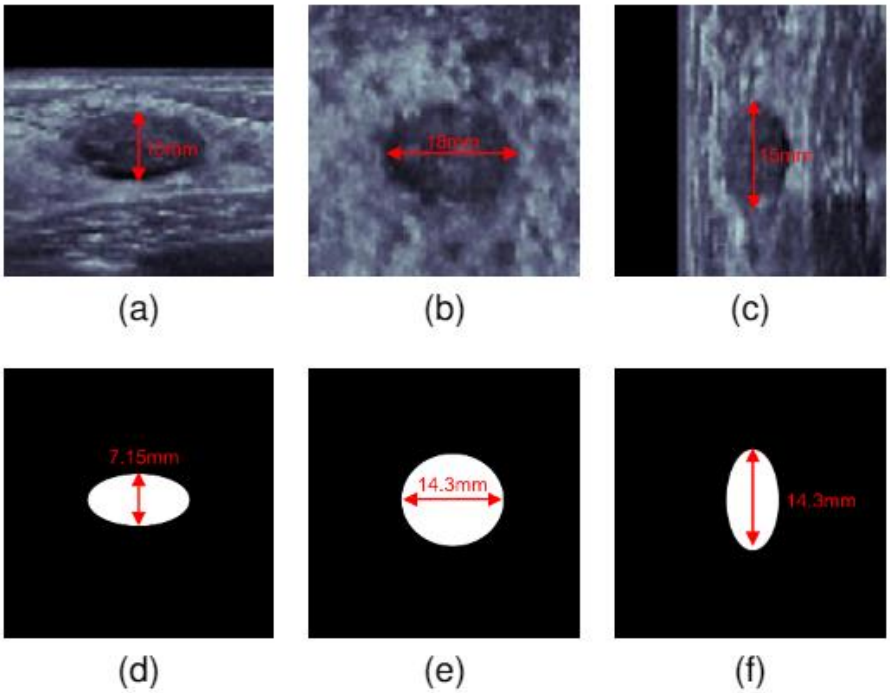


그림 13 템플릿 마스크의 한 예. 각 열은 transverse, coronal, sagittal 단면을 의미함. (행 1) 대상이 되는 유방병변. (행 2) 대상병변을 이용해 생성된 템플릿 마스크

2.4. 악성 유방병변 분류기

2.4.1 알고리즘 구조

악성 유방병변 분류기는 알고리즘 유방병변 유무 판별기와 유사하게 삼차원 DenseNet[25]의 구조로 이루어져 있다. DenseNet-121-BC[25]의 구조의 이차원 합성곱 신경망을 삼차원 합성곱 신경망의 구조로 변경하여 자동 유방초음파 영상에서 유방병변의 볼륨을 처리할 수 있도록 하였다. 유방병변 악성도 분류기의 경우는 다른 분류기에 비해서 더 많은 정보를 필요로 하기 때문에 영상에서의 손실은 최소화하고 모델에 저장되는 특징은 최대화해야 한다. 따라서 기존의 DenseNet-121-BC와 다른 점은 두번째 transition layer에서 pooling layer를 제거하였다. 그 이유는 유방병변 악성도 분류기의 경우에는 다른 분류기에 비해서 더욱 미세한 정보들을 요구하는 반면에 pooling layer가 많이 들어갈수록 데이터의 해상도가 떨어지고 또 입력 이미지의 크기가 48mm 밖에 되지 않기 때문에 pooling layer에 의한 정보손실을 최소화하기 위해서이다. 또 일반적으로는 모델에 입력되기 전에 전처리 과정을 통하여 x, y, z 축의 비율을 동일하게 맞추어주나, 자동 유방초음파 이미지의 경우에는 같은 x, y, z 세 방향의 pixel spacing이 약 0.2, 0.82, 0.5로 불균형이 심하기 때문에 그 과정에서 정보 손실이 발생할 수 있다. 따라서 먼저 전처리과정을 거치치 않은 상태로 모델에 입력시켜서 첫번째 합성곱 신경망을 통해서 특징을 추출한 후에 특징맵 상태에서 2×4×1 average pooling layer를 통해 세 축 간의 불균형을 조절해준다. 유방병변 악성도 분류기의 구조는 그림 14 와 같다.

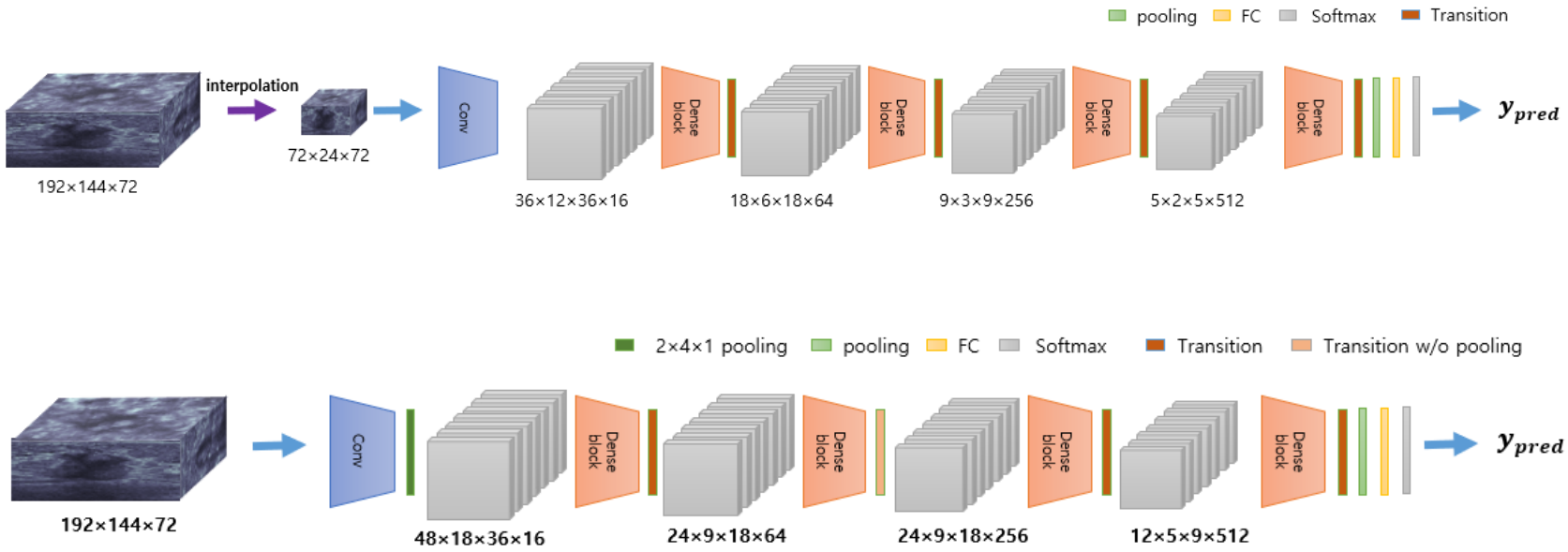


그림 14 유방병변 악성도 분류기의 알고리즘 구조. (행 1) 기존의 DenseNet을 활용한 분류기. (행 2) 본 연구에서 제안하는 분류기 구조.

2.4.2 훈련 및 예측과정

본 연구에서 유방병변 악성도 분류기에 입력 이미지는 검출된 유방병변의 중심부에서 192×144×72의 사이즈로 크롭한 이미지를 사용한다. 입력된 이미지는 인공신경망 알고리즘을 통과한 후 악성도를 예측하는 스칼라 값을 출력한다. 유방병변 악성도 분류기의 알고리즘의 마지막 단계에서 Softmax layer가 적용되므로 출력값은 0~1사이의 값을 가지게 되고, 설정된 역치값 이상의 대상에 대해서 악성 유방병변으로 판단하고 최종 출력물로 선정하게 된다.

2.5. 실험 방법

2.5.1. 실험 목적

본 연구를 통해 개발한 유방암 검출 시스템의 성능을 검증해야 한다. 시스템은 총 세 부분으로 나뉠 수 있다. 유방병변 후보 검출기, 유방병변 유무 판별기, 악성 유방병변 분류기이다. 따라서 본 연구의 실험에서는 각각의 단계에서의 성능을 평가한다. 그리고 세가지 기능을 통합한 전체적인 시스템의 성능도 평가한다. 이 모든 부분을 포함하는 실험을 설계하였다.

특히 유방병변 분류기에 적용된 템플릿 마스크와 마스크 네크워크의 효과를 확인하기 위해 기존의 다른 모델과 다양조건의 제안된 기법을 적용한 모델의 성능을 비교하여 현재 유방암 검출기에 적용된 기법의 효과성을 확인한다.

이 실험의 목적은 각각의 모듈과 전체 시스템의 성능측정을 위해 선정된 테스트 셋에 대하여 성능을 평가하고 그 유효성을

판별하는 데에 있다.

2.5.2. 데이터셋 구성

본 연구에서 사용된 자동 유방초음파 영상의 경우 분당 서울대병원에서 GE healthcare사의 ABUS invenia 시스템을 통해 스캔된 이미지를 사용하였다. 본 연구에는 2017년 5월부터 2019년 10월사이에 자동 유방초음파를 촬영한 환자 중 363명의 환자가 포함되었는데, 이 환자군은 286명의 유방병변을 가진 환자와 77명의 정상 환자로 구성되어 있다. 286명의 유방병변을 가진 환자군에서 BIRADS C2 이상으로 분류된 443개의 유방병변을 포함했다. 443개의 유방병변은 BIRADS에 따라 C2, C3, C4 또는 5, 그리고 C6로 구분하였고 각각 188, 94, 34, 127 개의 유방병변을 포함하고 있다. 본 연구에서 데이터셋은 무작위로 학습, 검증, 테스트 셋으로 구분되었다. 학습 세트의 병변 304개, 검증 세트의 병변 59개, 테스트 세트의 병변 80개로 구성되어 있다. 또한 데이터 세트에는 77명의 일반 환자에서 무작위로 잘라낸 3898개의 정상 영역이 포함되었다. 정상 영역은 무작위로 구분되었다. 3777개의 영역은 교육 세트로, 59개의 영역은 검증 세트로, 80개의 영역은 테스트 세트로 분할되었다. 악성도 분류기를 학습하고 평가할 때에는 C3~C5 영역은 악성 여부가 불분명하므로 제외하고 나머지 188, 127개의 C2, C6 유방병변을 사용하였고, 각 28개, 48개씩 검증 및 테스트셋으로 사용하고 나머지는 학습용으로 사용하였다. 이를 도식화 하면 그림 15와 같다.

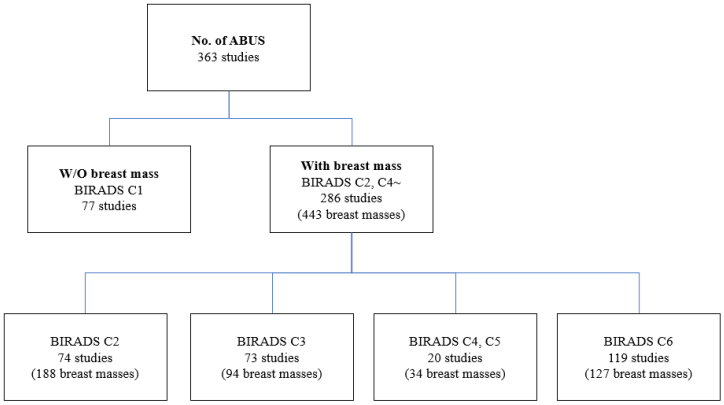


그림 15 훈련 및 검증용 데이터셋의 구성. BIRADS에 따라 환자군을 구분하여 각각의 모듈을 학습시키고 검증하는 데에 사용될 수 있도록 함.

