



Development of an Optimized Deep Learning Model for Medical Imaging

의료 영상에 최적화된 딥러닝 모델의 개발

Young Jae Kim, PhD*[†] , Kwang Gi Kim, PhD[†]

Department of Biomedical Engineering, Gachon University, Incheon, Korea

Deep learning has recently become one of the most actively researched technologies in the field of medical imaging. The availability of sufficient data and the latest advances in algorithms are important factors that influence the development of deep learning models. However, several other factors should be considered in developing an optimal generalized deep learning model. All the steps, including data collection, labeling, and pre-processing and model training, validation, and complexity can affect the performance of deep learning models. Therefore, appropriate optimization methods should be considered for each step during the development of a deep learning model. In this review, we discuss the important factors to be considered for the optimal development of deep learning models.

Index terms Deep Learning; Algorithms; Data Collection; Product Labeling; Diagnostic Imaging

서론

의료 영상 분야는 최근 인공지능, 특히 딥러닝 기술이 가장 활발하게 연구 개발되고 있는 대표적인 분야 중 하나이다. 의료 영상 분야는 1990년대 의료 영상 저장 전송 시스템(Picture Archiving and Communication System; 이하 PACS)의 보급이 이루어지면서 필름 방식의 의료 영상이 디지털화되고, 이후 대량의 의료 정보들이 생성되면서 본격적으로 컴퓨터를 이용한 분석 및 자동 진단에 대한 연구들이 활발하게 이루어지고 있다. 고전적 방법인 영상처리 기술을 의료 영상에 적용하려는 시도와 함께 인공지능 기법 중 기계학습을 적용하려는 시도 또한 계속되어 왔다. 인체에 대한 해부학적 지식 및 임상적 경험에 근거해서 비정형 데이터인 의료 영상을 정형적 특징(feature)으로 추출하고, Artificial Neural Network (이하 ANN), Support Vector Machine과 같은 기계학습 모델을 기반으로 의료 영상으로부터 추출된 특징들을 학습시켜 영상 진단을 하고자 하였다(1-3). 기계학습을 통한 영상 진단에 대한 연구가 활발하게 이루어졌음에도 불구하고, 결과는 좋지 않았다. 개발된 알고리즘들은 의학적 정확

Received October 3, 2020
Revised November 1, 2020
Accepted November 2, 2020

*Corresponding author
Young Jae Kim, PhD
Department of Biomedical Engineering, Gachon University, 21 Namdong-daero 774beon-gil, Namdong-gu, Incheon 21565, Korea.

Tel 82-32-458-2844
Fax 82-32-458-2844
E-mail youngjae@gachon.ac.kr

This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<https://creativecommons.org/licenses/by-nc/4.0>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

ORCID iDs

Young Jae Kim
<https://orcid.org/0000-0003-0443-0051>
Kwang Gi Kim
<https://orcid.org/0000-0001-9714-6038>

[†] These authors contributed equally to this work.

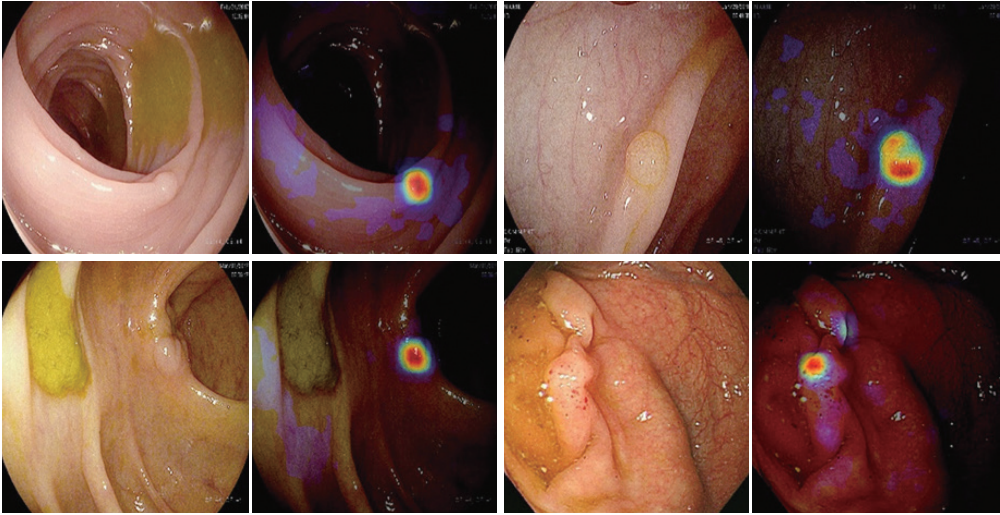
도와 다양한 데이터에서의 강건함(robustness)이 떨어져 실제 임상에 적용하기에는 한계가 있었다. 데이터를 수집하고, 모델을 학습하면, 해당 데이터에서는 좋은 결과를 보이거나 새로운 데이터에서는 성능이 떨어지는 과적합(overfitting) 현상이 발생하였고, 데이터의 변형에 대한 대처 능력이 떨어졌으며, 정확도 또한 높지 않았다. 의료 영상은 인체의 내부를 촬영한 영상으로서, 해부학적 구조가 매우 복잡하고 다양하며, 질환의 다양성까지 포함하고 있다. 또한 사람에 따라 해부학적 차이도 보이며, 촬영 영상의 종류에 따라서도 차이를 보인다. 고전적 영상처리 기술이나 기계학습 방법에서의 알고리즘은 개발자의 직관이나 경험에 근거하기 때문에 다양한 상황과 데이터의 형태 및 변형들을 모두 고려하기가 힘들고, 개발자의 지식을 넘어서는 복잡하고 추상적인 변환들을 모델링할 수 없는 것이 문제였다. 이는 의료 영상에서 기계학습의 한계로 여겨지며 실제 임상 적용이나 상용화까지 이어지지 못한 채 연구 수준에 머물러 있다. 이러한 인식과 상황은 딥러닝 기술의 등장 이후 많은 변화로 이어졌다.

딥러닝은 인공신경망(ANN)에 기반하여 설계된 개념으로 과적합의 문제로 한계를 보였으나 제프리 힌튼(Geoffrey Hinton) 교수가 2006년 사전 학습 개념과 2012년 드랍아웃 개념을 발표하면서 기존의 문제를 극복할 수 있는 것을 증명하였다(4, 5). 이후 하드웨어의 발전과 대량의 빅데이터의 생산이라는 요인이 맞물리면서 딥러닝은 급진적인 발전을 이루었고, 2012년 국제 이미지 인식 기술 대회(ImageNet Large Scale Visual Recognition Challenge)에서 제프리 힌튼 교수팀이 알렉스넷(AlexNet)을 사용하여 1000개의 카테고리에 대해 84.7%의 인식률로 경쟁자들과 10%의 에러율 차이로 압도적 우승을 하고(6), 2016년 구글 딥마인드의 알파고가 이세돌에게 바둑으로 압도적 승리를 하면서 대중에게 딥러닝이라는 개념이 널리 알려지고, 다양한 분야에서 딥러닝을 적용하는 계기가 되었다(7). 다양한 딥러닝 분야에서도 특히 Convolutional Neural Networks (이하 CNN)은 이미지 인식에서 높은 성능을 보이며, 다양한 비전 분야를 포함하여 의료 영상 분야에서도 CNN을 기반으로 많은 연구가 이루어지고 있다(8-10).

