

결절성 폐암 검출을 위한 상용 및 맞춤형 CNN의 성능 비교

박성욱[†], 김승현^{**}, 임수창^{***}, 김도연^{****}

Performance Comparison of Commercial and Customized CNN for Detection in Nodular Lung Cancer

Sung-Wook Park[†], Seunghyun Kim^{**}, Su-Chang Lim^{***}, Do-Yeon Kim^{****}

ABSTRACT

Screening with low-dose spiral computed tomography (LDCT) has been shown to reduce lung cancer mortality by about 20% when compared to standard chest radiography. One of the problems arising from screening programs is that large amounts of CT image data must be interpreted by radiologists. To solve this problem, automated detection of pulmonary nodules is necessary; however, this is a challenging task because of the high number of false positive results. Here we demonstrate detection of pulmonary nodules using six off-the-shelf convolutional neural network (CNN) models after modification of the input/output layers and end-to-end training based on publicly databases for comparative evaluation. We used the well-known CNN models, LeNet-5, VGG-16, GoogLeNet Inception V3, ResNet-152, DensNet-201, and NASNet. Most of the CNN models provided superior results to those of obtained using customized CNN models. It is more desirable to modify the proven off-the-shelf network model than to customize the network model to detect the pulmonary nodules.

Key words: Pulmonary Nodule, Computer Aided Detection, Deep Neural Network, Convolutional Neural Network

1. 서 론

암으로 인한 사망률은 전 세계에서 폐암(Lung Cancer)이 가장 높다. 미국에서 폐암은 암으로 인한 사망의 약 26%를 차지한다[1]. 2011년 미국의 국가 폐암검진연구(National Lung Screening Trial)에서는 방사선(Radiation)의 양이 적은, 저선량(Low Dose) 컴퓨터단층촬영(Computed Tomography, CT)

그룹에서 폐암으로 인한 사망률이 20% 감소되는 것을 관찰했다[2]. 이 연구결과로 인해, 현재 미국에서는 폐암의 조기 발견을 위해 저선량 CT를 꾸준히 사용하고 있다. 그러나 이런 검진 프로그램의 실행 결과인 CT 영상의 양은 막대하다[3]. 따라서 방사선 전문의가 분석하는데 무리가 따른다. 이런 문제를 완화하기 위해 컴퓨터 보조 탐지(Computer Aided Detection, CAD) 시스템이 개발됐다. CAD 시스템은 방

※ Corresponding Author : Do-Yeon Kim, Address: (57922) Jungang-ro 255, Suncheon, Jeonnam, Korea, TEL : +82-61-750-3628, FAX : +82-61-750-3620, E-mail : dykim@sunchon.ac.kr

Receipt date : Jun. 25, 2019, Revision date : Apr. 30, 2020
Approval date : May. 22, 2020

[†] Dept. of Computer Engineering, Suncheon National University (E-mail : 411050@sunchon.ac.kr)

^{**} Dept. of Computer Engineering, Suncheon National University E-mail : kimtmdgus22@gmail.com)

^{***} Dept. of Computer Engineering, Suncheon National University E-mail : suchangLim@sunchon.ac.kr)

^{****} Dept. of Computer Engineering, Suncheon National University

사선 전문의들의 영상 판독을 용이하게 하고, 환자의 폐암 검진을 효율적으로 할 수 있다[4-7]. CAD 시스템은 두 단계로 나뉜다. 첫 번째 단계에서 결절(Nodule) 후보는 높은 민감도(Sensitivity)로 탐지되는데 이는 거짓 양성(False Positive, FP)이 그만큼 많기 때문이다. 두 번째 단계에서 FP 수는 이전보다 감소한다[4]. FP 수가 얼마만큼 감소했느냐에 따라 해당 CAD 시스템의 성능이 결정된다. 최근까지 많은 CAD 시스템이 제안됐지만 공개 데이터 셋으로 객관적인 비교 평가를 제공하는 연구 결과는 찾아보기 힘들다[3, 8, 9].

딥러닝(Deep Learning)은 머신 러닝(Machine Learning)의 한 분야로 연속된 층에서 점진적으로 의미 있는 표현을 배우는데 강점이 있으며 데이터로부터 표현을 학습하는 알고리즘이다[10]. 최근 몇 년 동안, 딥러닝은 대부분의 영상 처리(Image Processing) 및 컴퓨터 비전(Computer Vision) 문제를 해결할 수 있는 선도적인 머신 러닝 방법으로 부상했다. 특히, 학습된 패턴이 평행 이동 불변성(Translation Invariant)을 가지고, 패턴의 공간적 계층 구조를 학습할 수 있는 컨벌루션 신경망(Convolutional Neural Network, CNN)은 대부분의 컴퓨터 비전 애플리케이션(Application)에 사용된다. 심층 CNN은 원시 데이터(Raw Data)에서 획득한 중간 수준 및 고 수준의 추상화(Abstraction)를 스스로 학습한다[11]. 의료 영상 커뮤니티에서 심층 CNN은 병변 분류 및 탐지(Lesion Classification and Detection)는 물론 장기 분할 및 등록(Organ Segmentation and Registration)을 위한 방법으로 사용하고 있다[12]. 본 논문에서는 공개 데이터 셋을 사용하여 후보자의 폐 결절을 탐지하는 데 CNN이 효율적으로 활용될 수 있음을 보인다. 검증된 기성 CNN 모델이 폐 결절 탐지를 위해 설계된 맞춤형 CNN 모델보다 더 나은 결과를 제공한다는 점도 입증한다. 여기서 깊이가 얇은 CNN 모델은 훈련 시간 및 메모리(Memory) 크기로 인해 모바일 장치 활용 측면에서 큰 이점을 가진다.

