

뇌종양 진단을 위한 자기공명영상(MRI) 이미지 분류 CNN 모델 비교분석

Comparative Analysis of Magnetic Resonance Imaging (MRI) Image Classification CNN Model for Brain Tumor Diagnosis

문 승 재¹
Seung-jae Moon

요 약

본 논문은 MRI 영상의 자동 분류를 통해 뇌종양 진단에 있어 오검출 빈도와 검출 소요 시간을 줄일 수 있는 CNN 알고리즘 기반의 VGG, ResNet, Inception V3과 추가로 제작한 커스텀 모델을 포함한 네 가지 딥러닝 기반 모델을 제시하여 비교 분석한다. 웹상에 공개된 뇌종양 MRI 이미지 데이터를 각 모델에 학습시키고 실험한 결과, Inception V3 모델이 다양한 평가지표에서 뛰어난 성능을 띄는 것을 확인하였으며, 커스텀 모델의 발전 가능성 또한 확인하였다.

☞ 주제어 : 뇌종양 진단, MRI, 자동 분류, CNN, 딥러닝

ABSTRACT

This paper presents and comparative analyzes four deep learning-based models: VGG, ResNet, Inception V3 and additionally created custom model based on CNN algorithms, which can reduce the frequency of false detections and detection time in brain tumor diagnosis through automatic classification of MRI images. As a result of training and testing each model on brain tumor MRI image data published on the web, it was confirmed that the Inception V3 model showed excellent performance in various evaluation metrics, and the possibility of developing a custom model was also confirmed.

☞ keyword : Brain Tumor Diagnosis, MRI, Automatic Classification, CNN, Deep-Learning

1. 서 론

뇌종양이란 뇌 조직이나 뇌를 싸고 있는 막에서 발생한 종양과 머리뼈나 주변 구조물로부터 멀리 떨어진 부위에서 뇌 조직이나 뇌막으로 전이된 종양을 의미한다. 뇌종양은 양성종양과 악성종양으로 나뉘지만, 조직학적으로 양성일 뿐, 신체의 중추신경계를 담당하는 뇌와 척수에 심각한 손상을 입힐 수 있는 위험한 질병이다[1]. 뇌종양에는 뇌의 조직에서 발생하는 원발성 종양과 인체의 다른 부분에서 혈류를 통해 뇌 조직으로 타고 들어가 확장된 이차성 종양으로 나뉜다. 뇌종양 중 가장 흔하게 발생하는 종양으로는 원발성 종양에 속하는 신경교종으로, 초기에 진단되지 않으면 사망까지 이를 수 있는 종양이다[2]. 또한, 몸에서 가장 중요한 신체 부위라고 할 수 있는 뇌는 어떠한 종류의 종양이더라도 치료가 늦으면 치명적일 수 있기 때문에, 빠른 진단과 조기 치료가 중요하다. 현재 뇌종양 진단을 위한 기술로 컴퓨터 단층 촬영(CT), 양전자 방출 단층 촬영(PET), 및 자기 공명 영상(MRI) 등 다양한 의료영상

기술이 존재한다[3]. 해당 방식 중 MRI는 뇌종양의 크기, 위치, 유형 등에 유의미한 정보를 단면뿐만 아니라 다차원 영상과, 높은 해상도와 같은 자유로운 형태로 제공할 수 있는 유일한 방식으로 가장 선호된다. 하지만 영상의학 전문가가 MRI 영상을 일일이 분석하여 종양을 검출하는 데는 오검출이 발생할 수 있고 긴 검출 시간이 소요된다. 이때, 뇌 MRI 영상의 자동 분류는 뇌종양 진단에 있어 오검출 빈도와 검출 소요 시간을 줄이는 데 중요한 역할을 한다[4].

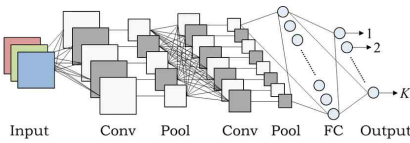
최근 딥러닝 기술이 발전하며 영상 처리 기법에 특화된 알고리즘인 CNN (Convolution Neural Network)을 기반으로 한 다양한 모델이 등장하고 있다. 이 모델들을 뇌종양 진단을 위한 MRI 영상 분류에 적용하여 자동화하면 빠르고 정확하게 진단을 할 수 있다.