사람의 지식과 경험에 의거하여 직접적으로 특징을 추출하는 기존의 머신러닝과 달리 딥러닝은 특징 추출과정 없이 데이터로부터 자체적으로 특징을 추출하고 학습하기 때문에 사람의 개입 없이 데이터 의존적 학습을 수행하게 된다. 이때 딥러닝이 사용하는 특징들은 사람이 직접적으로 정의하는 특징들에 비해 데이터에 근거하여 객관적이고 일관적인 성격을 띠며, 분석 목적에 부합하는 분별력을 갖추므로써 높은 분류 성능을 확보할 수 있다. 이와 같은 학습방법은 딥러닝의 가장 큰 장점이자 단점이라 할 수 있다. 현재의 딥러닝 학습 방법은 데이터로부터 어떠한 과정을 통해 결과를 유추하는지 명확한 원리를 알 수 없기 때문에 블랙박스라고 불린다. 이는 딥러닝이 내린 결정에 대한 신뢰성과 직결되며, 사람의 생명을 다루는 의료 분야에서 특히 문제가 된다. 이러한 문제를 해결하기 위해 딥러닝의 학습 과정을 시각화하기 위한 많은 연구들이 시도되고 있고, 최근에는 Class Activation Map (이하 CAM)을 통해 딥러닝의 결정을 유추할 수 있다(11, 12). Fig. 1은 대장 내시경 영상에서 용종 검출에 대한 딥러닝 모델의 CAM을 보여준다. 딥러닝 모델의 마지막 레이어에서의 가중치 값들을 가시화한 것으로, 용종의 위치에서 강하게 활성화된 것을 컬러로 확인할 수 있다.

딥러닝 모델을 개발하는 과정에서 많은 어려움이 따른다. 대부분의 연구들이 데이터를 많이 수집하고, 딥러닝 모델을 성능이 좋은 최신 모델을 사용하는 것에 집중하는 경향이 있다. 물론 딥러닝 모

Fig. 1. Visualization of deep learning models using Class Activation Map for colonoscopy images. Activation occurs at the location of the polyp.



델을 개발하는데 데이터의 양과 딥러닝 아키텍처(architecture)도 성능을 결정하는 중요한 요소들이지만, 그 외에도 최적의 딥러닝 모델을 개발하기 위해 고려해야 할 요소들이 존재한다. 본 논문에서는 저자의 경험을 기반으로 최적의 딥러닝 모델을 개발하는데 고려해야 할 중요한 요소들을 살펴보고자 한다.

데이터 수집

데이터는 딥러닝 모델을 개발하는데 가장 핵심적이고 중요한 요소라 할 수 있다. 현재 의료 분야에서 대부분의 딥러닝 모델은 지도 학습(supervised learning) 방법을 통해 개발되고 있으며, 지도 학습 방법에서 데이터는 입력 데이터로 사용될 영상의 원시 데이터(raw data)와 출력 데이터, 즉 정답 데이터로 사용될 레이블링 데이터(labeling data)로 구분된다.

의료 데이터는 병원에서만 생산되는 특수한 경우로서, 까다로운 수집절차와 개인정보보호 문제, 전문적 지식의 필요, 병원 외에서의 접근의 어려움 등으로 인해 의료 데이터의 수집에 제한이 많다. 이에 관련 기업 및 연구진들은 대량의 의료 데이터 확보를 위해 주로 오픈 데이터셋을 활용하고 있다. Table 1은 대표적인 오픈 데이터셋을 보여준다(13-20). 가장 큰 규모의 오픈 데이터를 제공하는 곳은 미국 국립 암 연구소(National Cancer Institute)와 미국 국립 보건원(National Institutes of Health; 이하 NIH)이다. 국립 암 연구소는 The Cancer Imaging Archive 데이터베이스를 통해 데이터를 수집 및 제공하고 있으며, 현재 28개 기관에서 수집된 100개 이상의 컬렉션에 대한 3000만 장 이상의 의료 영상을 보유하고 있다(13, 14). 미국 국립 보건원에서는 The National Lung Screening Trial 데이터베이스를 통해 폐암에 대한 CT 25000명의 영상 데이터를 제공하고 있으며(15), ChestX-ray 데이터베이스를 통해 10만 장 이상의 흉부 질환에 대한 X-ray 영상을 제공하고 있다(16). 오픈 데이터는 딥러닝 학습에 부족한 데이터를 보완할 수 있다는 점에서는 훌륭한 보완책이 될 수 있으나, 데이터의 퀄리티 및 신뢰성에 대한 문제점이 있다. 이에 오픈 데이터를 사용할 경우에는 데이터 검증을 통해 신뢰성을 확보할 필요가 있다.

Table 1. Representative Open Data Sets for Medical Image Collection

Provider	Data Name	Collection Type	Modality	Images
NCI	TCIA	Common disease (e.g. lung cancer, breast cancer, colorectal cancer, head and neck cancer)	CT, MR, MG, CR, DX, pathology, etc.	Over 100 unique data collections, 30 million images
NIH	NLST	Lung cancer	CT, pathology	25000 patients, 200000 images
NIH	Deep lesion	Common disease (e.g. lung nodules, liver tumors, enlarged lymph nodes, etc.)	CT	4427 patients, 32120 images
NIH	ChestX-ray8	Common thoracic diseases (atelectasis, cardiomegaly, effusion, infiltration, mass, nodule, pneumonia, pneumothorax)	X-ray	32717 patients, 108948 images
Washington University Knight Alzheimer Disease Research Center	OASIS3	Alzheimer disease	MRI, PET	1098 patients, 2000 MR, 1500 PET
University of South Florida	DDSM	Benign and malignant mass	Mammography	2620 images
University of Michigan, Deep Blue	RIGA	Glaucoma	Fundus image	750 images

CR = computed radiography, DX = digital X-ray, MG = mammography, NCI = National Cancer Institute, NIH = National Institutes of Health, NLST = The National Lung Screening Trial, TCIA = The Cancer Imaging Archive

원시 데이터 수집 시, 다양한 대표성을 띠는 데이터를 수집하여 추후 학습된 모델이 일반화될 수 있도록 하는 것이 중요하다. 영상 간에 유사한 데이터는 학습에 큰 도움이 되지 않는다. 또한 수집된 데이터의 개수 간의 불균형이 일어나지 않도록 하는 것이 좋다. 어쩔 수 없이 불균형이 발생하면 데이터 증강을 통해 보완해 주는 것이 좋다.

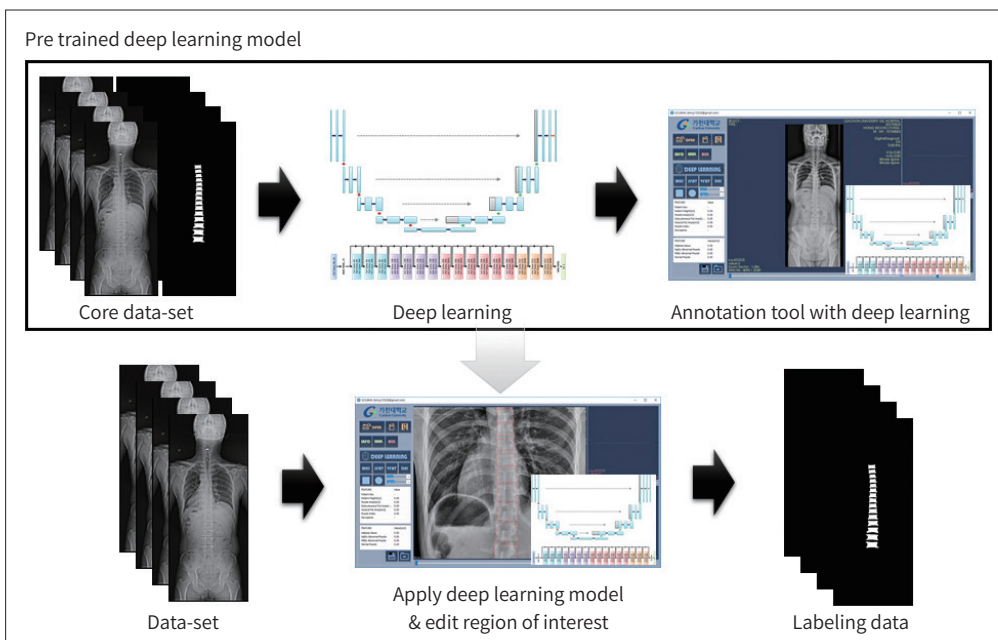
레이블링 데이터는 학습되는 모델의 목적에 따라 수집 방법에 차이가 발생하며, 목적은 크게 분류(classification), 검출(detection), 분할(segmentation)로 구분된다. 우선 카테고리에 따른 분류를 목적으로 하는 경우에는 학습 데이터에 각 카테고리에 대한 번호를 매겨 레이블링 한다. 비교적 레이블링 과정이 간단하기 때문에 빠른 시간에 학습 데이터를 구축할 수 있다는 장점이 있다. 폐 종양의 양/악성 진단, 퇴행성 관절염의 K-grade 예측 등이 분류를 목적으로 하는 경우에 해당된다. 검출은 병변이나 특정 장기 등의 위치를 찾고자 하는 경우이며, 일반적으로 검출 대상의 위치에 박스 또는 원 형태의 관심영역(region of interest)을 그려 레이블링 한다. 검출 대상을 영상에서 찾아 위치를 표시해 주어야 하기 때문에 레이블링에 다소 시간이 걸리지만 관심영역을 그리는 방법이 간단하기 때문에 비교적 빠른 시간에 학습 데이터를 구축할 수 있다. 유방 종괴 검출, 대장 용종 검출 등이 검출을 목적으로 하는 경우에 해당된다. 분할은 주로 특정 대상에 대한 정량적 측정이 필요한 경우에 이루어진다. 병변의 면적 또는 체적의 측정이 필요한 경우, 병변의 형태 관찰이 필요한 경우 등이 해당된다. 분할은 대상의 정확한 구조를 추출해내야 하기 때문에 자유곡선 방식을 통해 분할 대상의 외곽을 따라 관심영역을 그려야 한다. 분할 대상의 구조가 복잡할수록 관심영역을 그리는 시간은 증가하며, 분류 및 검출에 비해 레이블링에 가장 많은 시간이 소요된다.