2. 기성 CNN의 구조 및 알고리즘

본 논문에서는 여섯 가지 CNN 모델(LeNet-5, VGG-16, Inception-V3, ResNet-152, DenseNet-201, NASNet)을 사용하여 실험을 진행했다.

LeNet-5의 구조는 총 3개의 컨벌루션, 2개의 서브

샘플링(Subsampling), 1개의 완전연결(Fully-connected) 층으로 구성된다[13]. ‘C’는 컨벌루션을 의미한다. C1, C3, C5 층은 28×28 , 10×10 , 1×1 픽셀의 6, 16, 120개 특징맵(Feature Map)을 가진다. ‘S’는 서브샘플링을 의미한다. S2, S4 층은 14×14 , 5×5 픽셀의 6, 16개 특징맵을 가진다. ‘F’는 완전연결을 의미한다. F6 층은 C5 층의 결과를 84개 유닛(Unit)에 연결시킨다. LeNet-5 분류 층의 출력 클래스 수는 ‘10’이다.

VGG 아키텍처의 모든 층은 스트라이드(Stride)와 패딩(Padding) 크기가 1인 3×3 크기의 필터(Filter)와 스트라이드가 2인 2×2 크기의 맥스풀링(Max Pooling)을 사용한다[14]. 이로 인해 파라미터(Parameter) 개수가 감소한다. 맥스풀링 때문에 크기는 줄었지만 필터 수는 층과 함께 증가한다. VGG Net은 5가지 ConvNet 구성을 제공한다. 예를 들어, VGG-16은 13개의 컨벌루션 층과 3개의 완전연결 층으로 구성된다.

Inception은 크기가 다른 필터와 풀링을 여러 개 적용하여 그 결과를 결합하고, 그 구조를 하나의 모듈로 사용한다[15]. Inception은 1×1 크기 필터를 많은 컨벌루션 층에서 사용한다. 1×1 컨벌루션 연산은 차원이 축소되는 효과가 있기 때문에 파라미터 제거와 고속 처리에 크게 기여한다. Inception은 ILSVRC 2014에서 Top-5 에러율 6.7%를 달성하며 우승을 차지했다.

ResNet에서는 하위 층의 출력을 상위 층에 더한다(이때, 두 출력의 크기는 동일해야 한다)[16]. 크기가 일치하지 않을 경우 선형 변환을 사용하여 문제를 해결한다. 이렇게 하면 역전파(Back-propagation) 때 기울기 소실 및 폭주(Vanishing and Exploding Gradient) 문제가 완화되기 때문에 층의 깊이에 비례해 성능을 향상시킬 수 있다. ResNet은 ILSVRC2015에서 Top-5 에러율 3.5%를 달성하며 우승을 차지했다.

DenseNet은 합산(Addition)이 아닌 연결(Concatenate)을 사용하여 이전 층의 출력과 현재 층의 출력이 합쳐진다[17]. 이 방식은 내부 전체에서 특징 재사용이 가능하여 보다 구체적인 학습, 정확한 모델을 생성할 수 있다. 조밀한 내부 표현 및 특징 중복 감소로 인해 DenseNet은 컨벌루션 연산을 기반으로 한 다양한 컴퓨터 비전 작업에서 우수한 특징 추출기로 자리매김 했다.

NASNet은 동일한 차원의 특징맵을 반환하는 노말 셀(Normal cell)과 높이가 너비가 절반으로 줄어든 특징맵을 반환하는 리덕션 셀(Reduction cell)로 구성된다[18]. 노말 셀과 리덕션 셀의 구조는 신경 구조 탐색(Neural Architecture Search) 프레임워크(Framework) 방법이 결정한다. NASNet은 컨벌루션 필터 뱅크(Bank), 비선형성 및 신중한 연결 선택의 조합으로 구성된 반복 모티프(Motif)를 식별하여 좋은 성능을 보여준다.

3. 맞춤형 CNN의 구조 및 알고리즘

CUMedVis는 Multi-level Contextual 3D ConvNet을 사용했다[19]. Multi-level Contextual 3D ConvNet은 결절 크기, 유형 및 형상에 따른 문제 발생 원인을 규명하기 위해 Archi-a, Archi-b, Archi-c로 구성된다. 각 아키텍처는 다른 수용 영역(Receptive Field)을 가진 입력 영상으로서 특이점 주변의 정보를 획득한다.