전체 유방암 검출기의 성능을 평가하기 위해서는 앞서 언급한 데이터 셋과는 구분된 129명의 유방암을 포함한 환자군을 구성하여 시스템 전체의 성능을 평가하였다. 2017년 12월부터 2018년 2월까지 분당 서울대 병원에서 유방암을 가지고 있는 147명의 환자를 수집하였다. 그 중 ABUS 탐촉자로 촬영할 수 없는 큰 크기의 유방병변을 가진 3명의 환자와 ABUS 상으로 확인이 힘든 15개 케이스를 제외한 129개의 케이스가 대상으로 선정되었다. 비교 평가를 위하여 동일 데이터셋으로 기존의 ABUS CAD를 평가한 결과를 함께 비교하였다 참고하였다.[37] 비교 대상이 되는 CAD는 Qview CAD로 미국 FDA 승인을 받아 검증받은 CAD라 할 수 있다.

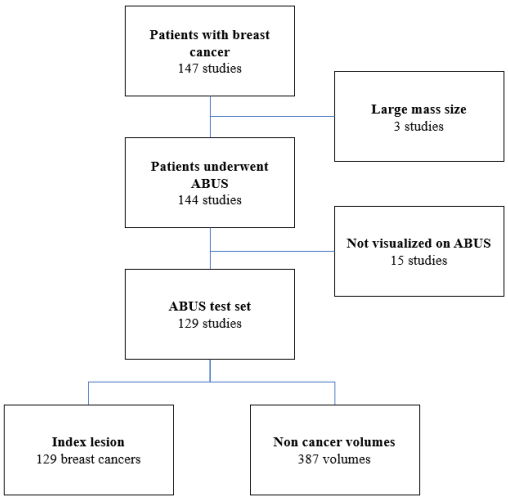


그림 16 유방암 검출 시스템의 성능 평가에 사용된 테스트 셋의 구성.

본 임상시험에 사용된 의료영상들은 분당서울대학교병원 임상 시험 심사위원회(Institutional Review Board, IRB)에서 안전하다고 판단되어 심의 면제 판정을 받아 진행하였다.

2.5.3 알고리즘 학습 및 검증

알고리즘 학습에 사용된 크롭 데이터는 유방병변으로 표시된 위치의 중심에서 10mm 거리 내에 있는 지점에서 일정 크기로 잘라낸 볼륨이다. 유방병변의 모든 중심 좌표는 숙련된 영상의학과 전문의가 분류했다. 크롭되는 크기는 유방병변 유무기 같은 경우는 0.5mm 로 픽셀 사이즈를 맞춰준 후 80×54×80 사이즈로 크롭하였다. 유방병변 유무 판별기의 경우는 1mm 사이즈로 픽셀 사이즈를 맞춘 후 48×32×48 사이즈로 크롭하였다.

유방병변 유무기와 유방병변 악성도 분류기는 크롭 데이터셋 중 테스트 셋으로 성능을 평가하였으며, 성능을 검증하기

위해 민감도(식 7), 특이도(식 8), 수신자 조작특성곡선(ROC)의 하위 면적(AUC)를 지표로 하여 평가하였다.

$$Sensitivity(Se) = \frac{TP}{TP + FN} \quad (7)$$

$$Specificity(Sp) = \frac{TN}{TN + FP} \quad (8)$$

유방병변 후보 검출기와 전체 시스템은 자동 유방 초음파 전체영상을 입력하여 성능을 평가하였다. 평가 지표로는 전체 유방암 또는 유방병변 중 검출에 성공한 비율을 뜻하는 검출률(Detection rate)과 위양성 검출의 수를 전체 볼륨개수로 나눈 위양성률(False positive rate, FPR)을 기준으로 삼았다. 유방병변 후보 검출기의 경우는 테스트셋 중 유방병변이 포함된 80개중 1개의 유방병변만 포함된 64개의 영상을 통하여 평가하였고, 전체 시스템의 경우는 별도로 구분한 129개의 케이스로 평가하였다. 129개를 평가할 때에 검출률은 유방암이 존재하는 자동 유방초음파영상내에서 가장 심각도가 높은 유방암을 index lesion으로 설정하고 전체 index lesion 중 검출에 성공한 index lesion의 비율을 의미한다.

2.5.3 실험 환경

본 연구에서 알고리즘은 Tensorflow framework 로 구현되었으며 Ubuntu 18.04 시스템의 엔비디아 2080-Ti에서 훈련되었다. 합성곱 레이어의 가중치는 Xavier initializer[29]에 의해 초기화되었다. 각 합성곱 계층 다음에는 항상 배치 정규화 함수(Batch Normalization)[30]와 활성화 함수(Relu)를 수행했다. 본 연구에서 사용한 모델은 동일한 크기의 미니 배치 2개를 가진 48개의 배치

크기로 훈련되었으며 손실은 RMSprob 최적화 함수에 의해 최적화 되었다. 30,000 step의 훈련 후, 검증 데이터 세트의 AUC 값이 가장 높은 모델을 테스트 세트에 의해 평가하였다.

결과 및 분석

3.1 유방병변 후보 검출기 성능

본 연구에서는 유방병변 후보 검출기에 적용된 알고리즘 이외에 기존의 3D Unet을 변형한 모델도 같이 유방병변 후보 검출기에 적용하여 평가하였다. 비교군에 해당하는 모델은 기존의 3D Unet에서 입력 영상의 크기와 딥러닝 환경의 메모리를 고려하여 가능한 범위내에서 모델의 합성곱 계층의 수를 조정하고, 데이터 특성상 decoder가 많을수록 불리하므로 마지막 역 합성곱 계층(de-convolution layer)를 삭제하였다. 비교군에 해당하는 Simple 3D-Unet V1과 V2는 그림 17과 그림 18에 나타나 있다.

표 1는 각 모델의 성능평가 결과를 나타낸다. Simple 3D-Unet V1의 경우는 민감도 95.3%로 비교적 높은 결과를 보였으나 위양성률이 444이 넘는 결과가 나와서 위양성률 문제를 가지고 있었다. Simple 3D-Unet V2의 경우는 위양성률은 117.5로 Simple 3D-Unet V1보다 낮은 결과를 보였으나 민감도가 54.7%이므로 놓치는 유방병변의 수가 너무 많았다. 마지막으로 본 연구에서 제안하는 네트워크는 민감도 98.4% 위양성률 108.8로 비교군보다 민감도와 위양성률 모두에서 높은 결과를 보였다.

표 1 유방병변 후보 검출기의 성능 비교표

Model	Sensitivity	FPR
Simple 3D Unet V1	95.3%	444.7
Simple 3D Unet V2	54.7%	117.5
Proposed Network	98.4%	108.8

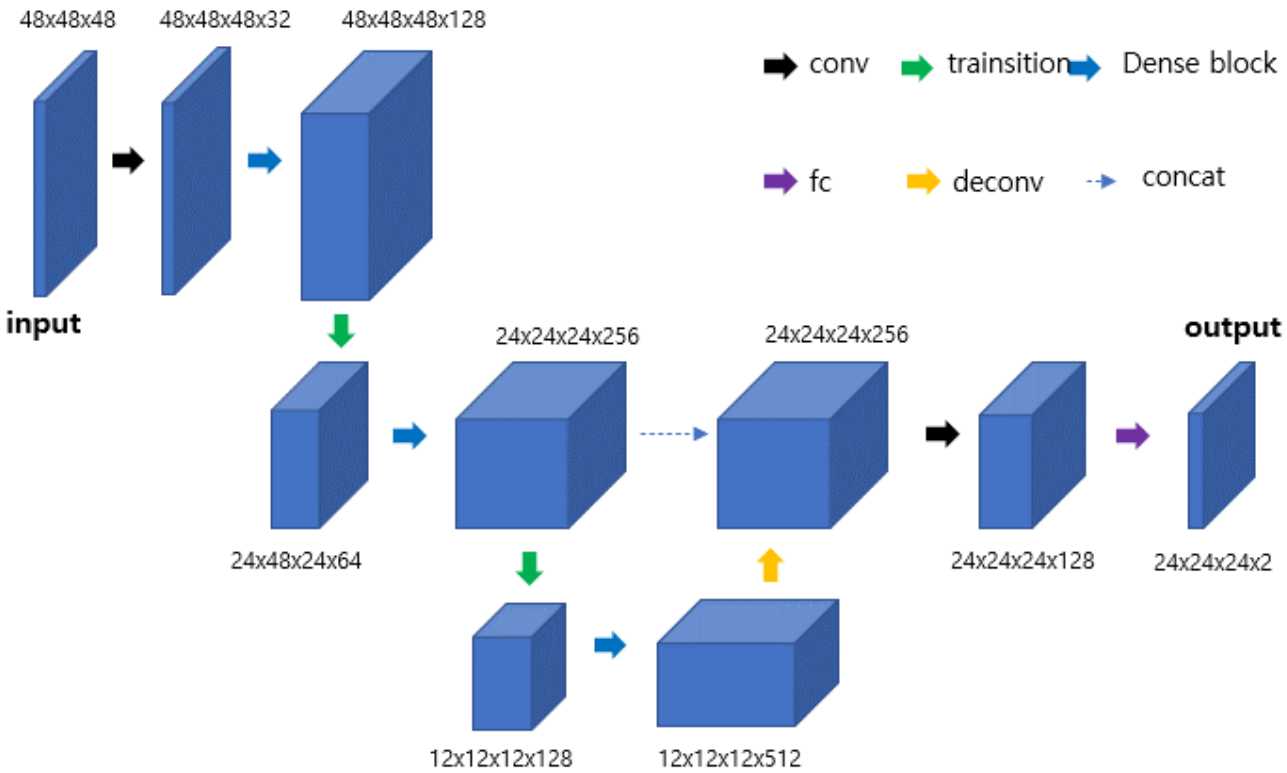


그림 17 Simple 3D Unet V1의 알고리즘 구조도. 2개의 dense block의 encoder와 1개의 deconvolution layer의 decoder로 이루어져 있음.