레이블링 작업은 딥러닝 모델의 개발 과정에서 가장 많은 시간과 노력이 요구된다. 이러한 문제는 선형 학습된 딥러닝 모델을 활용하여 일부 해결할 수 있다. 모델 학습에 큰 기여를 할 수 있는 소량의

코어 데이터셋(core data-set)을 우선적으로 수집한 후, 이를 통해 선형 모델을 학습한다. 이후 나머지 데이터들은 선형 학습된 딥러닝 모델에 적용하여 결과를 얻고, 이를 수정하여 레이블링 데이터를 확보함으로써, 레이블링에 소요되는 시간의 단축이 가능하다(Fig. 2). 소량의 코어 데이터셋으로 학습된 딥러닝 모델은 많은 수정을 필요로 할 수 있으나, 점차 레이블링 데이터가 쌓일수록 딥러닝 모델의 성능도 향상되며, 레이블링 수정에 소요되는 시간 또한 단축될 수 있다.

대부분의 병원에서 사용하는 PACS에서 박스, 원, 자유곡선 등의 관심영역 기능을 제공하고는 있으나 레이블링 데이터에 대한 별도의 내보내기(export) 기능을 제공하고 있지 않기 때문에, 일반적으로 딥러닝 학습 데이터 구축을 위해서는 별도의 레이블링 소프트웨어를 사용해야 한다. 의료 영상을 사용한다는 특수함으로 인해 타 분야에 비해 소프트웨어 선택의 폭이 다소 좁다. Table 2에는 의료 영상을 지원하는 무료 및 유료 레이블링 소프트웨어들을 보여준다(21-25). 무료로 제공되는 소프트웨어 중 가장 대표적인 것은 NIH에서 제공하는 ImageJ 소프트웨어이다. ImageJ는 two-dimensional/three-dimensional (이하 2D/3D) 의료 영상 뷰어 및 영상 처리 기능을 제공하며, Box, Oval, Polygon, Freehand 등의 관심영역 기능을 제공한다(21). 또한 자바 기반의 코드가 공개되어 있고, 플러그인 기능을 제공하여 원하는 기능을 직접 구현하여 추가할 수 있어 의료 영상 분야에서 널리 사용되고 있다. 유료로 제공되는 소프트웨어에는 코어라인의 Aview, 메디컬아이피의 MEDIP 등이 있다. Aview는 의료 영상을 이용하는 연구자들에 특화된 소프트웨어로, 2D/3D 의료 영상 데이터 관리 및 다양한 전처리 기능을 제공하며, Box, Brush 등의 관심영역 기능과 다양한 포맷으로의 저장 기능을 통해 인공지능 학습 데이터를 효과적으로 생산할 수 있다. 메디컬아이피의 MEDIP 또한 기본적인 관심영역 기능 외에 다양한 영상 처리 및 딥러닝 기반의 반자동 분할 기능을 제공하여, 쉽고 빠른 레이블링 작업이 가능하다.

Fig. 2. Labeling using a pre-trained deep learning model based on the core dataset.



데이터 증강

딥러닝 학습을 위한 데이터의 수집 과정에서 의료 분야의 특성상 딥러닝 학습에 사용하기에 데이터가 부족하거나 각 레이블 간 데이터 불균형이 자주 발생한다. 데이터 증강(data augmentation) 기법은 소량의 데이터를 바탕으로 컴퓨터 알고리즘을 통해 데이터의 양을 늘리는 기술로서, 데이터 부족 및 불균형 문제를 해결하기 위한 최적의 수단이다. 학습 데이터의 규모는 딥러닝 모델의 성능에 중요한 영향을 미치는 요소 중 하나이다. 데이터가 부족한 경우 데이터셋의 특징을 잘 반영하지 못하거나 과소적합 및 과적합에 빠질 위험이 증가한다. 또한 레이블 간 데이터의 양이 불균형한 상태에서는 딥러닝 모델의 학습이 데이터가 많은 레이블 쪽으로 편향될 위험이 증가한다. 따라서 데이터 증강 기법을 통해 임의로 데이터의 양을 늘려 부족한 데이터의 양을 보완하거나 레이블 간 데이터의 양의 균형을 맞춰주는 것이 적절하다.

데이터 증강은 기존 영상을 변환하여 증강하는 방법과 새로운 데이터를 생성하여 증강하는 방법으로 구분될 수 있다. Table 3에서는 다양한 데이터 증강 방법들을 적용한 연구들의 사례를 보여준다.

기존 영상의 변환을 통해 증강하는 방법은 데이터 증강에서 일반적으로 사용되는 방법으로, 영상

Table 2. Free and Commercial Labeling Software for Medical Images

Name	Region of Interest	URL
Free software		
ImageJ	Rectangle, oval, polygon, freehand	https://imagej.nih.gov/ij/index.html
ITK-snap	Polygon, paintbrush, smart brush	http://www.itksnap.org/pmwiki/pmwiki.php
Label me	Polygon, rectangle, circle, line and point	https://github.com/wkentaro/labelme
CVAT	Box, polygon, points & line, point	https://github.com/opencv/cvat
Medtagger	Slicing, brush, polygon, pointing	https://github.com/medtagger/MedTagger
Commercial software		
Aview	Brush, polygons, magic cut	https://www.corelinesoft.com/
MEDIP	Polygon, free-draw, pixel-wised, draw cut, region growing	http://www.medicalip.com/
OsiriX	Rectangle, oval, polygon, point, brush, region growing	https://lionbridge.ai/
Labelbox	Box, polygon, points & line, custom attributes	https://labelbox.com/
Trainingdata	Segmentation with brush & eraser, bounding box, polygon tool, freehand with sculpter, key point tool, magnify, 2D growth tool for DICOM	https://www.trainingdata.io/

Table 3. Studies on Data Augmentation Methods

Reference	Modality	Method	Techniques
Mehta & Arbel (26)	MR	Standard	Translation, rotation, scaling, shear transformation
Isensee et al. (27)	MR	Standard	Rotation, scaling, elastic deformation, gamma correction, mirroring
Rahman et al. (28)	X-ray	Standard	Rotation, translation
Yang et al. (29)	X-ray	Standard	Rotation, translation, zoom
Han et al. (31)	MR	GAN	Conditional PGGAN
Sandfort et al. (32)	CT	GAN	CycleGAN
Frid-Adar et al. (33)	CT	GAN	Deep convolutional GAN, auxiliary classifier GAN
Xin et al. (34)	MR	GAN	Pix2pix

GAN = Generative Adversarial Network

의 기하학적 구조는 유지하면서 수학적 변환을 주어 데이터를 늘리는 방법이다. 주로 뒤집기(flip), 회전(rotation), 이동(translation), 배율(scale) 등의 위치적 정보를 변환하는 방법과 영상 밝기(brightness) 및 감마 보정(gamma correction), 블러링(blurring), 샤프닝(sharpening) 등의 픽셀 강도 정보를 기반으로 변환하는 방법 등이 있다(26-29). 다양한 변환 방법 및 변환 파라미터 수준을 랜덤하게 적용하여 한 장의 영상으로부터 기하급수적으로 데이터를 증강할 수 있다. 다만, 증강된 모든 영상들은 기하학적 구조가 그대로 유지되기 때문에 데이터의 대표성이 낮고, 이는 모델의 일반화 성능을 떨어뜨릴 수 있다. 하지만 데이터 불균형에서 발생하는 딥러닝 모델의 편향 문제를 일부 방지 가능하다는 점에서 학습에 분명한 도움이 된다는 것은 사실이다.