Archi-a의 수용 영역은 $20 \times 20 \times 6$ 이다. 첫 번째부터 세 번째 컨벌루션 층의 커널(Kernel) 크기는 $5 \times 5 \times 3$, $5 \times 5 \times 3$, $5 \times 5 \times 1$ 이고, 필터 개수는 64개다. 이후 150개 출력 유닛을 갖는 완전연결 및 소프트맥스(Softmax) 층이 연결된다. Archi-b의 수용 영역은 $30 \times 30 \times 10$ 이다. 첫 번째 컨벌루션 층의 커널 크기는 $5 \times 5 \times 3$ 이고, 필터 개수는 64개다. 두 번째 맥스풀링 층의 커널 크기는 $2 \times 2 \times 1$ 이다. 세 번째와 네 번째 컨벌루션 층의 커널 크기는 $5 \times 5 \times 3$ 이고, 필터 개수는 64개다. 이후 250개 출력 유닛을 갖는 완전연결 및 소프트맥스 층이 연결된다. Archi-c의 수용 영역은 $40 \times 40 \times 26$ 으로 Archi-a, Archi-b보다 크다. 첫 번째 컨벌루션 층의 커널 크기는 $5 \times 5 \times 3$ 이고, 필터 개수는 64개다. 두 번째 맥스풀링 층의 커널 크기는 $2 \times 2 \times 2$ 이다. 세 번째와 네 번째 컨벌루션 층의 커널 크기는 $5 \times 5 \times 3$ 이고, 필터 개수는 64개다. 이후 250개의 출력 유닛을 갖는 완전연결 및 소프트맥스 층이 연결된다. Archi-a, Archi-b, Archi-c로부터 출력된 예측 확률은 가중 선형 결합(Weighted Linear Combination)과 융합되어 입력 영상의 최종 예측을 생성한다.

JackFPR은 CUMedVis(Archi-a, Archi-b, Archi-c)의 구조와 하이퍼파라미터(Hyperparameter)를 일부 수정하여 사용했다. 첫 번째는 활성화 함수(Activation

Function)를 지수함수형 선형 유닛(Exponential Linear Unit)으로 변경했다[20]. 두 번째는 Archi-a, Archi-b, Archi-c 출력력을 완전연결 층으로 연결했다. 마지막은 128개 출력 유닛을 갖는 완전연결 및 소프트맥스 층으로 연결했다.

DIAG CONVNET는 Multi-view ConvNet을 사용했다[21]. Multi-view ConvNet은 후보별 서로 다른 시점에서 65×65 크기의 패치(Patch)를 9개 추출한다. 시점의 간격은 50×50 mm다. 각 시점은 큐브(Cube) 내 다른 대칭면(Plane of Symmetry)에 해당하고, 2D ConvNet 스트림(Stream)을 사용하여 처리한다. ConvNets 스트림은 컨벌루션 층 3개, 맥스풀링 층 1개로 구성된다. 첫 번째 컨벌루션 층의 커널 크기는 5×5 이고, 필터 개수는 24개다. 두 번째 컨벌루션 층의 커널 크기는 3×3 이고, 필터 개수는 32개다. 세 번째 컨벌루션 층의 커널 크기는 3×3 이고, 필터 개수는 48개다. 맥스풀링 층은 입력 패치의 크기를 절반으로 줄인다. 이후 16개 출력 유닛을 갖는 완전연결 및 소프트맥스 층이 연결된다. 가중치(Weight)의 경우 무작위로 초기화한다. 활성화 함수는 정규화 선형 유닛(Rectified Linear Unit, ReLU)[22]을 사용했다. DIAG CONVNET은 컨벌루션 층의 모든 파라미터를 공유하기 때문에 여러 3D 특성을 학습할 수 있다.

ZNET은 최근에 발표된 넓은 잔류 네트워크(Wide Residual Network)를 일부 수정하여 사용했다[23]. ZNET은 후보별 횡단면(Axial Plane), 시상면(Sagittal Plane), 관상면(Coronal Plane)에서 64×64 크기의 패치를 추출했다. 횡단면은 직각이 되게 가로로 잘라 생긴 면, 시상면은 측면에서 수직으로 자른 면, 관상면은 정면에서 수직으로 자른 면이다. ZNET은 각 패치를 별도로 처리하고, 횡단면, 시상면, 관상면에서 추출한 패치들의 출력값을 평균화하여 최종 예측했다. ZNET은 연이은 컨벌루션 층을 4세트 사용했다. 첫 번째 세트의 컨벌루션 층은 하나로 커널 크기는 3×3 , 필터 개수는 16개다. 두 번째부터 네 번째 세트의 컨벌루션 층은 열 개로 커널 크기는 3×3 , 필터 개수는 96, 192, 38개다. 스트라이드는 2로 모두 같다. 각 세트는 또한 스킵 연결(Skip Connection)에서 $1 \times 1 \times N$ 의 투영(Projection) 컨벌루션을 가진다. 여기서 N 은 해당 세트의 필터 개수다. 이후 전역 평균풀링(Global Average Pooling) 및 소프트맥스 층

으로 연결된다[24].

CADIMI는 Multi-slice ConvNet을 사용했다. CADIMI는 후보별 횡단면, 시상면, 관상면에서 세 개의 패치를 추출했다. 정확한 후보 위치와 나머지 자유 축(x, y, z) 양방향의 평면 크기는 2 mm다. 패치는 3차원 배열(Three-dimensional Array)로 연결되고, 후보 위치를 중심으로 $52 \times 52 \times 3$ mm 크기의 패치가 생성된다. Multi-slice ConvNet은 컨벌루션 층 3개, 맥스풀링 층 1개로 구성된다. 첫 번째 컨벌루션 층의 커널 크기는 5×5 이고, 필터 개수는 24개다. 두 번째 컨벌루션 층의 커널 크기는 3×3 이고, 필터 개수는 32개다. 세 번째 컨벌루션 층의 커널 크기는 3×3 이고, 필터 개수는 48개다. 이후 맥스풀링 층을 거치고, 512개 출력 유닛을 갖는 완전연결 및 소프트맥스 층이 연결된다. 활성화 함수는 ReLU를 사용했다.