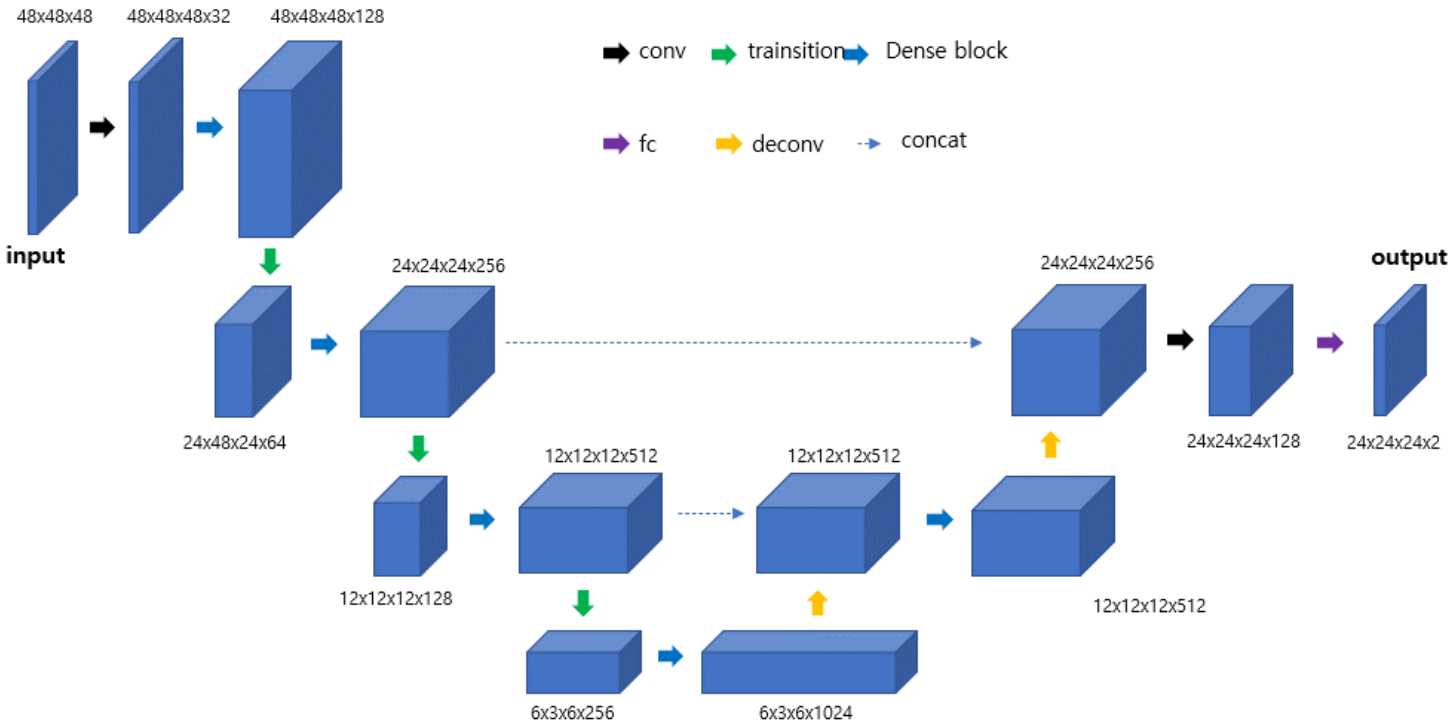
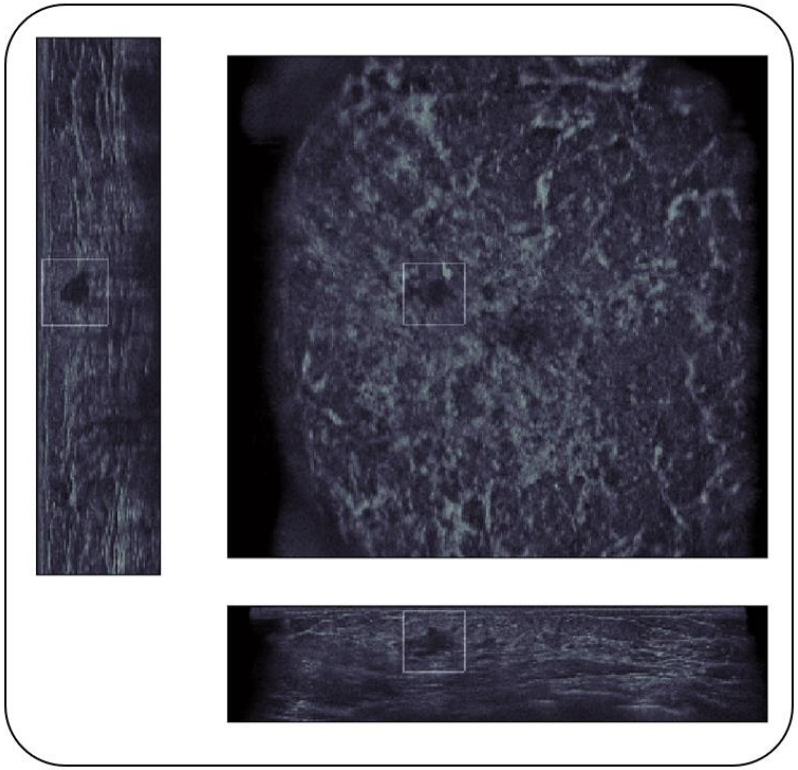
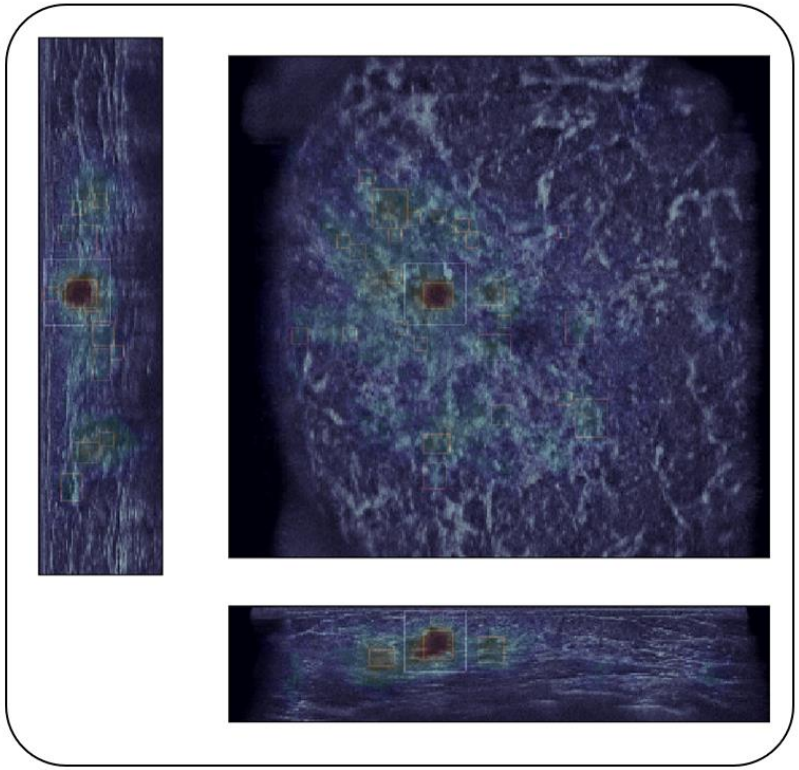


그림 18 Simple 3D Unet V2의 알고리즘 구조도. 3개의 dense block의 encoder 와 2개의 deconvolution layer의 decoder로 이루어져 있음





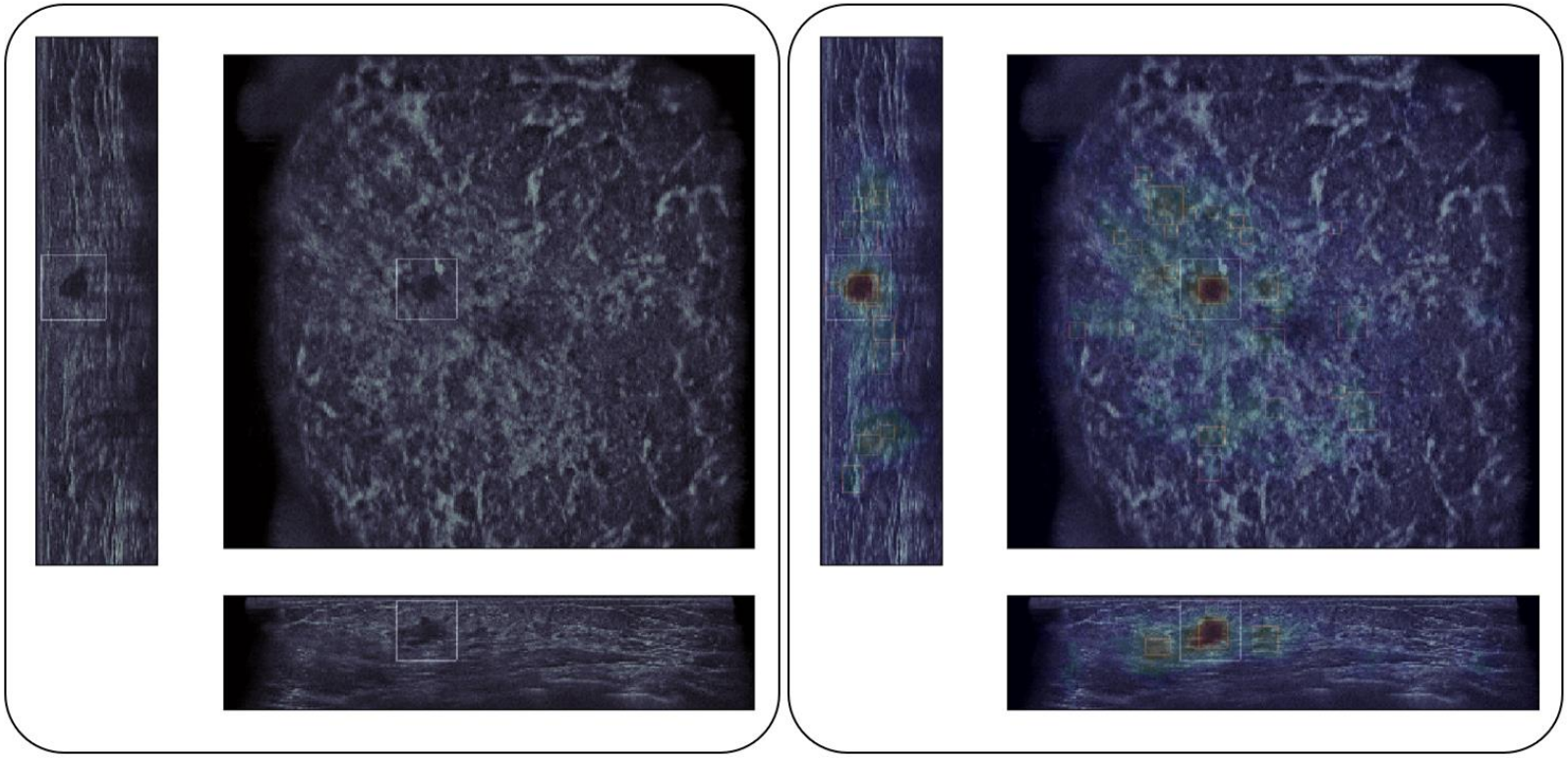


그림 19 유방병변 후보 검출기의 결과 예시1. (왼쪽) 자동 유방초음파 영상의 유방병변이 존재하는 위치의 단면으로 하얀색으로 표시된 부분이 유방병변의 위치 (오른쪽) 유방병변 후보 검출기의 결과로 주황색으로 표시된 부분이 검출결과임