본 논문은 CNN 알고리즘 기반의 다양한 모델을 구축하며, 구축된 모델 중 가장 좋은 성능의 모델을 찾기 위해 직접 제작한 커스텀 모델과 ResNet, VGG, Inception V3 모델을 정확도, 손실률 등 다양한 평가지표를 통해 비교 분석한다.

2. 관련 연구

2.1 CNN 알고리즘

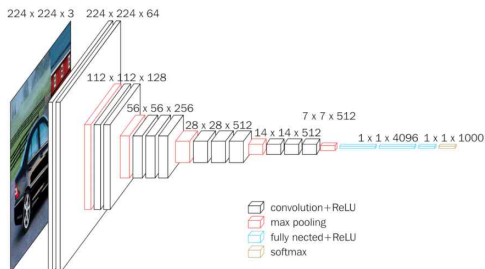
CNN 이란 Convolution Neural Networks의 약자로, 동물의 뇌 구조를 모방하여 만들어진 딥러닝 신경망 모델이다. CNN은 합성곱(Convolution) 계층과 풀링(Pooling) 계층의 반복 연결을 통해 이미지의 특징을 추출한다. 합성곱 계층에서는 2차원 배열 형태의 이미지를 입력하고, 일정 크기의 필터를 가중치로 연산하는 특징 추출을 통해 특징 지도(feature map)를 생성한다. Pooling 계층에서는 기본적으로 생성된 특징지도를 Down Sampling 하여 크기를 감소시킨다. 이러한 차원 축소는 특징점의 위치 변화의 영향을 줄이고, 연산량을 줄이는 역할을 한다. 이렇게 두 층을 통해 특징 추출과 차원 축소를 반복하여 1차원 특징 배열을 출력값으로 가지고, 이를 전결합층(Fully-Connected Layer)의 입력값으로 분류 알고리즘 등을 통해 모델이 생성된다.



(그림 1) 합성곱 신경망
(Figure 1) Convolution Neural Network

2.2 VGG

VGG 모델은 2012년 이미지넷 대회에서 준우승을 한 모델이다. VGG 모델은 필터와 커널의 크기를 3x3으로 고정하여 네트워크층의 깊이에 따른 모델의 성능을 확인하기 위한 모델이다. VGG 모델에는 네트워크 층수에 따라 VGG-11, VGG-13, VGG-16, VGG-19 등으로 나뉜다.



(그림 2) VGG-16 모델 구조
(Figure 2) Structure of VGG-16

그림 2는 VGG16 모델의 구조이다. Convolution 커널의 사이즈는 3x3, Maxpooling 커널의 크기는 2x2이며 stride는 2이다. Pooling 레이어를 제외하고 Convolution 레이어와 Fully Connected 레이어의 개수를 합하면 총 16개의 층으로 이루어져 있고, 맨 마지막 레이어를 제외하면 활성화 함수로 ReLU 함수를 적용한다. VGG 개발연구논문에 따르면 마지막 층은 총 1,000개의 뉴런까지 구성할 수 있어 총 1,000가지 클래스를 분류할 수 있다[5]. 본 논문에서는 실험 환경 문제로 VGG11 모델을 사용한다.

2.3 ResNet

CNN의 레이어를 깊게 쌓을수록 성능이 좋아질 것이라는 예상과는 달리, 계속되는 ReLU 함수의 역전과 과정으로 인해 기울기가 점차 작아져 소실되는 Gradient Vanishing 현상과 기울기가 커지며 가중치가 비정상적으로 커지는 Gradient Exploding 현상이 나타나 모델의 성능이 떨어지는 현상이 발생하였다. 레이어 개수를 늘려와 동시에 성능을 높이며 이러한 현상을 예방하기 위한 모델로 ResNet이 등장한다.

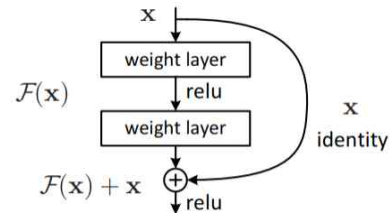


Figure 2. Residual learning: a building block.