4. 학습 데이터 셋 및 실험 환경 구성

4.1 학습을 위한 폐 결절 데이터 셋 구성

객관적인 비교 평가를 위해 무료로 다운로드 가능한 LIDC/IDRI(Lung Image Database Consortium/Image Database Resource Initiative) 폐 결절 데이터 셋을 사용했다. LIDC/IDRI 데이터 셋은 흉부 CT 영상에서 폐 결절 탐지 알고리즘의 성능을 평가할 수 있는 공개 리소스다[25]. 초기에 획득한 데이터 셋은 1,018개 CT 영상으로 구성됐지만 슬라이스(Slice) 간격이 일치하지 않거나 슬라이스 되지 않은 9개 영상, 두께가 3 mm 이상인 121개 영상은 데이터 셋에서 제외했다. 결과적으로 영상은 888개가 된다. 본 연구에서는 LUNA16(Lung Nodule Analysis 2016) 웹 사이트에서 위 데이터 셋을 다운로드하여 사용했다[3]. 후보 *.csv(Comma Separated Value) 파일에는 Fig. 1과 같이 x, y, z 세 좌표로 구성된 754,975개의 결절 위치가 포함되어 있다. Fig. 1은 인덱스(Index) 기준으로 상위 5개 행만 출력한 그림이다.

인덱스 0부터 754,974까지 총 754,975개의 결절 후보 행이 있고, 각 행은 널(Null)값이 아닌 5개 컬럼(Column)으로 구성된다. 컬럼 coordX, coordY, coordZ는 3차원 공간에서 결절의 좌표를 나타내는 실수다. seriesuid는 CT 영상 후보자며 각 행의 결절 후보가 소속된 *.mhd(MetaImage header) 파일을 가리킨다. class는 결절 또는 비결절을 의미한다.

해당 파일은 클래스 레이블(Label) 1(결절)의 1,557개 후보 위치, 클래스 레이블 0(비결절)의 753,418개 후보 위치뿐만 아니라 각 후보자에 해당되는 레이블(결절은 1, 비결절은 0) 정보도 제공한다. 각 후보 영상은 x, y, z 좌표 중심에서 48×48 픽셀로 추출했다. 픽셀 강도는 [-1000, 400 하운스필드단위(Hounsfield Unit)] 범위에서 클리핑(Clipping)했고, [0, 1] 범위로 정규화(Normalization)했다. 예시 영상은 Fig. 2와 같고, 왼쪽부터 횡단면, 시상면, 관상면이다.

그리고 결절과 비결절 후보 사이의 데이터 불균형(1,557:753,418) 문제를 완화하기 위해 데이터 증대 기법(확대/축소, 회전, 플립핑, 잘라내기)을 결절만 약 480배 수행했다. 결과적으로, 선택한 CNN 모델을 훈련시키기 위해 총합 150만 개의 데이터를 확보했다. 본 연구에 사용된 데이터 셋 및 후보 *.csv 파일은 LUNA16 챌린지(Challenge)에서 사용된 데이터 셋 구성과 동일하다.

훈련 시 각 네트워크의 영상 크기는 48×48 픽셀로 리사이징(Resizing)했고, 대표 영상은 Fig. 3과 같다. Fig. 3은 무작위 5개 결절 영상의 횡단면, 시상면, 관상면을 각 축 중심에서 다섯 단계로 설정하고, 출력한 것이다.

4.2 실험 환경 구성

분류 층의 출력 클래스 수는 '2'로 변형시켜 사용했

	seriesuid	coordX	coordY	coordZ	class
0	1.3.6.1.4.1.14519.5.2.1.6279.6001.100225287222...	68.420000	-74.480000	-288.700000	0
1	1.3.6.1.4.1.14519.5.2.1.6279.6001.100225287222...	-95.209361	-91.809406	-377.426350	0
2	1.3.6.1.4.1.14519.5.2.1.6279.6001.100225287222...	-24.766755	-120.379294	-273.361539	0
3	1.3.6.1.4.1.14519.5.2.1.6279.6001.100225287222...	-63.080000	-65.740000	-344.240000	0
4	1.3.6.1.4.1.14519.5.2.1.6279.6001.100225287222...	52.946688	-92.688873	-241.067872	0

Fig. 1. Configuration of candidate *.csv file.

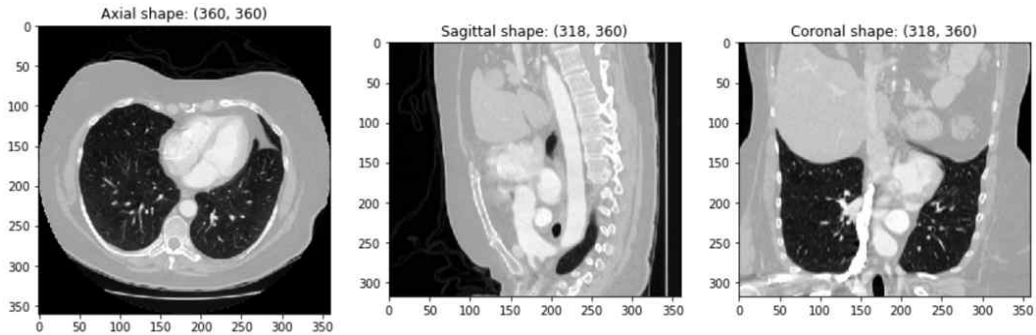


Fig. 2. Examples of clipping and normalized nodule images.