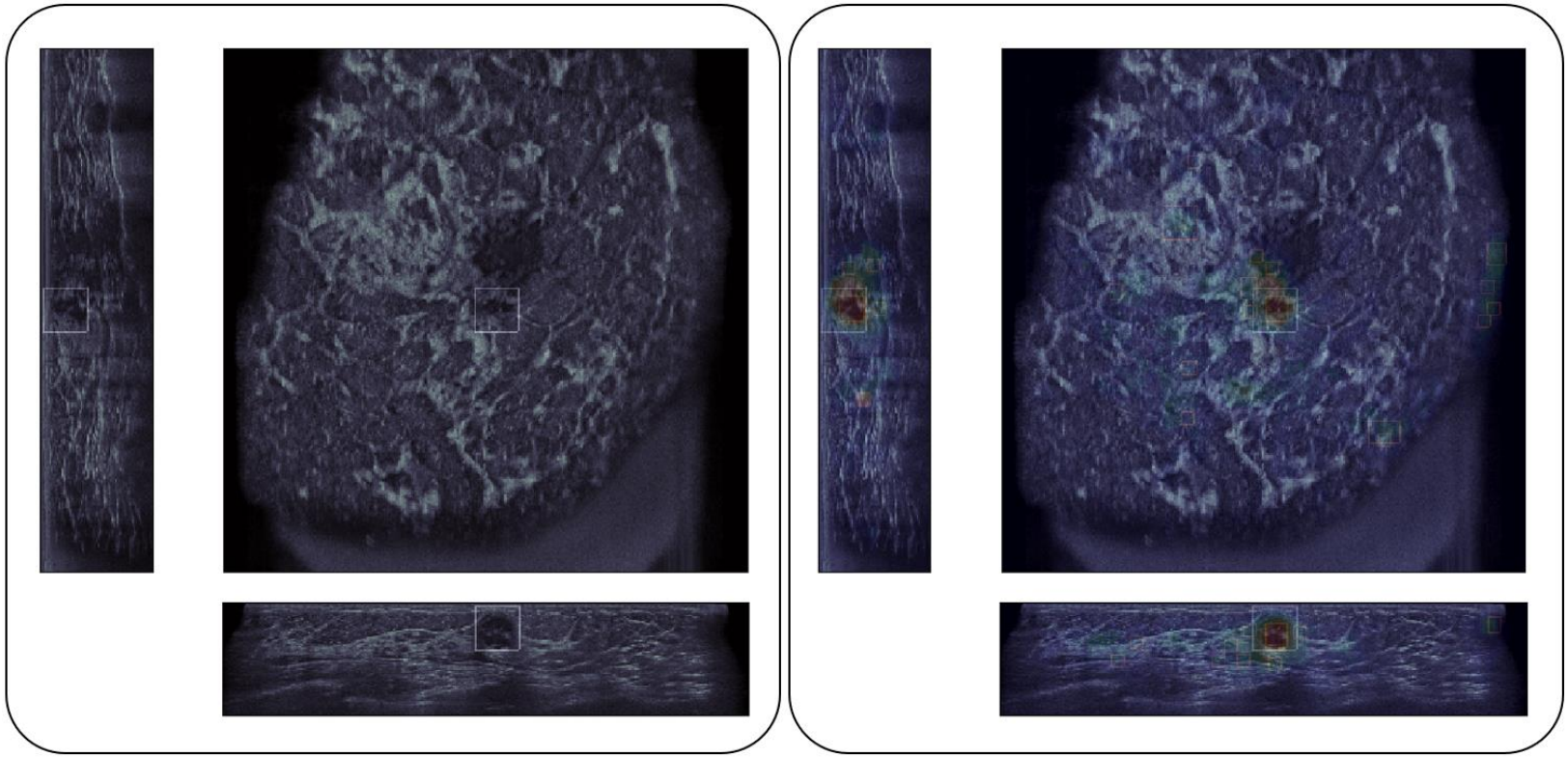


그림 20 유방병변 후보 검출기의 결과 예시2.(왼쪽) 자동 유방초음파 영상의 유방병변이 존재하는 위치의 단면으로 하얀색으로 표시된 부분이 유방병변의 위치 (오른쪽) 유방병변 후보 검출기의 결과로 주황색으로 표시된 부분이 검출결과임

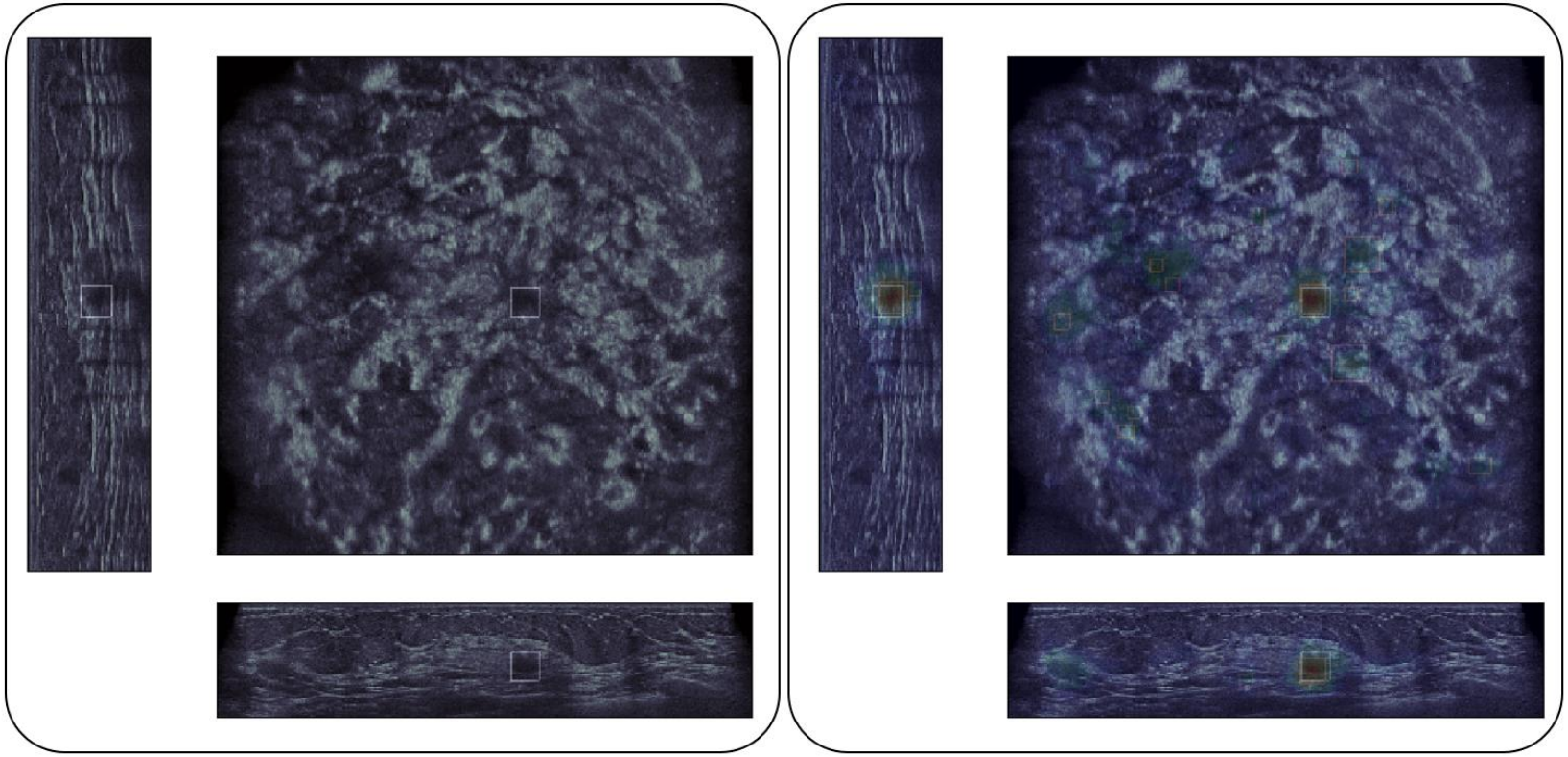


그림 21 유방병변 후보 검출기의 결과 예시3. (왼쪽) 자동 유방초음파 영상의 유방병변이 존재하는 위치의 단면으로 하얀색으로 표시된 부분이 유방병변의 위치 (오른쪽) 유방병변 후보 검출기의 결과로 주황색으로 표시된 부분이 검출결과

3.2 유방병변 유무 분류기 성능

3.2.1. 알고리즘 구조에 따른 성능

본 연구에서는 최종적으로 유방병변 유무분류기에 적용된 DenseNet-BC[25] 뿐만 아니라 ResNetV2-101[31]에도 제안된 방법을 적용하여 평가하였다. 표 2는 DenseNet에 각 버전의 마스크 네트워크를 적용했을 때 민감도, 특이도 및 AUC 실험 결과를 요약한다. 마스크 네트워크를 적용한 모든 케이스들은 기존과 비교하여 전반적으로 더 높은 성능을 달성하였다. 최상의 성능을 보이는 마스크 네트워크의 구조는 모델에 따라 다르다. DenseNet의 경우 MBN-V2가 모든 지표에 대해 최고의 성능을 달성했다. 표 3에는 ResNet을 사용한 방법의 결과가 성능 비교를 위해 표시된다. DenseNet의 경우와 유사하게, 제안된 모든 방법은 일반 Resnet 모델을 능가한다. DenseNet을 기반으로 한 실험과 달리, 마스크 네트워크는 민감도에 MBN-V2를, 특이도와 AUC에 MBN-V3를 사용하여 최고의 성능을 달성했다. 그림 22는 유방병변 유무 분류기에 대한 ROC 곡선과 마스크 분기 네트워크를 비교한다.

표 2 DenseNet에 마스크 네트워크를 적용한 유방병변 유무 분류기의 성능 비교표

Model	Sensitivity	Specificity	AUC
DenseNet	56.25%	86.25%	0.8553
DenseNet + MBN-V1	75.00%	92.50%	0.9252
DenseNet + MBN-V2	87.75%	93.75%	0.9491
DenseNet + MBN-V3	82.50%	92.50%	0.9472
DenseNet + MBN-V4	77.50%	92.50%	0.9483

표 3 ResNet에 마스크 네트워크를 적용한 유방병변 유무 분류기의 성능

비교표

Model	Sensitivity	Specificity	AUC
ResNet	65.00%	70.00%	0.7675
ResNet + MBN-V1	68.75%	77.50%	0.8270
ResNet + MBN-V2	71.25%	86.25%	0.8795
ResNet + MBN-V3	70.00%	91.25%	0.9261
ResNet + MBN-V4	67.50%	90.00%	0.8797

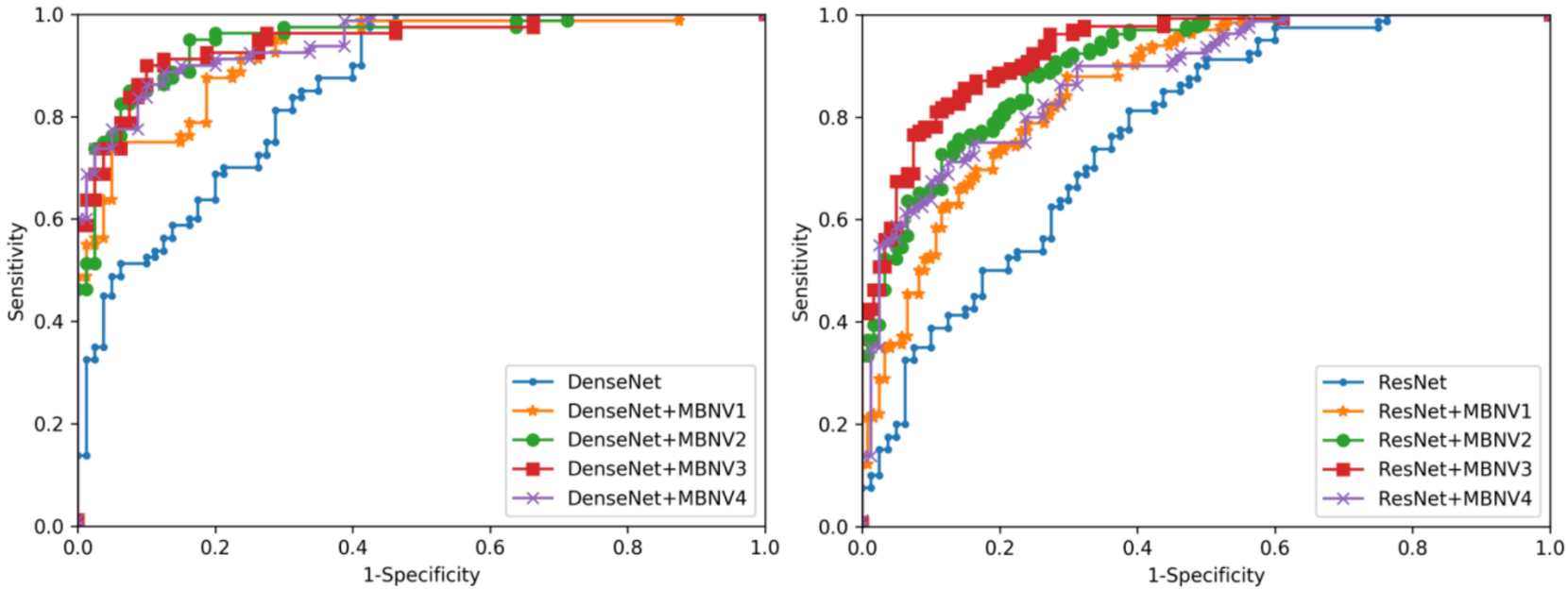


그림 22 마스크 네트워크를 적용한 DenseNet과 ResNet 기반의 유방병변 유무 분류기 ROC 그래프. 네트워크 구조마다 성능이 다르지만 전체적인 성능을 보았을 때에 기존 대비 높은 성능을 보임.