새로운 데이터를 생성하여 증강하는 방법은 샘플 데이터를 기반으로 새로운 데이터를 인위적으로 만들어내는 방법으로, 일반적으로 Generative Adversarial Network (이하 GAN)이라는 기법이 사용된다. GAN은 생성자(generator)와 구분자(discriminator)를 경쟁적으로 학습시켜 진짜 데이터와 구분이 어려운 가짜 데이터를 생성해내는 기법이다(30). 최근에는 머신 러닝 분야에서도 GAN에 대한 관심이 가파르게 증가하면서 빠른 발전을 보이고 있으며, GAN의 훈련 성능을 높이기 위한 다양한 알고리즘들이 보고되고 있다(31-34). 2019년 Han 등(31)은 MR에서 2813개의 뇌 전이 암을 기반으로, 조건부로 점차 해상도를 높여 나가는 GAN 학습법인 Conditional PGGAN으로 새로운 뇌 전이 암 증강을 시도하였다. 그 결과, 4000개의 생성된 뇌 전이 암 데이터를 함께 학습하였을 때 뇌 전이 암 검출에 대한 민감도(sensitivity)가 83%에서 91%로 증가하였다고 보고하고 있다. 또한 2019년 Sandfort 등(32)은 10681개의 조영증강된 CT와 603개의 비 조영증강 CT 영상을 통해 Cycle GAN을 이용하여 조영증강된 CT에서 비 조영증강된 CT로의 변환을 시도하였다. 그리고 복부의 신장, 간, 비장을 각각 분할하기 위한 U-Net 모델의 학습에 대해, 원본 데이터만 사용하여 학습한 결과와 Cycle GAN을 통해 생성된 데이터를 원본 데이터와 함께 사용하여 학습한 결과를 비교하였다. 그 결과, 신장은 Cycle GAN을 활용하였을 때 Dice score가 0.059에서 0.664로 향상되었으며, 간은 0.207에서 0.887, 비장은 0.038에서 0.691로 대폭 향상된 결과를 얻을 수 있었다. GAN을 이용한 증강 방법은 영상의 기하학적 구조 자체가 다른 새로운 형태의 영상을 생성해내기 때문에 영상의 위치 또는 밝기 등을 변환하는 방법에 비해 데이터의 대표성이 상대적으로 높다는 장점이 있다. 하지만 GAN을 통해 생성된 데이터는 임상적 검증이 결여된 가짜 데이터로써, 이러한 데이터를 통해 학습된 딥러닝 모델은 임상적 신뢰성이 떨어진다. 환자의 생명과 직결될 수 있는 의료 분야에서 학습 데이터의 임상적 신뢰성은 매우 중요한 문제이다. 딥러닝의 목적에 따라 GAN을 통해 생성된 데이터의 사용 여부를 고려해야 한다. 딥러닝의 목적이 조직 검사의 확진이 필요한 암에 대한 진단과 관련된 경우, GAN을 통해 생성된 데이터는 조직 검사를 시행할 수 없기 때문에 사용을 가급적 피하는 것이 좋다. 따라서 딥러닝의 목적이 양/악성의 감별이 필요하지 않은 경우 또는 장기 영역의 분할이나 검출인 경우, 조직 검사나 추가적인 검사 없이 영상 소견만으로 확진이 가능한 경우에만 GAN을 이용한 데이터 증강 방법을 사용할 것을 권하며, 새롭게 생성된 데이터에 대해 의료진을 통해 충분한 임상적 검증을 거친 후 학습에 사용해야 한다.

데이터 전처리

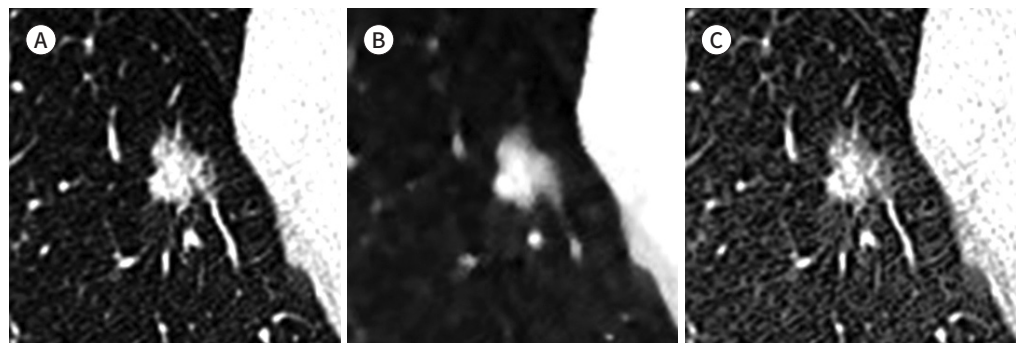
전처리는 딥러닝 모델의 학습 전에 학습 데이터를 대상으로 처리되는 모든 과정을 의미한다. 전처리 과정은 기존의 데이터를 학습에 용이하도록 바꾸는 과정으로, 딥러닝 모델의 성능 향상에 직접적인 영향을 미치기 때문에 적절한 전처리 방법들을 선택하고 적용하는 것은 딥러닝 모델 개발에서 중요한 과정 중 하나이다. 데이터 전처리에는 목적에 따라 다양한 방법들이 존재한다. 본 종설에서는 다양한 전처리 방법들 중 의료 딥러닝 모델의 개발에서 주로 사용되고 있는 비트 변환, 노이즈 제거 및 영상 개선, 정규화, 일괄적 영상 크기 조정 방법들에 대해 살펴보고자 한다.

비트 변환은 의료 영상에서 주로 사용되는 특수한 전처리 방법이다. 하나의 픽셀이 8bit를 가지는 일반적인 영상들에 비해 의료 영상은 종류에 따라 8bit, 12bit, 16bit 등으로 구성된다. 초음파, 안저 영상 등은 8bit의 값을 가지지만 X-ray, CT, MR 등의 대부분의 의료 영상들은 12bit 이상의 값을 가진다. 12bit 이상의 영상은 Window Width/Level에 따라 8bit 영상으로 재구성하여 모니터에 표현된다. 따라서 판독 시 모니터를 통해 보는 영상은 8bit로 변환된 영상이며, 판독 대상에 따라 Window Width/Level을 조절하여 육안 식별에 최적화하여 판독이 이루어진다. 비트 변환은 이러한 특수성을 반영한 방법으로, 12bit 이상의 영상을 최적의 육안 식별이 가능하도록 특정 Window Width/Level을 조절하여 8bit 영상으로 변환하는 방법이다. 물론, 12bit의 영상을 8bit로 변환할 경우 픽셀 손실이 발생한다. 12bit 원본 데이터를 사용하면 더 많은 정보를 통해 딥러닝 모델을 학습할 수는 있겠지만 상황에 따라서는 픽셀 범위를 좁혀주는 것이 도움이 되는 경우도 있다. Window Level을 기준으로 Window Width 범위 밖에 위치한 픽셀 값들은 손실이 되기 때문에, 이를 활용하면 학습 대상 외의 불필요한 부분을 제거해 줌으로써 학습에 긍정적인 영향을 미칠 수 있다. 따라서 학습 대상의 특성에 따라 비트 변환을 통한 전처리 방법의 적용 여부를 결정해야 한다.