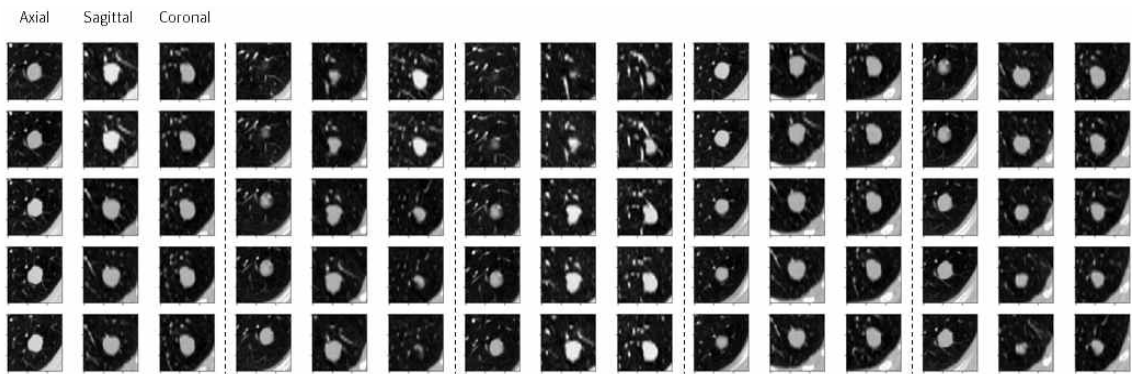


Fig. 3. Representative nodule images used in experiment.

다. 선택된 CNN 모델들은 모두 기성 모델들이었지만 훈련 시 전이 학습(Transfer Learning) 및 미세 조정(Fine Tuning) 방법을 사용하지 않았다. 따라서 훈련은 단대단(End-to-End)으로 진행했다. 가중치의 경우 경험상 잘 작동한다고 알려져 있는 방법으로 초기화했다. LeNet-5, VGG-16, Inception-V3 그리고 DenseNet-201 모델은 ‘Xavier 초기화’, ResNet-152와 NASNet 모델은 ‘He 초기화’ 방법을 사용했다 [26-27]. 최적화기(Optimizer)의 경우 확률적 경사 하강법(Stochastic Gradient Descent)을 사용했고, 학습률(Learning Rate)=‘0.01’, 가중치 감쇠(Weight Decay)=‘0’, 모멘텀(Momentum)=‘0’으로 설정했다. 손실 함수(Loss Function)는 교차 엔트로피(Cross Entropy)를 사용했고, 훈련은 손실이 3에폭(Epoch) 동안 떨어지지 않으면 멈춘다.

하드웨어 사양의 경우 CPU(Central Processing Unit)는 Intel Core i7 7세대 Kaby Lake 7700K, Graphics Card는 NVIDIA TITAN Xp 12GB, RAM은 삼성 DDR4 48GB, SSD(Solid State Drive)는 삼성

성전자 850 Pro 512GB를 사용했다. 소프트웨어 사양의 경우 운영체제(Operating System)는 우분투(Ubuntu) 16.04.4 LTS, CUDA(Compute Unified Device Architecture)는 9.0.176, cuDNN(cuda Deep Neural Network library)은 7.1, 텐서플로(Tensorflow)는 1.12.0, 케라스(Keras)는 2.2.4, 파이썬(Python)은 3.5.2를 사용했다. 케라스는 텐서플로 외에도 다양한 딥러닝 라이브러리(Library)를 호출하여 그에 맞는 알고리즘을 수행한다. 텐서플로는 머신 러닝과 딥러닝을 위한 오픈소스(Open Source) 라이브러리다.

다운로드된 888개의 데이터 셋은 10개의 서브셋으로 나누어져 있다. 본 논문에서는 첫 번째로, 데이터 셋을 훈련 셋과 테스트 셋으로 나눴다. 서브셋 N을 테스트 셋으로 사용하고, 나머지 셋을 훈련 셋으로 사용했다. 두 번째로, 학습은 훈련 셋에서 미니 배치(Mini Batch) 크기만큼 추출한 패치로 진행된다. 최종 결과는 모든 폴드에서 위 과정을 반복한 후 평균했다. 데이터 셋의 10-폴드 교차 검증 수행 방법

은 5장에서 설명한다.

5. 실험 결과

성능은 수신자 조작 특성(Receiver Operating Characteristic)의 곡선 아래 면적(Area Under the Curve, AUC)과 경쟁 성능 지표(Competition Performance Metric, CPM)를 사용하여 평가했다. 자유 반응 수신자 조작 특성(Free-response Receiver Operating Characteristic, FROC)의 민감도는 탐지된 진짜 양성(True Positive) 수를 결절 수로 나눈 값이고, 스캔 당 평균 거짓 양성 수(FPs/scan)로 표시했다. CPM 점수는 사전에 정의된 7가지 거짓 양성 비율(False Positive Rate) 1/8, 1/4, 1/2, 1, 2, 4, 8의 민감도 평균으로 산출했다[28]. 거짓 양성 비율은 전체 음성 샘플 중에서 거짓 양성으로 잘못 분류한 비율이고, 식(1)과 같다.

$$False\ Positive\ Rate = \frac{False\ Positive}{False\ Positive + True-negative} \quad (1)$$

AUC는 후보군을 결절 또는 비결절로 분류할 때의 성능을 나타낸다. CPM은 실제로 사용되는 작동 점(Operating Point)에서의 CAD 성능을 나타낸다. AUC는 항상 0과 1 사이의 값을 가진다. 따라서 데이터 셋에 담긴 클래스가 아무리 불균형하더라도 무작위로 예측한 AUC 값이 0.5가 된다. 그러므로 불균형한 데이터 셋에서는 정확도보다 AUC가 훨씬 나은 지표다. Table 1은 사용된 6개 모델의 벤치마크 성능(훈련 시간, 모델 크기, AUC 및 CPM 점수)을 나타낸다. Table 1에서 볼 수 있듯이 AUC와 CPM 점수는 비례 관계가 아니다.