3.2.2. 손실 통합 기법에 따른 성능

본 연구에서는 손실을 통합하는 방법에 따른 결과를 비교하였다. 멀티태스킹 학습(Multi Task Learning)에서 다중 손실을 효율적으로 결합하기 위해 다양한 방법이 시도되었다[32]. 가장 간단한 방법은 균일한 가중치이다. 다중 손실은 단순히 하나의 스칼라 손실 값을 생성하기 위해 합산된다. 다중 손실을 통해 모델을 최적화하는 과정에서 그래디언트가 충돌하면 모델 성능이 저하될 수 있기 때문에 이런 충돌을 최소화하기 위해 유동적으로 손실들 간의 균형을 조절할 수 있는 동적인 최적화 기법이 중요하다. 켄달은 불확실성 기반 가중치 접근법을 제안하고 CNN에 적용했다. [28]. 또 다른 접근법은 켄달의 불확실성 기반 방법에 정규화하는 계산을 추가하였다. 그리고 불확실성에 기반하지 않은 가중치 조절 기법인 동적 중량 평균(DWA)도 제안되었다.[34] 본 연구에서는 이러한 방법을 MBN-V1과 동일한 데이터 세트를 사용하여 제안된 방법에 맞게 조정하고 성능을 비교했다. 우리는 동일한 가중치, 불확실성 기반 가중합 방식, 수정된 불확실성 가중합 방식 및 DWA를 마스크 손실에 적응시켰다. 결과는 표 4에 나타나 있으며, 다른 설정에 비해 불확실성 가중합 방식과 수정된 불확실성 가중합 방식이 더 높은 성능을 달성함을 보여준다. 결과는 불확실성 가중합 방식과 수정된 불확실성 가중합 방식 사이의 많은 차이를 보여주지는 않지만, 다른 방식들보다 상당히 우수함을 보였다. 이 결과를 고려하여, 우리는 최종 결과를 위해 우리의 방법에 불확실성 가중합 방식을 적용했다.

표 4 다중 손실 가중 방식에 따른 성능 비교표.

Model	Left side area	Right side area	AUC
Equal	65.00%	82.50%	0.8720
Uncertainty Weighting	75.00%	92.50%	0.9252
Revised Uncertainty Weighting	71.25%	93.75%	0.9369
Dynamic Weight Average	70.00%	77.50%	0.8416

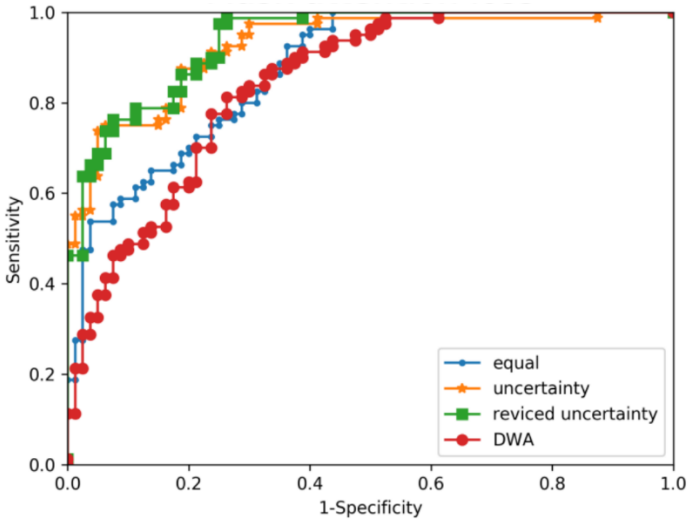


그림 23 다중 손실 가중방식에 따른 ROC 그래프 비교. 불확실성에 기반한 손실 통합 방식이 가장 높은 성능을 보임.

3.2.3. 적용된 유방병변 특성에 따른 성능

본 연구에서는 다양한 특성을 활용한 템플릿 마스크의 성능을 비교했다. 우리는 마스크 레이블을 만드는 데 쉽게 활용할 수 있는 몇 가지 특성을 제안한다. 템플릿 마스크의 변화는 사용된 특성에 따라 TM-V1에서 TM-V5로 분류되었다. 표 5는 유방병변과 다른 특성 세트를 사용하는 사례와 그 결과를 보여줍니다. TM-V1은 위치만 고려하며 그 모양은 직경 20mm의 구에 고정된다. 병변의 유일한 위치를 고려하는 TM-V1에서도 일반

모델에 비해 성능이 향상되었다. TM-V2는 구 모양의 병변 위치와 직경을 모두 고려하여 마스크 라벨의 양의 영역을 설정하여 TM-V1보다 높은 감도를 산출한다. TM-V2에 더해, TM-V3는 유방병변의 orientation 특성을 활용하여 마스크의 양의 영역의 모양을 구 또는 타원체로 설정한다. 추가적인 특징 중 후향 에코 패턴의 경우 병변 아래 영역의 특성에 따라 'no posterior feature', 'enhancement', 'shadowing', 'combined echo pattern'으로 구분된다. TM-V4에서는 양성 영역을 피부 표면 반대 방향으로 반직경의 길이만큼 길게 확장하여 후방 에코 패턴을 활용할 수도 있다. 가장 성능이 좋은 템플릿 마스크는 방향 또는 후방 에코 패턴이 있는 마스크이지만, 위치와 직경만 있는 TM-V2 또한 그림 24에서 보는 바와 같이 성능향상을 보였다. 따라서 방향 또는 후방 음향 기능과 같은 추가 정보를 사용할 수 없는 경우에는 병변의 중앙점과 크기만으로 마스크 레이블을 만들 수 있다.

표 5 템플릿 마스크에 활용된 유방병변의 특징에 따른 성능비교

Model	Left side area	Right side area	AUC
None	56.25%	86.25%	0.8553
TM-V1(only pos)	67.50%	90.00%	0.9278
TM-V2(pos+ dia)	72.50%	90.00%	0.9214
TM-V3(pos+ dia+ ori)	75.00%	92.5%	0.9252
TM-V4(pos+ dia+ eco)	70.00%	90.00%	0.9344
TM-V5(pos+ dia+ ori+ eco)	76.25%	88.75%	0.9267

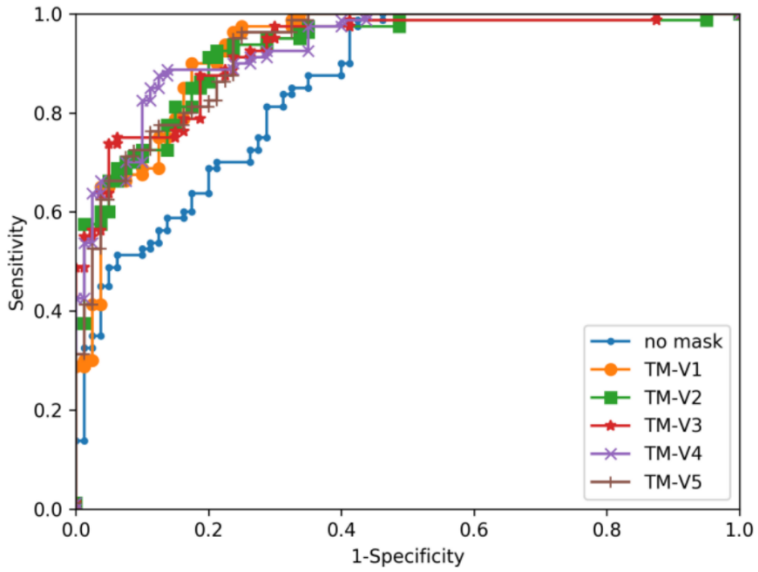


그림 24 템플릿 마스크에 활용된 유방병변의 특징에 따른 ROC 그래프 비교

3.2.4. 시각적 평가

그림 25는 각 모델의 grad-CAM을 표시하여 마스크 손실의 효과를 정성적으로 시각화했다. Grad-CAM는 마지막 합성곱 계층의 출력에 대해 계산된다. 각 행은 유방병변이 있는 입력 이미지, DenseNet의 grad-CAM, 마스크 네트워크가 있는 DenseNet의 grad-CAM을 의미한다. 두 번째 행에서 활성화된 영역이 잘못된 위치에 있거나 뚜렷한 활성화된 영역 없이 넓게 퍼져 있다. 이는 네트워크가 제대로 훈련되지 않았으며 어디에서 특성을 추출해야 할 지 잘 학습하지 못했다는 것을 의미한다고 할 수 있다. 반면에 번째 행에서는 템플릿 마스크와 마스크 네트워크가 적용되어 대상 개체 영역에 집중하여 학습된 것을 알 수 있다. 본 연구에서 제안한 템플릿 마스크와 마스크 네트워크를 적용한 딥러닝 모델은 더욱 효과적으로 학습하였다.