저선량 CT의 경우 영상에 많은 잡음이 발생하게 된다. 잡음은 영상 인식을 위한 학습에 불필요한 영향을 미칠 수 있기 때문에 잡음 제거를 위한 필터링 알고리즘을 통해 잡음을 제거하는 것이 도움이 될 수 있다. 잡음 제거에 주로 사용되는 알고리즘으로는 Median Filter (35), Gaussian Filter (36) 등이 있다. 또한 학습 대상이 주변 구조물과 대비(contrast)가 적어 경계의 구분이 어려운 경우 학습이 제대로 이루어지지 않을 가능성이 크다. 따라서 대비 및 선명도를 높이기 위한 전처리 방법을 적

Fig. 3. Example of CT image pre-processing.

- A. Original.
- B. Result of the median filter.
- C. Result of histogram equalization.



용해 주는 것이 좋다. 영상 대비 및 선명도를 높이기 위한 방법으로는 히스토그램 평활화(histogram equalization) (37), 대비 제한 적응 히스토그램 평활화(contrast-limited adaptive histogram equalization) (38) 등이 있다. Fig. 3은 CT에서 잡음 제거와 영상 대비를 높이는 전처리 방법의 결과에 대한 예시를 보여준다.

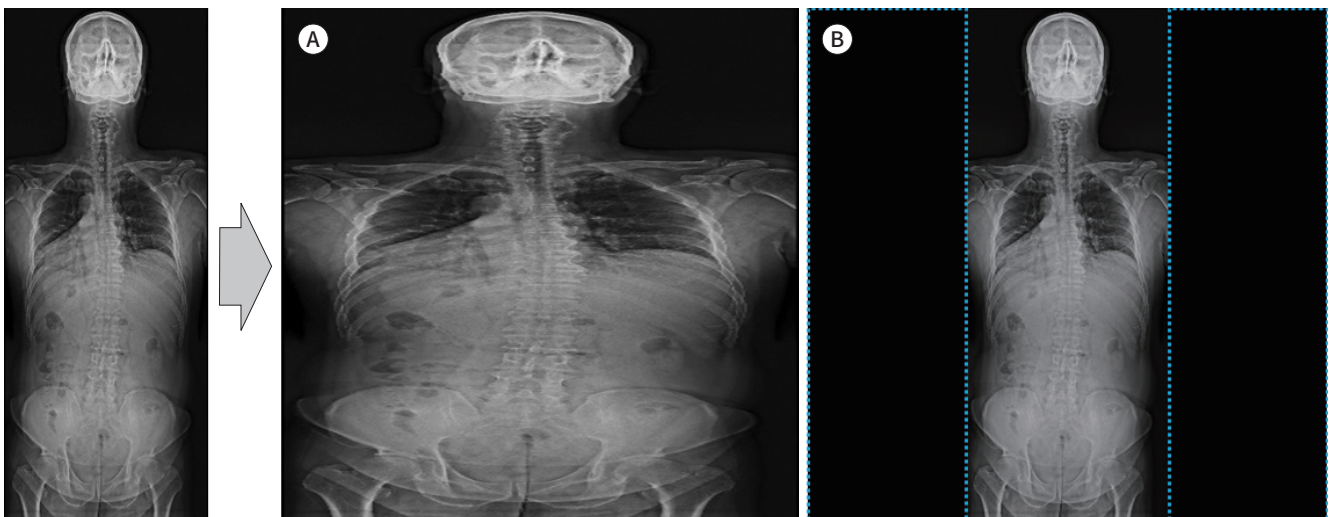
정규화는 입력 영상의 화소값 범위를 제한하는 것이다. 8bit 영상의 경우 0~255, 12bit 영상의 경우 0~4095의 화소 범위를 가지는데 이를 0~1 또는 -1~+1 사이의 범위로 좁혀 딥러닝 모델의 수렴 속도를 높일 수 있다. 일괄적 영상 크기 조정은 학습 데이터의 크기가 서로 다를 경우, 동일한 크기로 조정해 주는 과정을 의미한다. 이때 동일한 크기로 크기 수정(resize) 하는 과정에서 Fig. 4와 같이 가로세로 비율(aspect ratio)을 유지하거나 무시할 수 있다. 가로세로 비율을 무시할 경우에는 원본 영상에서의 형태학적 구조들이 손실된 채 크기 조정이 발생하기 때문에 가급적 가로세로 비율을 유지해 주는 것이 적절하다. 서로 크기가 다른 영상들의 가로세로 비율을 유지한 채 동일한 크기로 수정하기 위해서는 짧은 변을 맞추고 잘라내거나 긴 변을 맞추고 빈칸을 제로 패딩 하여 채울 수 있다 (39). 일반적으로는 긴 변을 맞추는 방법이 사용된다.

딥러닝 모델의 학습

전처리 과정이 끝난 후 사용할 딥러닝 모델을 선정하고, 학습을 위한 하이퍼 파라미터(hyper parameter)에 따라 학습이 이루어진다. 하이퍼 파라미터는 가중치(weight) 같이 모델이 스스로 설정하고 갱신하는 변수가 아닌, 사람이 직접 설정해 주는 매개변수를 의미한다. 대표적인 하이퍼 파라미터로는 배치 (batch)의 크기, 에폭(epochs), 학습률(learning rate) 등이 있다(40). 이러한 하이퍼 파라미터는 값에 따라 모델의 성능이 크게 좌우될 수 있기 때문에 매우 중요하나, 사람이 직접 결정해야 하기 때문에 많은 경험을 필요로 한다. 따라서 하이퍼 파라미터의 최적값이 존재하는 범위를 좁히면서 찾아내는 것이 효율적이다.

Fig. 4. Comparison of image resizing methods.

- A. Interpolation-based resize (ignore aspect ratio).
- B. Zero padding-based resize (maintain aspect ratio).



딥러닝 모델의 학습에는 대량의 학습 데이터가 필요하지만, 의료분야에서는 대량의 학습 데이터를 확보하기 어려운 경우가 많다. 이러한 경우, 유용하게 사용될 수 있는 방법이 전이 학습(transfer learning)이다. 전이 학습이란 학습 데이터가 부족한 경우, 데이터가 풍부한 타 분야에서 사전 훈련된 모델을 재사용하는 학습 기법이다(41, 42). 전이 학습은 학습 데이터의 수가 적을 때 효과적이며, 학습 속도와 정확도를 높일 수 있다는 장점이 있다.

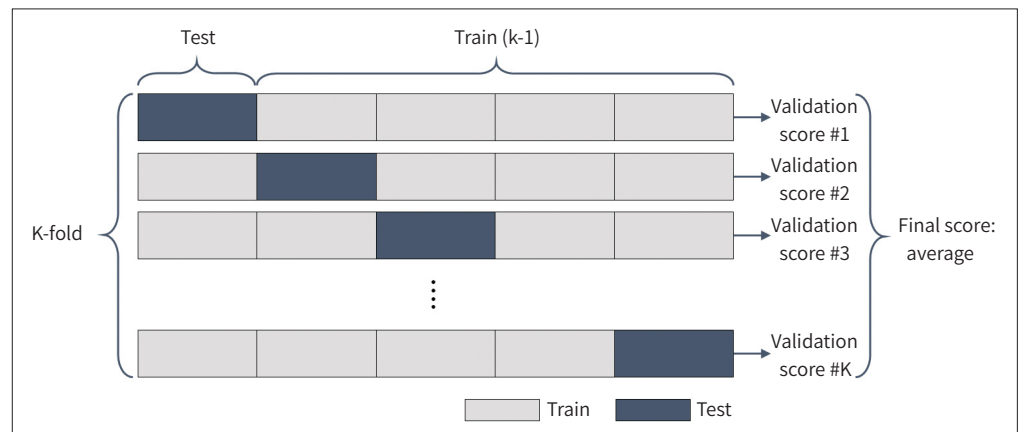
학습 모델의 검증

학습이 완료된 모델은 별도로 구축한 테스트 데이터를 통해 검증이 이루어진다. 이는 학습된 모델을 대상으로 새로운 데이터에서의 일반화 성능을 평가하기 위한 과정이다. 가장 최적의 검증 결과는 학습 데이터에서의 결과와 유사한 성능을 보이는 것이다. 만약 테스트 데이터에서의 결과가 학습 데이터에서의 결과와 상이한 차이를 보이는 경우, 모델이 학습 데이터에 과적합 되었다고 유추할 수 있다.