Fig. 4(a)는 본 실험에서 사용된 6개 CNN 모델의 FROC 곡선이고, Fig. 4(b)는 비교를 위해 삽입한 LUNA16 챌린지의 FROC 곡선으로 결절성 폐암 검출을 위해 자체 제작한 맞춤형 CNN 구조다. CPM

Table 1. Results of off-the-shelf CNN models featured with training time per epoch, model size, AUC and CPM value. Numbers in parentheses indicate the number of subsets used in calculation of the AUC and CPM

Model	Training time / Epoch [sec]	Model Size [MB]	AUC [%]	CPM [%]
LeNet-5	101	2.45	0.999 (8)	0.956 (8)
NASNet	4231	681.8	0.997 (10)	0.954 (10)
VGG-16	755	256.4	0.999 (9)	0.951 (9)
DenseNet-201	1959	147.8	0.999 (8)	0.943 (8)
Inception-V3	1115	175.2	0.996 (7)	0.893 (7)
ResNet-152	2081	257.4	0.979 (7)	0.846 (7)

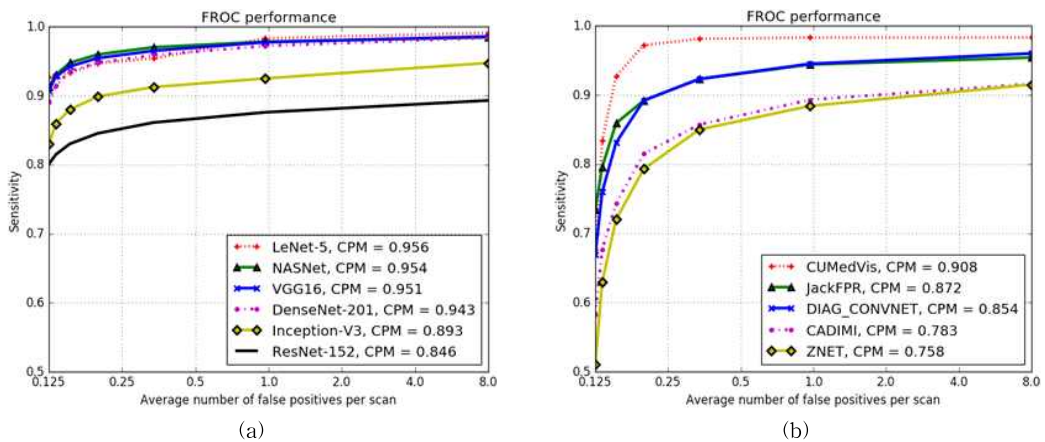


Fig. 4. FROC curve of the pulmonary nodule detection by (a) off-the-shelf network models and (b) customized network models in LUNA 16.

Table 2. CPM values of each subset with the usage status that determined the CPM values of the network models (-x- : unused)

Model	Subset									
	0	1	2	3	4	5	6	7	8	9
LeNet-5	0.896	0.922	×	×	0.966	0.975	0.968	0.975	0.992	0.958
NASNet	0.948	0.951	0.951	0.956	0.958	0.955	0.961	0.957	0.956	0.955
VGG-16	0.967	0.934	-	0.925	0.958	0.955	0.965	0.963	0.962	0.935
DenseNet-201	0.940	0.934	-	-	0.944	0.949	0.961	0.954	0.951	0.914
Inception-V3	0.946	0.924	-	-	-	0.901	0.710	0.920	0.950	0.903

값은 LUNA16 챌린지와 동일한 데이터 셋을 사용하여 산출했다. 이때, CPM 값은 k -폴드 교차 검증 방법(k -Fold Cross Validation)을 사용했고, k 는 '10'으로 설정했다. 그러나 사전에 정의된 FPs 간격의 민감도는 근사값조차 모든 서브셋에 대해 획득하기 어렵다. 따라서 특정 서브셋은 평가에서 제외했다. 미니 배치 크기는 '100'으로 설정했고, 각 모델 훈련에 사용한 알고리즘은 테스트 세트의 오류가 최저인 지점을 기준으로 n 번 동안 성능 향상이 없으면 멈춘다. 본 연구에서는 n 값을 '3'으로 설정했다. 또한, CPM 계산에는 10개 서브셋 중 과반인 7개 이상을 사용했다.

Fig. 4(a)에서 Inception-V3 및 ResNet-152 모델의 CPM 값은 다른 네 개 모델보다 낮지만 기존 LUNA 챌린지의 다른 팀들과 비교할 수 있다. Fig. 4(a) Inception-V3와 ResNet-152 모델의 CPM 값은 Fig. 4(b) JackFPR과 DIAG_CONVNET 모델의 CPM 값과 각각 0.021, 0.008% 차이로 근소하다. 그 외 LeNet-5, VGG-16, DenseNet 그리고 NASNet 모델은 Fig. 4(b)의 CUMedVis 보다 높은 성능을 보였다.