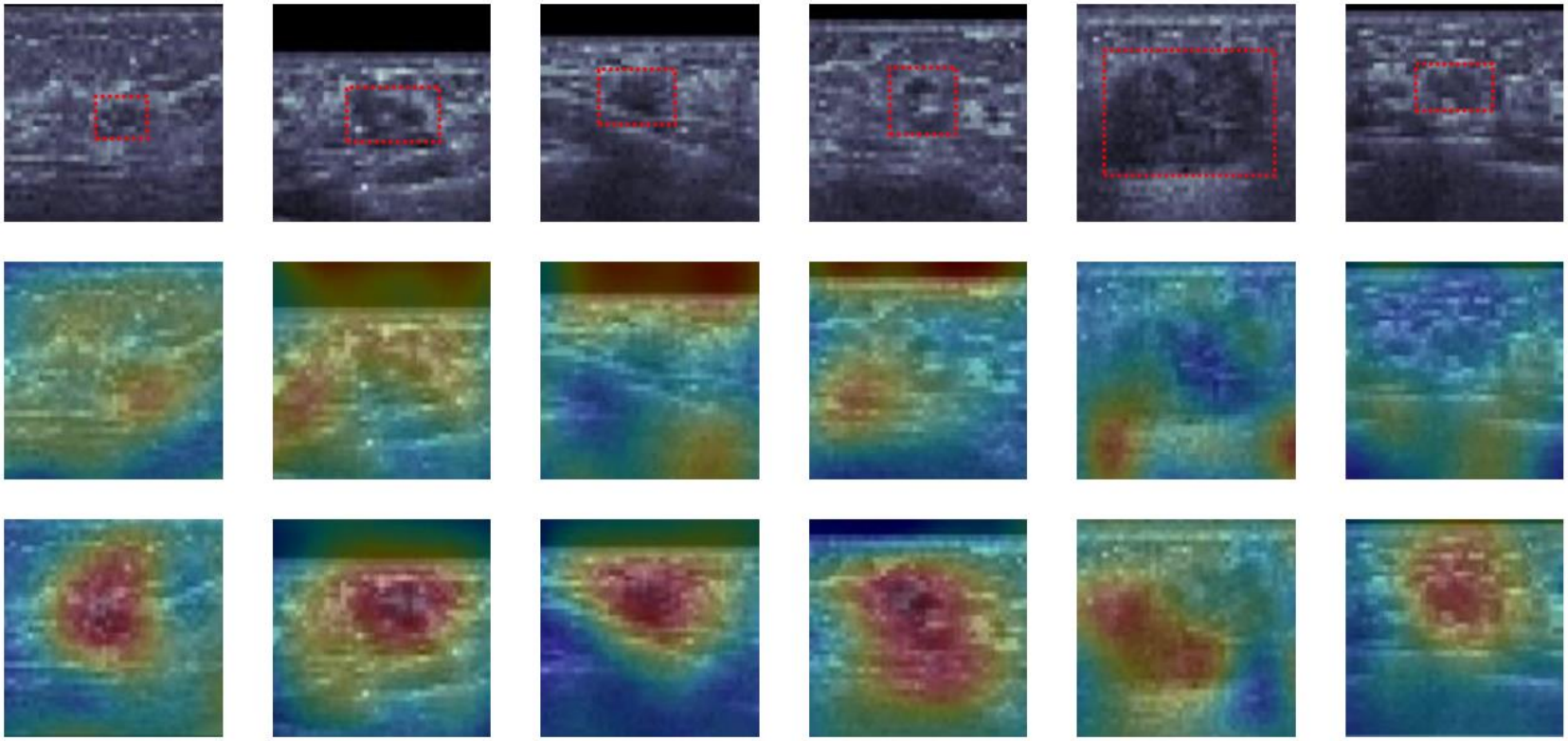


그림 25 템플릿 마스크와 마스크 네트워크 적용 전후 CAM. (행 1) 유방병변의 단면 (행 2) 기존 모델의 CAM (행 3) 템플릿 마스크와 마스크 네트워크가 적용된 모델의 CAM

3.3. 유방병변 악성도 분류기 성능

유방병변 악성도 분류기는 2.4.1 절에서 설명한 바와 같이 전처리를 하지 않고 알고리즘에 입력된다. 이때에 전처리를 하지 않은 크롭이미지의 사이즈가 커서 모델이 필요로 하는 메모리가 GPU 메모리로 수용하지 못할 정도로 커지기 때문에 x, y, z 축 세가지 중 한가지 방향으로 축소시킬 수 밖에 없었다. 따라서 입력되는 크롭이미지의 사이즈를 그림 22와 같이 96×288×72, 192×144×72, 192×288×36 세가지로 나누어 함께 성능을 평가하였다. 표 6은 전처리를 한경우와 하지 않은 세가지 경우에 대한 성능평가 결과를 나타낸다. 성능 평가에 사용된 테스트 셋은 악성병변 40 case 양성 병변 40 case로 모두 모델의 트레이닝 셋과 완전히 분리된 데이터셋이었다. AUC를 보았을 때에 전처리를 하지않고 본 연구에서 제안하는 알고리즘 구조를 쓴 경우가 전처리를 한 일반적인 경우보다 높은 성능을 보였다. 그 중에서 y 축 방향으로 크기를 줄인 경우가 가장 성능이 높았고 최종적으로 유방암 검출 시스템에 적용되었다.

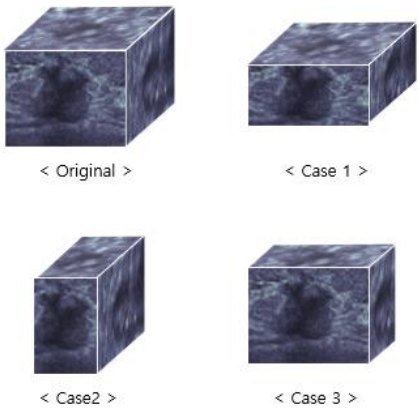


그림 26 유방병변 악성도 분류구에 입력되는 입력 이미지 후보군

표 6 유방병변 악성도 분류기의 성능

Case	Sensitivity	Specificity	AUC
Normal	91.67%	53.18%	0.858
x-axis slice	83.30%	**86.11%**	0.8982
y-axis slice	**93.33%**	75.00%	**0.9019**
z-axis slice	83.30%	66.67%	0.8590

3.4. 유방암 검출 시스템 정확도

표 7은 성능 비교 결과를 보여준다. 첫번째 행의 결과는 QVCAD의 매뉴얼상에 기록된 성능인데, FPR 은 매우 낮지만 검출률은 낮은 결과를 보였다. 두번째와 세번째 행의 결과가 동일한 테스트 셋으로 평가한 결과인데, 검출률의 경우 본연구에서 개발한 유방암 검출기는 0.876(95% 신뢰구간, 0.819-0.933)의 성능을 보였고 비교군의 0.845(95% 신뢰구간, 0.783-0.907)보다 약 0.03 더 높은 결과를 보였다. 위양성 검출률 0.442(95% 신뢰구간 0.356-0.528)의 경우 비교군의 0.45(95% 신뢰구간 0.364-0.536)보다 약 0.008 낮은 결과를 보였다.

표 7 비교군 CAD와 본 연구에서 제안한 유방암 검출 시스템의 성능비교

Case	Detection rate	FPR
QV CAD	0.71	0.12
QV CAD(same testset)	0.845	0.450
ours	**0.876**	**0.442**

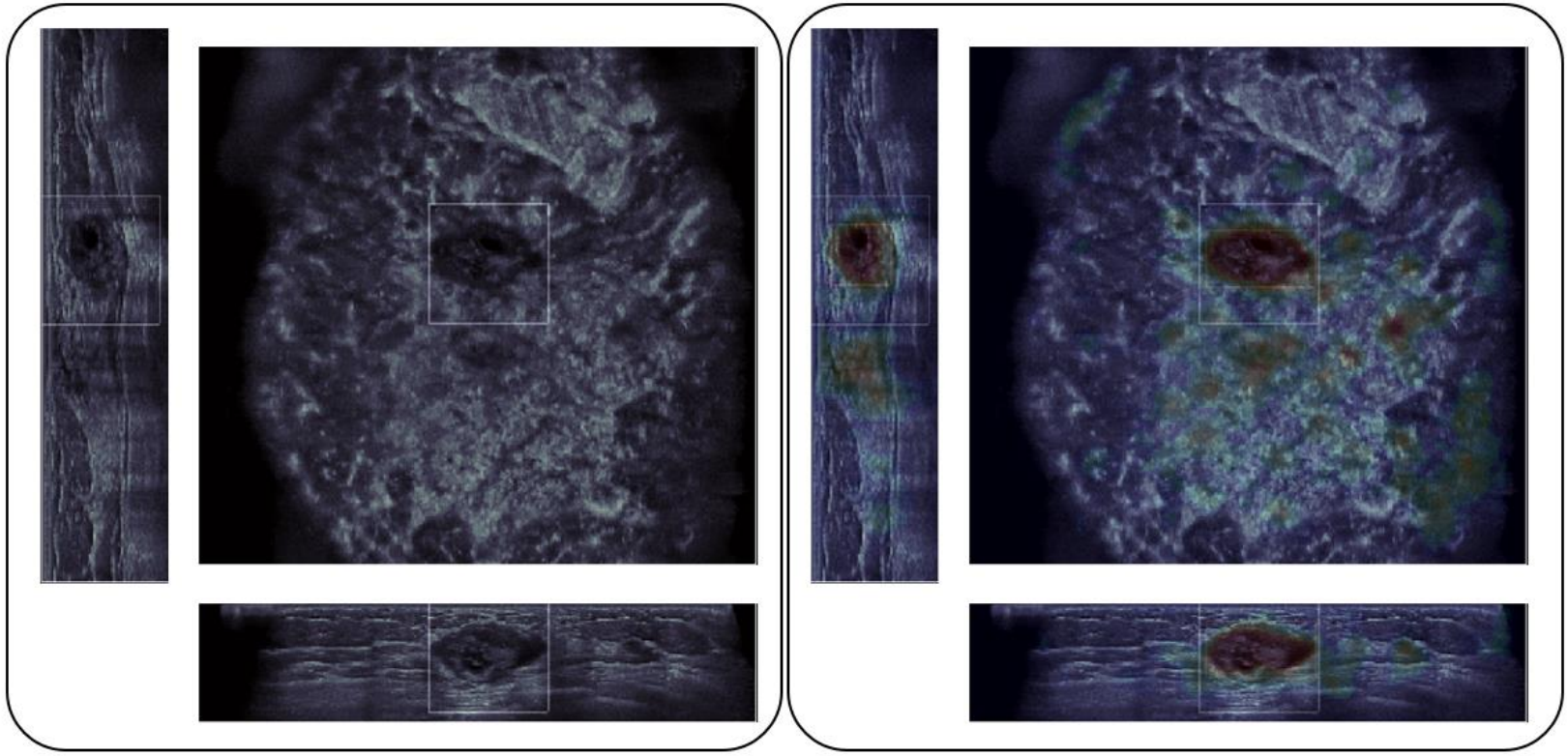


그림 27 본 연구에서 개발한 유방암 검출 시스템의 결과 예시1. (왼쪽) 자동 유방초음파 영상과 하얀색으로 표시된 유방암 (오른쪽) 유방암 검출시스템에서 활성화된 정도를 나타내는 CAM과 주황색으로 표시된 최종결과