테스트 데이터가 충분하지 않은 경우에는 검증 결과만으로 모델의 일반화 성능을 단정 짓기는 어렵다. 특히, 의료 데이터는 특성상 모델의 일반화 성능을 확신할 정도의 충분한 데이터를 구축하는 것이 어려운 경우가 많다. 이때 고려할 수 있는 방법이 교차검증이며, 대표적으로 사용되는 교차검증 방법은 K-fold 교차검증이다. 이는 데이터셋의 모든 샘플들이 한 번씩 테스트 될 기회를 갖도록 하는 방법이다(43). Fig. 5와 같이 K-fold 교차검증은 전체 데이터를 사이즈가 동일한 k개의 하부집합으로 나누고 k 번째의 하부집합을 검증용 데이터로, 나머지 k-1개의 하부집합을 훈련용 데이터로 사용한다. 이를 k번 반복 측정하고 각각의 반복 측정 결과를 평균 낸 값을 최종 평가로 사용한다. 하부집합의 분포에 따라 적절한 k의 선정이 필요하다. K의 크기가 너무 작아질수록 학습 데이터의 개수가 적어져 추정 값의 편향이 증가될 위험성이 있고, 데이터의 배분에 따라 모델이 민감해져 추정의 분산도 증가될 위험이 있다. 이에 k의 값은 5 또는 10이 주로 사용되고 있고, 많은 연산과 시간이 걸리지 않는 5-fold 교차검증이 주로 사용되고 있다.

의료 분야에서 교차검증은 학습된 모델의 일반화 성능과 과적합 여부를 검증할 수 있는 좋은 방법이지만, 교차검증만으로 일반화 성능을 확신하기에는 무리가 있다. 테스트 데이터도 결국 동일 기관

Fig. 5. K-fold cross-validation.



에서 확보된 데이터일 가능성이 크기 때문이다. 이에 최근에는 타 기관으로부터 수집된 데이터를 통해 모델의 일반화 성능을 검증하기 위한 외부 검증(external validation)의 필요성이 증가하고 있다(44).

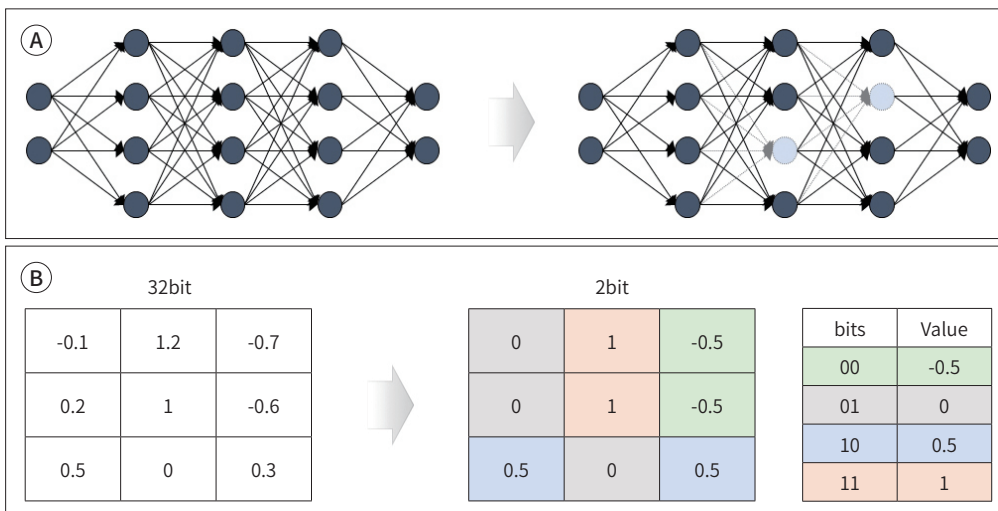
딥러닝 모델의 경량화

딥러닝 모델은 학습 과정에서뿐만 아니라 서비스 활용 시에도 많은 양의 연산이 필요하고, 이를 위해서는 GPU 기반의 고성능 컴퓨팅 파워(computing power)가 요구된다. 그럼에도 불구하고 실시간 처리가 필요하거나 모바일 환경과 같이 리소스 사용에 제한이 있는 경우에는 딥러닝 모델을 활용하기 위한 자원이 부족할 가능성이 높다. 이에 최근에는 모델 경량화를 통해 저사양의 환경에서 적은 연산으로도 정확도를 유지하면서 딥러닝 모델을 학습하고 구동할 수 있는 경량화 기법들에 대한 연구가 활발히 진행되고 있다.

딥러닝 모델의 경량화를 위해 우선적으로 적용할 수 있는 방법은 모델의 네트워크 구조나 합성곱 필터 등이 효율적으로 설계된 딥러닝 모델을 활용하는 것이다. 잔여 블록이나 병목 블록, 밀집 블록 등의 구조 변경을 통해 연산량과 파라미터 수를 줄인 신경망 모델을 활용하는 것으로, ResNet (45), DenseNet (46), SqueezeNet (47) 등의 모델이 있다. 또한 채널별로 합성곱을 수행하고, 점별로 연산을 나누어 연산량과 파라미터를 줄인 모델을 활용할 수 있으며, MobileNet (48), ShuffleNet (49) 등이 있다. 효율적인 구조로 설계된 딥러닝 모델을 활용하는 방법 외에 기존 알고리즘을 다양한 파라미터를 줄임으로써 경량화하는 방법을 활용할 수 있다. 가장 대표적인 방법으로는 모델 압축을 들 수 있다. 모델 압축 방법에는 Fig. 6과 같이 기존 신경망 네트워크의 가중치 중 작은 가중치 값들을 모두 0으로 설정하여 모델의 크기를 줄이기 위한 가중치 가지치기(weight pruning), 신경망의 32bit 부동 소수점 수준을 16bit나 8bit 등으로 줄여서 연산의 효율성을 높이는 양자화 기법 등이 있다(50). 가중치 가지치기는 가중치를 0으로 만들어 뉴런을 삭제하는 드롭 아웃(dropout)과 동일한 효과를 갖는다. 하지만 드롭 아웃은 학습 과정에서 무작위로 가중치를 삭제하지만, 가중치 가지치

Fig. 6. Model compression method for lightweight deep learning models.

- A. Weight pruning.
- B. Quantization.



기는 삭제할 가중치를 선택할 수 있다는 차이점이 있다. 이때 어떠한 뉴런이 관련성이 있는지 여부를 선택하여 불필요한 가중치를 삭제하는 효율적인 알고리즘을 구현하는 것이 가중치 가지치기의 주요 과제라 할 수 있다. 그 외에도 지식 증류 방법과 전이 학습 또한 딥러닝 모델을 경량화하기 위한 방법으로 많이 활용되고 있다(51, 52). 지식 증류 방법은 더 많은 파라미터와 연산량을 기반으로 선행 학습된 큰 규모의 모델(전문가 모델, teacher model)로부터 필요한 정보들을 작고 단순한 규모의 모델(숙련자 모델, student model)에서 전달받아 학습하는 기법으로 적은 규모의 모델에서 큰 규모의 모델과 비슷한 성능을 낼 수 있다는 장점이 있다. 파라미터의 수와 연산량이 많을수록 높은 컴퓨팅 파워(computing power)가 요구되기 때문에 지식 증류 기법을 통해 딥러닝 모델을 작은 규모로 경량화하면서도 기존 모델의 성능을 유지할 수 있다. 전이 학습 또한 지식 증류 방법과 유사하다. 전이 학습은 다른 데이터로 학습된 모델의 결과를 옮겨와 이를 기반으로 학습하는 방법으로, 부족한 데이터와 훈련 시간을 단축할 수 있다는 장점이 있다. 이미 잘 학습된 모델의 결과를 기반으로 학습하기 때문에 적은 파라미터와 연산량에서도 기존보다 높은 수준에서 학습 결과가 수렴되기 때문에 딥러닝 모델의 경량화에 많이 활용되고 있다.