Table 2는 Fig. 4(a) 6개 모델의 CPM 값을 결정할 각 서브셋 CPM 값과 CPM 산출에 사용되지 않은 서브셋을 나타낸다. 예를 들어, Inception-V3의 경우 서브셋 6번의 CPM 값은 0.71%이다. 특히, Inception-V3의 서브셋 6번은 나머지 서브셋의 CPM 값보다 낮다. Fig. 4(a) Inception-V3의 CPM을 서브셋 6번을 제외하여 산출하면 0.9237%가 되고, 그렇게 하면 Fig. 4(b)의 CUMedVis 모델보다 CPM이 높아진다.

6. 결 론

본 논문에서는 결절성 폐암 탐지에 사용할 수 있는 다양한 CNN 모델의 성능을 비교하고, 입증된 기성 CNN 모델의 성능이 맞춤형 CNN 모델보다 우수

함을 보였다. Fig. 4(a)의 6개 모델을 앙상블 하면 더 높은 CPM 값 획득이 가능할 것이라 사료된다[29]. 결과적으로, 폐 결절 탐지 문제는 맞춤형 모델을 직접 설계하는 것보다 입증된 기성 모델들을 사용하는 것이 더 바람직하다고 결론지었다. 특히, LeNet-5 모델은 비교적 훈련 시간이 짧고, 메모리 크기가 작아 모바일 장치에 사용해도 적합할 것이라 판단된다. 그리고 대부분의 CPM 산출 과정에서 서브셋 2번과 3번이 제외됐는데 향후 서브셋 2번과 3번에 어떤 문제점이 있는지 명확하게 밝혀 유사 실험에서 도움이 될 수 있는 연구가 필요하다.

REFERENCE

- [1] American Cancer Society, *Cancer Facts and Figures*, 2016.
- [2] D.R. Aberle, A.M. Adams, C.D. Berg, W.C. Black, J.D. Clapp, R.M. Fagerstrom, et al., "Reduced Lung-cancer Mortality with Low-dose Computed Tomographic Screening," *The New England Journal of Medicine*, Vol. 365, No. 5, pp. 395-409, 2011.
- [3] A.A.A. Setio, A. Traverso, T. deBel, M.S.N. Berens, C.V.D. Bogaard, P. Cerello, et al., "Validation, Comparison, and Combination of Algorithms for Automatic Detection of Pulmonary Nodules in Computed Tomography Images: The Luna16 Challenge," *Medical Image Analysis*, Vol. 42, pp. 1-13, 2017.
- [4] A.A.A. Setio, F. Ciompi, G. Litjens, P. Gerke, C. Jacobs, S.V. Riel, et al., "Pulmonary Nodule Detection in CT Images: False Positive Reduction Using Multi-view Convolutional

- Networks,” *IEEE Transactions Medical Image*, Vol. 35, No. 5, pp. 1160-1169, 2016.
- [5] K. Murphy, B.V. Ginneken, A.M.R. Schilham, B.J.D. Hoop, H.A. Gietema, and M. Prokop, “A Large Scale Evaluation of Automatic Pulmonary Nodule Detection in Chest CT Using Local Image Features and K-nearest Neighbor Classification,” *Medical Image Analysis*, Vol. 13, No. 5, pp. 757-770, 2009.
- [6] T. Messay, R.C. Hardie, and S.K. Rogers, “A New Computationally Efficient CAD System for Pulmonary Nodule Detection in CT Imagery,” *Medical Image Analysis*, Vol. 14, No. 3, pp. 390-406, 2010.
- [7] C. Jacobs, E.M.V. Rikxoort, T. Twellmann, E.T. Scholten, P.A.D. Jong, J.M. Kuhnigk, et al., “Automatic Detection of Subsolid Pulmonary Nodules in Thoracic Computed Tomography Images,” *Medical Image Analysis*, Vol. 18, No. 2, pp. 374-384, 2014.
- [8] M. Firmino, A.H. Morais, R.M. Mendona, M.R. Dantas, H.R. Hekis, and R.A. Valentim, “Computer-aided Detection System for Lung Cancer in Computed Tomography Scans: Review and Future Prospects. Biomed,” *Biomedical engineering online*, Vol. 13, No. 1, pp. 41, 2014.
- [9] C. Jacobs, E.M.V. Rikxoort, K. Murphy, M. Prokop, C.M.S. Prokop, and B.V. Ginneken, “Computer-aided Detection of Pulmonary Nodules: A Comparative Study Using the Public LIDC/IDRI Database,” *European Radiology*, Vol. 26, No. 7, pp. 2138-2147, 2016.
- [10] Y. Lecun, Y. Bengio, and G. Hinton, “Deep Learning,” *Nature*, Vol. 521, No. 7553, pp. 436-444, 2015.
- [11] H. Greenspan, B.V. Ginneken, and R.M. Summers, “Guest Editorial Deep Learning in Medical Imaging: Overview and Future Promise of an Exciting New Technique,” *IEEE Transaction on Medical Imaging*, Vol. 35, No. 5, pp. 1153-1159, 2016.
- [12] G. Litjens, T. Kooi, B.E. Bejnordi, A.A.A. Setio, F. Ciompi, M. Ghafoorian, et al., “A Survey on Deep Learning in Medical Image Analysis,” *Medical Image Analysis*, Vol. 42, pp. 60-88, 2017.
- [13] Y. Lecun, L. Bottou, Y. Bengio, and P. Haffner, “Gradient-based Learning Applied to Document Recognition,” *Proceeding of the IEEE*, pp. 2278-2324, 1998.
- [14] K. Simonyan and A. Zisserman, “Very Deep Convolutional Networks for Large-scale Image Recognition,” *arXiv Preprint arXiv 1409.1556*, 2015.
- [15] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, “Rethinking the Inception Architecture for Computer Vision,” *Proceeding of IEEE Conference on Computer Vision and Pattern Recognition*, pp. 2818-2826, 2016.
- [16] K. He, X. Zhang, S. Ren, and J. Sun, “Deep Residual Learning for Image Recognition,” *Proceeding of IEEE Conference on Computer Vision and Pattern Recognition*, pp. 770-778, 2016.
- [17] G. Huang, G. Liu, L. Maaten, and K.Q. Weinberger, “Densely Connected Convolutional Networks,” *Proceeding of IEEE Conference on Computer Vision and Pattern Recognition*, pp. 2261-2269, 2017.
- [18] B. Zoph, V. Vasudevan, J. Shlens, and Q.V. Le, “Learning Transferable Architecture for Scalable Image Recognition,” *Proceeding of IEEE Conference on Computer Vision and Pattern Recognition*, pp. 8697-8710, 2018.
- [19] Q. Dou, H. Chen, L. Yu, J. Qin, and P.A. Heng, “Multi-level Contextual 3D CNNs for False Positive Reduction in Pulmonary Nodule Detection,” *IEEE Transactions on Biomedical Engineering*, Vol. 64, No. 7, pp. 1558-1567, 2017.
- [20] D.A. Clevert, T. Unterthiner, and S. Hochreiter, “Fast and Accurate Deep Network Learning by Exponential Linear Units (ELUs),” *arXiv Preprint arXiv:1511.07289*, 2015.
- [21] A.A.A. Setio, F. Ciompi, G. Litjens, P. Gerke,