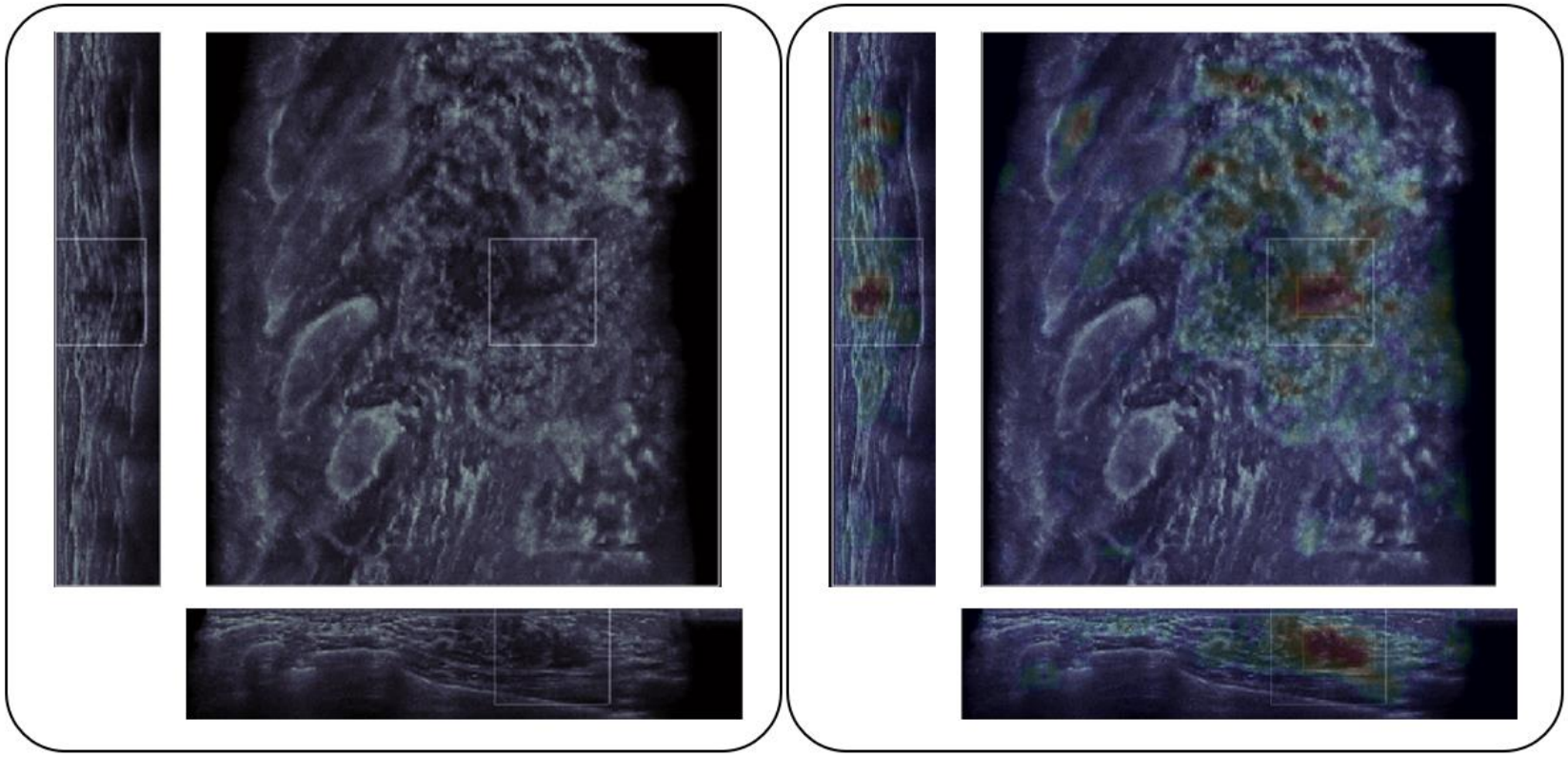


그림 28 본 연구에서 개발한 유방암 검출 시스템의 결과 예시2. (왼쪽) 자동 유방초음파 영상과 하얀색으로 표시된 유방암 (오른쪽) 유방암 검출시스템에서 활성화된 정도를 나타내는 CAM과 주황색으로 표시된 최종결과

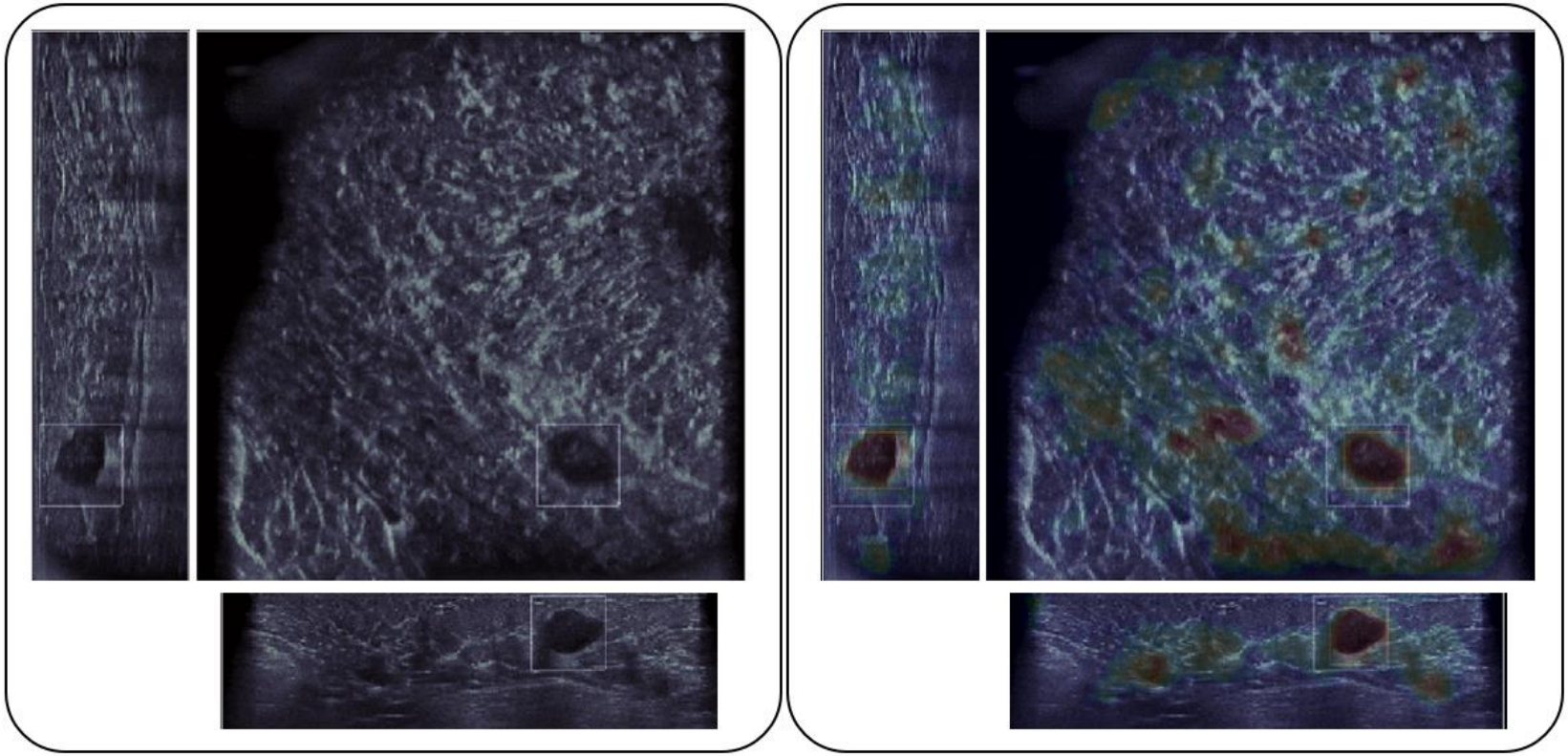


그림 29 그림 30 본 연구에서 개발한 유방암 검출 시스템의 결과 예시3. (왼쪽) 자동 유방초음파 영상과 하얀색으로 표시된 유방암 (오른쪽) 유방암 검출시스템에서 활성화된 정도를 나타내는 CAM과 주황색으로 표시된 최종결과

고찰

4.1. 실험 결과 고찰

자동 유방초음파 영상은 삼차원이라는 영상의 특성상 유용한 데이터 구축이 어렵고, 초음파 영상 특성상 영상의 품질이 비교적 떨어지고 병변의 특성이 뚜렷하지 않기 때문에 진단 보조시스템을 개발하기에 다른 의료영상에 비해 상대적으로 어려움이 있다고 할 수 있다. 따라서 높은 성능의 진단 보조 시스템을 개발하기 위해서는 효과적으로 데이터를 구축해야 하고, 유방병변의 특징을 효과적으로 학습할 수 있는 새로운 접근법이 필요하다. 본연구에서는 이러한 어려움을 극복하기 위해서 여러가지 기법을 제안하였고 실험을 통해서 그 효과를 검증하였다.

특히 유방병변 유무 판별기 알고리즘을 학습하는 과정에서 초음파영상의 한계를 극복하기 위하여 템플릿 마스크의 정보를 마스크 네트워크를 적용하여 한정된 데이터내에서 효과적으로 학습하는 방법을 제안하였다. 이 방법을 적용하였을 때에 성능향상을 확인하였고, 제안한 방법은 알고리즘의 학습 과정에서만 필요하므로 실제로 시스템을 사용하는 데에는 추가적인 메모리나 시간이 필요하지 않기 때문에 더욱 의미가 있다고 할 수 있다.

전체 유방암 검출시스템의 성능평가결과를 종합적으로 판단해 보았을 때에 비교군에 비해서 높은 성능을 보였고, 검출률 향상 관점에서 성과를 이룩하였고, 컴퓨터 보조 발견시스템은 민감도에 더욱 중점을 두고 개발한다는 점을 고려하였을 때에 본 연구의 유방암 검출시스템은 진단 보조시스템으로서 잠재적인

효과를 가지고 있다고 볼 수 있다.

4.2. 발전 방향

본 연구에서 제안한 유방암 검출 시스템은 비교군과 비교해 보았을 때에 높은 성능을 보여주었지만, 위양성의 관점에서 더욱 보완해야 할 부분이 있다. 현재 위양성률은 0.442이고 일반적으로 한 스터디당 자동 유방초음파 볼륨을 6개씩 스캔하므로 스터디 단위로는 2개이상의 위양성이 발생한다고 볼 수 있다. 위양성 검출은 판독시간을 지연시킬 수 있는 요소이므로 추가 연구 진행시 위양성의 문제 보완에 집중해야 할 것이다.

본 연구에서 유방암 검출 시스템의 성능을 평가할 때에는 유방암 검출시스템만으로 진단하는 조건(Standalone)으로 성능을 평가하였지만, 일반적으로 진단 보조시스템은 전문의의 판독과 병행하여 진행되므로, 실용적인 관점에서 효과성을 확인하기 위해서는 판독의와 병행하였을 때에 판독시간과 판독결과를 확인하는 임상시험이 진행될 필요가 있다.

결론

본 연구에서는 자동 유방초음파 영상에서 유방암을 검출하는 역할을 하는 딥러닝 알고리즘을 기반으로 한 진단 보조 시스템을 개발하였다. 유방암 검출 시스템은 '유방병변 후보 검출기', '유방병변 유무 판별기', '유방병변 악성도 분류기'의 세단계로 이루어져 있으며 입력된 삼차원 자동 유방초음파 영상에서 유방병변의 위치를 검출하고 유방병변의 악성도를 판단하여 유방암을 검출하는 기능을 한다.

이런 유방암 검출 시스템을 개발하는 과정에서 딥러닝 알고리즘 학습을 위한 데이터셋 구축을 필요로 하는데, 삼차원 의료영상의 특성상 라벨링 작업은 시간적 인적 낭비를 발생한다. 이런 장애물을 극복하기 위하여 추가적인 라벨링 작업 없이 기존의 정보를 최대한 활용하는 효율적인 데이터 수집 방법을 사용하였다.

진단보조 시스템에 사용된 딥러닝 알고리즘은 기존에 많은 연구를 통해 검증된 알고리즘을 기초로 하여 템플릿 마스크와 마스크 분기 네트워크 등의 다양한 기법을 적용하였다. 다양한 시도를 통해 유방병변에 대한 특징을 효과적으로 학습할 수 있도록 하여 시스템의 성능을 높일 수 있었다.

시스템의 각각의 모듈의 성능을 평가하여 시스템 개발에 사용된 여러가지 기법의 효과성을 확인하였고, 시스템 전체의 성능을 기존의 CAD와 비교 평가하여 본 연구에서 개발된 시스템의 우수성을 검증하였다. 보완해야 할 위양성의 단점이 남아있고 실제 판독에서 얼마나 효과가 있을지는 추가 임상시험을 통해서 검증해보아야 하겠지만 본 연구에서 보여준 결과만으로 잠재성은

확인할 수 있었다.