결론

지금까지 의료 영상 분야에서의 딥러닝 모델을 개발하는 과정에서 각 단계별 고려해야 할 중요 요소들과 이를 위한 최적화 방법들에 대해 살펴보았다. 일반화된 성능의 딥러닝 모델을 만들기 위해서는 방대한 데이터가 요구된다. 하지만 방대한 데이터를 모았다고 필히 일반화된 딥러닝 모델이 만들어지는 것은 아니다. 데이터 준비 과정에서부터 데이터 증강, 전처리, 딥러닝 모델의 학습과 검증, 경량화까지 모든 과정이 딥러닝 모델의 일반화 성능을 결정하는데 영향을 미칠 수 있기 때문에 딥러닝 모델의 개발 과정 각각에서 적절한 최적화 방법을 고려해야 한다. 딥러닝의 목적과 데이터의 규모, 사용하는 영상의 특성 등을 우선적으로 고려하고, 이에 필요한 최적화 방법들을 적절하게 사용해야 한다.

현재 인공지능 기술은 빠른 속도로 발전하고 있고, 다양한 분야에서 도입되어 우리의 생활 속에 빠르게 적용되고 있다. 의료 분야 또한 데이터 규제 개선과 적극적인 투자로 인공지능 기술들을 상용화하려는 시도가 점차 증가하고 있다. 의료 분야에서 인공지능이 의사의 역할을 대체하는 것이 아닌 의사를 보조하는 수단으로서 개발이 주를 이루고 있으나 환자들의 생명을 다루는 분야인 만큼 인공지능의 역할이 중요하다고 할 수 있다. 여기서 인공지능의 역할은 임상적 신뢰성을 바탕으로 정확한 예측과 일관된 성능을 보이는 것이다. 그리고 이를 위해서는 의료 영상의 임상적 관점을 공학적으로 풀어내어 딥러닝 모델의 개발과정에 녹여내고, 이를 최적화하기 위한 방법들을 고려해야 할 것이다.

Author Contributions

Conceptualization, all authors; data curation, all authors; formal analysis, all authors; investigation, all authors; methodology, K.Y.J.; project administration, all authors; resources, all authors; software, K.Y.J.; supervision, K.K.G.; validation, all authors; visualization, all authors; writing—original draft, K.Y.J.; and writing—review & editing, all authors.

Conflicts of Interest

The authors have no potential conflicts of interest to disclose.

Acknowledgments

This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (No. NRF-2019R1C1C1008381), and Institute of Information & communications Technology Planning & Evaluation (IITP) grant funded by the Korea government (MSIT) (No. 2019-0-01750-002, Development of an optimal limb-compressing cardiovascular treatment device using deep learning technique) (No. 2020-0-00161-001, Active Machine Learning based on Open-set training for Surgical Video).

REFERENCES

1. Amato F, López A, Peña-Méndez EM, Vaňhara P, Hampl A, Havel J. Artificial neural networks in medical diagnosis. *J Appl Biomed* 2013;11:47-58
2. Ren J. ANN vs. SVM: which one performs better in classification of MCCs in mammogram imaging. *Knowl Based Syst* 2012;26:144-153
3. Çınar M, Engin M, Engin EZ, Ateşçi YZ. Early prostate cancer diagnosis by using artificial neural networks and support vector machines. *Expert Syst Appl* 2009;36:6357-6361
4. Hinton GE, Salakhutdinov RR. Reducing the dimensionality of data with neural networks. *Science* 2006; 313:504-507
5. Hinton GE, Srivastava N, Krizhevsky A, Sutskever I, Salakhutdinov RR. Improving neural networks by preventing co-adaptation of feature detectors. *ArXiv Preprint* 2012;arXiv:1207.0580
6. Krizhevsky A, Sutskever I, Hinton GE. *ImageNet classification with deep convolutional neural networks*. New York: Communications of the ACM 2012:1-9
7. Silver D, Huang A, Maddison CJ, Guez A, Sifre L, Van den Driessche G, et al. Mastering the game of go with deep neural networks and tree search. *Nature* 2016;529:484-489
8. Tajbakhsh N, Shin JY, Gurudu SR, Hurst RT, Kendall CB, Gotway MB, et al. Convolutional neural networks for medical image analysis: full training or fine tuning? *IEEE Trans Med Imaging* 2016;35:1299-1312
9. Chen J, Wang Y, Wu Y, Cai C. An ensemble of convolutional neural networks for image classification based on LSTM. Proceedings of 2017 International Conference on Green Informatics (ICGI); 2017 Aug 15-17; Fuzhou, China: ICGI; 2017:217-222
10. Anwar SM, Majid M, Qayyum A, Awais M, Alnowami M, Khan MK. Medical image analysis using convolutional neural networks: a review. *J Med Syst* 2018;42:226
11. Wang Z, Yan W, Oates T. Time series classification from scratch with deep neural networks: a strong baseline. Proceedings of IJCNN 2017 : International Joint Conference on Neural Networks; 2017 May 14-19; Anchorage, AK, USA: IJCNN; 2017:1578-1585
12. Philbrick KA, Yoshida K, Inoue D, Akkus Z, Kline TL, Weston AD, et al. What does deep learning see? Insights from a classifier trained to predict contrast enhancement phase from CT images. *AJR Am J Roentgenol* 2018;211:1184-1193
13. Bona JP, Prior FW, Zozus MN, Brochhausen M. Enhancing clinical data and clinical research data with biomedical ontologies - Insights from the knowledge representation perspective. *Yearb Med Inform* 2019; 28:140-151
14. Basu A, Warzel D, Eftekhari A, Kirby JS, Freymann J, Knable J, et al. Call for data standardization: lessons learned and recommendations in an imaging study. *JCO Clin Cancer Inform* 2019;3:1-11
15. Aberle DR, Berg CD, Black WC, Church TR, Fagerstrom RM, Galen B, et al. The National Lung Screening Trial: overview and study design. *Radiology* 2011;258:243-253
16. Yan K, Wang X, Lu L, Zhang L, Harrison AP, Bagheri M, et al. Deep lesion graphs in the wild: relationship learning and organization of significant radiology image findings in a diverse large-scale lesion database. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition; 2018 Jun 18-22; Salt Lake City, UT, USA: IEEE; 2018:9261-9270
17. Wang X, Peng Y, Lu L, Lu Z, Bagheri M, Summers RM. 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 (CVPR); 2017 Jul 21-26; Honolulu, HI, USA: IEEE; 2017:2097-2106
18. LaMontagne PJ, Benzinger TL, Morris JC, Keefe S, Hornbeck R, Xiong C, et al. OASIS-3: longitudinal neuroimaging, clinical, and cognitive dataset for normal aging and Alzheimer disease. *MedRxiv* 2019 [in press] doi: <https://doi.org/10.1101/2019.12.13.19014902>
 19. Benndorf M, Herda C, Langer M, Kotter E. Provision of the DDSM mammography metadata in an accessible format. *Med Phys* 2014;41:051902
 20. Yu S, Xiao D, Frost S, Kanagasigam Y. Robust optic disc and cup segmentation with deep learning for glaucoma detection. *Comput Med Imaging Graph* 2019;74:61-71
 21. Schneider CA, Rasband WS, Eliceiri KW. NIH image to ImageJ: 25 years of image analysis. *Nat Methods* 2012;9:671-675
 22. Yushkevich PA, Gao Y, Gerig G. ITK-SNAP: an interactive tool for semi-automatic segmentation of multi-modality biomedical images. Proceedings of the 2016 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC); 2016 Aug 17-20; Orlando, FL, USA: IEEE; 2016:3342-3345
 23. Russell BC, Torralba A, Murphy KP, Freeman WT. LabelMe: a database and web-based tool for image annotation. *Int J Comput Vis* 2008;77:157-173
 24. Hansen US, Landau E, Patel M, Hayee BH. Novel artificial intelligence-driven software significantly shortens the time required for annotation in computer vision projects. *MedRxiv* 2020 [in press] doi: <https://doi.org/10.1101/2020.09.11.20192500>
 25. Liu H, Bielinski SJ, Sohn S, Murphy S, Waghlikar KB, Jonnalagadda SR, et al. An information extraction framework for cohort identification using electronic health records. *AMIA Jt Summits Transl Sci Proc* 2013; 2013:149-153
 26. Mehta R, Arbel T. *3D U-Net for brain tumour segmentation*. In *International MICCAI Brainlesion Workshop*. Cham: Springer 2018:254-266
 27. Isensee F, Kickingereder P, Wick W, Bendszus M, Maier-Hein KH. *No new-net*. In *International MICCAI Brainlesion Workshop*. Cham: Springer 2018:234-244
 28. Rahman T, Khandakar A, Kadir MA, Islam KR, Islam KF, Mazhar R, et al. Reliable tuberculosis detection using chest X-ray with deep learning, segmentation and visualization. *IEEE Access* 2020;8:191586-191601
 29. Yang S, Kweon J, Roh JH, Lee JH, Kang H, Park LJ, et al. Deep learning segmentation of major vessels in X-ray coronary angiography. *Sci Rep* 2019;9:13897
 30. Antoniou A, Storkey A, Edwards H. Data augmentation generative adversarial networks. *ArXiv Preprint* 2017;arXiv:1711.04340
 31. Han C, Murao K, Noguchi T, Kawata Y, Uchiyama F, Rundo L, et al. Learning more with less: conditional PGGAN-based data augmentation for brain metastases detection using highly-rough annotation on MR images. Proceedings of the 28th ACM International Conference on Information and Knowledge Management; 2019 Nov 3-7; Beijing, China: CIKM; 2019:119-127
 32. Sandfort V, Yan K, Pickhardt PJ, Summers RM. Data augmentation using generative adversarial networks (CycleGAN) to improve generalizability in CT segmentation tasks. *Sci Rep* 2019;9:16884
 33. Frid-Adar M, Diamant I, Klang E, Amitai M, Goldberger J, Greenspan H. GAN-based synthetic medical image augmentation for increased CNN performance in liver lesion classification. *Neurocomputing* 2018;321:321-331
 34. Xin B, Hu Y, Zheng Y, Liao H. Multi-modality generative adversarial networks with tumor consistency loss for brain MR image synthesis. Proceedings of 2020 IEEE 17th International Symposium on Biomedical Imaging (ISBI); 2020 Apr 3-7; Iowa City, IA, USA: IEEE; 2020:1803-1807
 35. Ning CY, Liu SF, Qu M. Research on removing noise in medical image based on median filter method. Proceedings of 2009 IEEE International Symposium on IT in Medicine & Education (ITME2009); 2009 Aug 14-16; Jinan, China: IEEE; 2009:384-388
 36. Cadena L, Zotin A, Cadena F, Korneeva A, Legalov A. Noise reduction techniques for processing of medical images. Proceedings of the World Congress on Engineering 2017; Jul 5-7; London, UK: World Congress on Engineering; 2017:5-9
 37. Wang Q, Chen L, Shen D. Fast histogram equalization for medical image enhancement. Proceedings of 2008 30th Annual International Conference of the IEEE Engineering in Medicine and Biology Society; 2008