C. Jacobs, S.J.V. Riel, et al., "Pulmonary Nodule Detection in CT Images: False Positive Reduction Using Multi-view Convolutional Networks," *IEEE Transactions on Medical Imaging*, Vol. 35, No. 5, pp. 1160-1169, 2016.

[22] V. Nair and G. Hinton, "Rectified Linear Units Improve Restricted Boltzmann Machines," *Proceedings of International Conference on Machine Learning*, pp. 807-814, 2010.

[23] S. Zagoruyko and N. Komodakis, "Wide Residual Networks," *Computer Vision and Pattern Recognition, arXiv Preprint arXiv:1605.07146*, 2016.

[24] M. Lin, Q. Chen, and S. Yan, "Network in Network," *arXiv Preprint arXiv:1312.4400v3*, 2014.

[25] S.G. Armato, G. McLennan, L. Bidaut, M.F.M. Gray, C.R. Meyer, and A.P. Reeves, et al., "The Lung Image Database Consortium (LIDC) and Image Database Resource Initiative (IDRI) : A Completed Reference Database of Lung Nodules on CT Scans," *Medical Physics*, Vol. 38, No. 2, pp. 915-931, 2011.

[26] G. Xavier and B. Yoshua, "Understanding the Difficulty of Training Deep Feedforward Neural Networks," *Proceeding of International Conference on Artificial Intelligence and Statistics*, pp. 249-256, 2010.

[27] H. Kaiming, Z. Xiangyu, R. Shaoqing, and S. Jian, "Delving Deep into Rectifiers: Surpassing Human-level Performance on ImageNet Classification," *arXiv Preprint arXiv 1502.01852*, 2015.

[28] M. Niemeijer, M. Loog, M.D. Abramoff, M.A. Viergever, M. Prokop, and B. Ginneken, "On Combining Computer-aided Detection Systems," *IEEE Transaction on Medical Imaging*, Vol. 30, No. 2, pp. 215-223, 2011.

[29] S.W. Park, J.C. Kim, D.Y. Kim, "A Study on Classification Performance Analysis of Convolutional Neural Network using Ensemble Learning Algorithm," *Journal of Korea Mul-*

timedia Society, Vol. 22, No. 6, pp. 665-675, 2019.



박 성 옥

2018년 순천대학교 컴퓨터공학과 (공학사)
 2020년 순천대학교 대학원 컴퓨터공학과 졸업(공학석사)
 관심분야 : 컴퓨터비전, 기계학습



김 승 현

2014년 순천대학교 컴퓨터공학과 졸업(공학사)
 2016년 순천대학교 대학원 컴퓨터공학과 졸업(이학석사)
 2019년 순천대학교 대학원 컴퓨터공학과 박사 수료

관심분야 : 인공지능, 기계학습, 빅데이터



임 수 창

2015년 순천대학교 컴퓨터공학과 졸업(공학사)
 2017년 순천대학교 대학원 컴퓨터공학과 졸업(공학석사)
 2019년 순천대학교 대학원 컴퓨터공학과 박사 수료

관심분야 : 컴퓨터비전, 영상처리, 인공지능, 딥러닝



김 도 연

1986년 충남대학교 계산통계학과 졸업(이학사)
 2000년 충남대학교 대학원 정보통신공학과 졸업(공학석사)
 2003년 충남대학교 대학원 컴퓨터공학과 졸업(공학박사)

1986년~1996년 한국원자력연구원 선임연구원
 1997년~2008년 한국전력기술(주) 책임연구원
 2008년~현재 순천대학교 컴퓨터공학과 교수
 관심분야 : 컴퓨터비전, 컴퓨터보안, 기계학습