향후 사용성이 확대될 가능성이 높은 자동 유방초음파 영상에서 진단 보조 시스템에 대한 필요성은 점차 증가할 것으로 보며, 딥러닝을 활용한 유방암 검출 시스템이 자동 유방초음파 영상 판독에 적용될 경우 의료 서비스 향상 및 진단 프로세스 효율성 증대에 기여할 수 있을 것이다.

참고 문헌

[1] Youlden DR, Cramb SM, Yip CH, Baade PD. Incidence and mortality of female breast cancer in the Asia-Pacific region. Cancer Biol Med 2014;11:101-115

[2] A. B. Miller, C. Wall, C. J. Baines, P. Sun, T. To, and S. A. Narod,"Twenty five year follow-up for breast cancer incidence and mortalityof the canadian national breast screening study: randomised screeningtrial,"Bmj, vol. 348, 2014

[3] D. Saslow, C. Boetes, W. Burke, S. Harms, M. O. Leach, C. D.Lehman, E. Morris, E. Pisano, M. Schnall, S. Seneret al., "Americancancer society guidelines for breast screening with mri as an adjunct tomammography,"CA: a cancer journal for clinicians, vol. 57, no. 2, pp.75–89, 2007

[4] A. C. on Breast Cancer Screening, "Screening for breast cancer inengland: past and future,"Journal of medical screening, vol. 13, no. 2,pp. 59–61, 200

[5] Boyd, Norman F., et al. "Mammographic density and the risk and detection of breast cancer." New England journal of medicine 356.3 (2007): 227-236.

[6] Kolb, Thomas M., Jacob Lichy, and Jeffrey H. Newhouse. "Comparison of the performance of screening mammography, physical examination, and breast US and evaluation of factors that influence them: an analysis of 27,825 patient evaluations." Radiology 225.1 (2002): 165-175.

[7] Nazari, Shayan Shaghayeq, and Pinku Mukherjee. "An overview of mammographic density and its association with breast cancer." Breast cancer 25.3 (2018): 259-267.

[8] Berg, Wendie A., et al. "Lesion detection and characterization in a breast US phantom: results of the ACRIN 6666 Investigators." Radiology 239.3 (2006): 693-702.

[9] Nothacker, Monika, et al. "Early detection of breast cancer: benefits and risks of supplemental breast ultrasound in asymptomatic women with mammographically dense breast tissue. A systematic review." BMC cancer 9.1 (2009): 1-9.

[10] A. Vourtsis and A. Kachulis, "The performance of 3d abus versus hhusin the visualisation and bi-rads characterisation of breast lesions in alarge cohort of 1,886 women,"European radiology, vol. 28, no. 2, pp.592–601, 2018

[11] Rella, Rossella, et al. "Automated breast ultrasonography (ABUS)

in the screening and diagnostic setting: indications and practical use." Academic Radiology 25.11 (2018): 1457-1470.

[12] S. Yang, X. Gao, L. Liu, R. Shu, J. Yan, G. Zhang, Y. Xiao, Y. Ju,N. Zhao, and H. Song, "Performance and reading time of automatedbreast us with or without computer-aided detection,"Radiology, vol.292, no. 3, pp. 540–549, 201

[13] Y. Jiang, M. F. Inciardi, A. V. Edwards, and J. Papaioannou, "Inter-pretation time using a concurrent-read computer-aided detection systemfor automated breast ultrasound in breast cancer screening of womenwith dense breast tissue,"American Journal of Roentgenology, vol. 211,no. 2, pp. 452–461, 2018

[14] J. C. van Zelst, T. Tan, P. Clauser, A. Domingo, M. D. Dorrius,D. Drieling, M. Golatta, F. Gras, M. de Jong, R. Pijnappelet al.,"Dedicated computer-aided detection software for automated 3d breastultrasound; an efficient tool for the radiologist in supplemental screeningof women with dense breasts,"European radiology, vol. 28, no. 7, pp.2996–3006, 2018

[15] P. Crystal, S. D. Strano, S. Shcharynski, and M. J. Koretz, "Usingsonography to screen women with mammographically dense breasts,"American Journal of Roentgenology, vol. 181, no. 1, pp. 177–182, 20

[16] K. M. Kelly, J. Dean, W. S. Comulada, and S.-J. Lee, "Breast cancerdetection using automated whole breast ultrasound and mammographyin radiographically dense breasts,"European radiology, vol. 20, no. 3,pp. 734–742, 201.

[17] X. Shi, H. Cheng, and L. Hu, "Mass detection and classification inbreast ultrasound images using fuzzy svm," in9th Joint InternationalConference on Information Sciences (JCIS-06). Atlantis Press, 200.

[18] S. Murali and M. Dinesh, "Classification of mass in breast ultrasoundimages using image processing techniques," 20

[19] C.-M. Lo, R.-T. Chen, Y.-C. Chang, Y.-W. Yang, M.-J. Hung, C.-S. Huang, and R.-F. Chang, "Multi-dimensional tumor detection inautomated whole breast ultrasound using topographic watershed,"IEEEtransactions on medical imaging, vol. 33, no. 7, pp. 1503–1511, 201

[20] C. Ye, V. Vaidya, and F. Zhao, "Improved mass detection in 3dautomated breast ultrasound using region based features and multi-viewinformation," in2014 36th Annual International Conference of the IEEEEngineering in Medicine and Biology Society. IEEE, 2014, pp. 2865–2868

[21] T.-C. Chiang, Y.-S. Huang, R.-T. Chen, C.-S. Huang, and R.-F. Chang,“Tumor detection in automated breast ultrasound using 3-d cnn and pri-oritized candidate aggregation,”IEEE transactions on medical imaging,vol. 38, no. 1, pp. 240–249, 201
[22] W. K. Moon, Y.-S. Huang, C.-H. Hsu, T.-Y. C. Chien, J. M. Chang, S. H.Lee, C.-S. Huang, and R.-F. Chang, “Computer-aided tumor detection inautomated breast ultrasound using a 3-d convolutional neural network,”Computer Methods and Programs in Biomedicine, vol. 190, p. 105360,2020
[23] 박영미. "표준화된 판독지의 작성." 대한유방검진의학회지 3.1 (2006): 80-83.
[24] Lazarus, Elizabeth, et al. "BI-RADS lexicon for US and mammography: interobserver variability and positive predictive value." Radiology 239.2 (2006): 385-391.
[25] Wang, Chengjia, et al. "A two-stage 3D Unet framework for multi-class segmentation on full resolution image." arXiv preprint arXiv:1804.04341 (2018).
[26] Baid, Ujjwal, et al. "Deep learning radiomics algorithm for gliomas (drag) model: a novel approach using 3d unet based deep convolutional neural network for predicting survival in gliomas." International MICCAI Brainlesion Workshop. Springer, Cham, 2018.
[27] Huang, Gao, et al. "Densely connected convolutional networks." Proceedings of the IEEE conference on computer vision and pattern recognition. 2017.
[28] S. Teerapittayanon, B. McDanel, and H.-T. Kung, “Branchynet: Fastinference via early exiting from deep neural networks,” in2016 23rdInternational Conference on Pattern Recognition (ICPR). IEEE, 2016,pp. 2464–2469
[29] H. Fukui, T. Hirakawa, T. Yamashita, and H. Fujiyoshi, “Attentionbranch network: Learning of attention mechanism for visual explana-tion,” inProceedings of the IEEE/CVF Conference on Computer Visionand Pattern Recognition, 2019, pp. 10 705–10 714.
[30] A. Kendall, Y. Gal, and R. Cipolla, “Multi-task learning using uncer-tainty to weigh losses for scene geometry and semantics,” inProceedingsof the IEEE conference on computer vision and pattern recognition,2018, pp. 7482–7491
[31] X. Glorot and Y. Bengio, “Understanding the difficulty of trainingdeep feedforward neural networks,” inProceedings of the thirteenthinternational conference on artificial intelligence and statistics, 2010,pp. 249–256.
[32] S. Ioffe and C. Szegedy, “Batch normalization: Accelerating deepnetwork training by reducing internal covariate shift,”arXiv preprintarXiv:1502.03167, 2015.

[33] K. He, X. Zhang, S. Ren, and J. Sun, “Identity mappings in deep residual networks,” inEuropean conference on computer vision. Springer, 2016,pp. 630–645

[34] . Gong, T. Lee, C. Stephenson, V. Renduchintala, S. Padhy, A. Ndi-rango, G. Keskin, and O. H. Elibol, “A comparison of loss weightingstrategies for multi task learning in deep neural networks,”IEEE Access,vol. 7, pp. 141 627–141 632, 2019.

[35] L. Liebel and M. K ¨orner, “Auxiliary tasks in multi-task learning,”arXivpreprint arXiv:1805.06334, 2018.

[36] S. Liu, E. Johns, and A. J. Davison, “End-to-end multi-task learningwith attention,” inProceedings of the IEEE Conference on ComputerVision and Pattern Recognition, 2019, pp. 1871–1880.

[37] Kim, Youngjune, et al. "False-negative results on computer-aided detection software in preoperative automated breast ultrasonography of breast cancer patients." Ultrasonography 40.1 (2021): 83.

Abstract

A Study on the Deep-learning based Computer-aided Detection system for Breast cancer in the Automated Breast Ultrasound

Daekyung Kim

Interdisciplinary Program in Bioengineering

The Graduate School

Seoul National University

The goal of this study is to develop computer aided detection system for breast cancer in three-dimensional automated breast ultrasound images and evaluate the performance. Automated Breast Ultrasound(ABUS), which has various advantages over mammography or handheld breast ultrasound(HHUS) requires relatively more time and effort to read by radiologists due to the image characteristics of ultrasound and the three-dimensional problem. Recent studies have employed computer vision technology and deep learning algorithms to develop computer aided diagnosis systems, and if it is applied to

automatic breast ultrasound images, it will contribute greatly to the efficiency and quality of the medical service.

Breast cancer detection systems developed in this study consist of three modules, each based on a three-dimensional convolutional neural network that has shown high effectiveness in image analysis through a number of recent studies. The breast cancer detection system first scans the whole automated breast ultrasound image through the "region proposal module for breast mass" to estimate the location of suspicious lesions. To determine whether breast mass is actually present at the estimated location, we classify the area where the actual tumor exists with a ‘breast mass classifier’. Results classified as breast mass are classified as benign and malignant through a "breast malignancy classifier" that predicts the malignancy of breast masses, resulting in lesions suspected of malignant breast cancer except benign tumors.

To train the deep learning algorithm used in this study, we built datasets in an efficient way without additional labeling work from automated breast ultrasound images. To effectively train deep learning algorithms, we utilized weakly supervised learning methods and proposed neural network architectures, which showed the performance improvement.

In this study, experiments were proceeded to verify the performance of both each modules and the overall system that. Single module evaluation showed higher performance compared to conventional methods, especially template masks and mask

networks used in breast mass classifier showed significant improvements. The systems developed in this study showed higher performance compared to other existing systems with a detection rate of 84.4% (95% confidence interval, 75.5%-93.3%), and a false positive detection rate of 0.8 (95% confidence interval 0.6-1.0).

In this study, breast cancer detection system was developed using self-built datasets and algorithms, and showed high performance. Thus, breast cancer detection systems utilizing deep learning has the potential to improve the diagnostic performance of automated breast ultrasound.

Keywords: Deep learning algorithm, Breast cancer, Automated breast ultrasound, Computer aided Detection system

Student Number: 2019-23660