Aug 20-25; Vancouver, BC, Canada: IEEE; 2008:2217-2220

38. Reza AM. Realization of the contrast limited adaptive histogram equalization (CLAHE) for real-time image enhancement. *J Signal Process Sys* 2004;38:35-44
39. Hashemi M. Enlarging smaller images before inputting into convolutional neural network: zero-padding vs. interpolation. *J Big Data* 2019;6:98
40. Smith LN. A disciplined approach to neural network hyper-parameters: part 1 -- learning rate, batch size, momentum, and weight decay. *ArXiv Preprint* 2018;arXiv:1803.09820
41. Shin HC, Roth HR, Gao M, Lu L, Xu Z, Nogues I, et al. Deep convolutional neural networks for computer-aided detection: CNN architectures, dataset characteristics and transfer learning. *IEEE Trans Med Imaging* 2016;35:1285-1298
42. Shi Z, Hao H, Zhao M, Feng Y, He L, Wang Y, et al. A deep CNN based transfer learning method for false positive reduction. *Multimed Tools Appl* 2019;78:1017-1033
43. Wong TT, Yeh PY. Reliable accuracy estimates from k-fold cross validation. *IEEE Trans Knowl Data Eng* 2020;32:1586-1594
44. Bleeker SE, Moll HA, Steyerberg EW, Donders AR, Derksen-Lubsen G, Grobbee DE, et al. External validation is necessary in prediction research: a clinical example. *J Clin Epidemiol* 2003;56:826-832
45. Szegedy C, Ioffe S, Vanhoucke V, Alemi A. Inception-v4, inception-resnet and the impact of residual connections on learning. *ArXiv Preprint* 2016;arXiv:1602.07261
46. Iandola F, Moskewicz M, Karayev S, Girshick R, Darrell T, Keutzer K. Densenet: implementing efficient convnet descriptor pyramids. *ArXiv Preprint* 2014;arXiv:1404.1869
47. Iandola FN, Han S, Moskewicz MW, Ashraf K, Dally WJ, Keutzer K. SqueezeNet: AlexNet-level accuracy with 50x fewer parameters and < 0.5 MB model size. *ArXiv Preprint* 2016;arXiv:1602.07360
48. Li Y, Huang H, Xie Q, Yao L, Chen Q. Research on a surface defect detection algorithm based on MobileNet-SSD. *Appl Sci* 2018;8:1678
49. Zhang X, Lin M, Sun J. Shufflenet: an extremely efficient convolutional neural network for mobile devices. Proceedings of the IEEE conference on computer vision and pattern recognition; 2018 Jun 19-21; Salt Lake City, UT, USA: IEEE; 2018:6848-6856
50. Han S, Mao H, Dally WJ. Deep compression: compressing deep neural networks with pruning, trained quantization and Huffman coding. *ArXiv Preprint* 2015;arXiv:1510.00149
51. Chen G, Choi W, Yu X, Han T, Chandraker M. Learning efficient object detection models with knowledge distillation. Proceedings of Advances in Neural Information Processing Systems; 2017 Dec 4-7; Long Beach, CA, USA: NIPS; 2017:1-10
52. Yim J, Joo D, Bae J, Kim J. A gift from knowledge distillation: fast optimization, network minimization and transfer learning. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition; 2017 Jul 21-26; Honolulu, HI, USA: IEEE; 2017:4133-4141

의료 영상에 최적화된 딥러닝 모델의 개발

김영재* · 김광기

최근, 의료 영상 분야에서 딥러닝은 가장 활발하게 연구되고 있는 기술 중 하나이다. 충분한 데이터와 최신의 딥러닝 알고리즘은 딥러닝 모델의 개발에 중요한 요소이다. 하지만 일반화된 최적의 딥러닝 모델을 개발하기 위해서는 데이터의 양과 최신의 딥러닝 알고리즘 외에도 많은 것을 고려해야 한다. 데이터 수집부터 가공, 전처리, 모델의 학습 및 검증, 경량화까지 모든 과정이 딥러닝 모델의 성능에 영향을 미칠 수 있기 때문이다. 본 종설에서는 의료 영상에 최적화된 딥러닝 모델을 위해 개발 과정 각각에서 고려해야 할 중요한 요소들을 살펴보고자 한다.

가천대학교 의용생체공학과