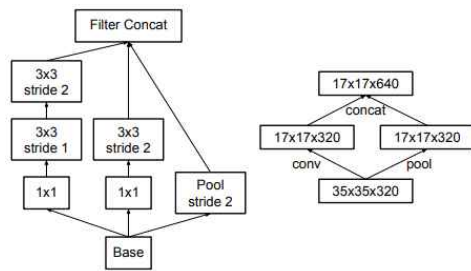
(그림 3) Residual Block & skip connection
(Figure 3) Residual Block & skip connection

ResNet은 그림 3과 같이 skip connection 개념을 사용하는 Residual Block을 연속적으로 쌓아 구성된 네트워크이다. skip connection이란 하나의 레이어의 출력을 몇 개의 레이어를 건너뛰고 다음 레이어의 입력으로 하는 방식이다. 이때 그림 3의 F(x)는 0이 학습되도록 진행한다. 이러한 방식은 역전파를 수행하여 미분을 하더라도 1이 남아 기울기 소실 문제를 방지할 수 있다. ResNet은 매우 깊은 모델임에도 파라미터 개수를 줄여 전체적인 연산량을 줄이는 동시에 오버피팅을 예방하고, 레이어 수를 늘려 높은 성능을 보여준다[6].

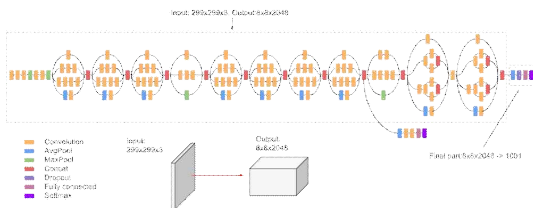
2.4 Inception V3

Inception V3 모델은 구글에서 발표한 모델로, 기존 GoogLeNET (Inception V1)을 기반으로 모델의 크기가 커지며 생기는 과적합으로 인한 정확도 문제와 높은 연산량으로 인한 속도 문제를 해결하기 위해 고안한 모델이다. 1x1의 더욱 작은 단위의 Convolution 레이어를 이용하여 모델의 깊이를 더하고, 차원을 축소하여 연산량을 크게 줄였다. 또한, 비대칭 Convolution 레이어를 사용하여 연산량을 1/3로 줄인다.

하지만, Pooling과 Convolution을 할 때, bottleneck 현상으로 인해 정보손실이 일어나거나, 연산량의 증가로 인한 성능의 저하가 발생하였다. 이러한, 현상을 해결하기 위해 그림 5와 같은 두 구조를 병렬로 진행 후 합치는 방식으로 설계하였다.



(그림 5) Inception V3 module
(Figure 5) Inception V3 module



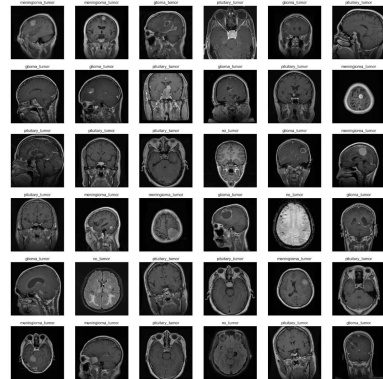
(그림 6) Inception V3 구조
(Figure 6) Structure of Inception V3

Inception V3는 그림 5와 같은 module을 그림 6과 같이 여러 번 이어 붙여 생성한 모델로, Target 값을 과적합 방지를 위해 원-핫 인코딩을 사용하지 않는다는 특징이 있다. 최적화 함수로는 RMSProp을 사용하고, 마지막 Fully Connected layer에서 배치 정규화 과정을 거쳐 출력된다[7].

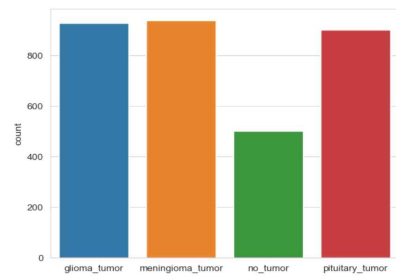
3. 데이터

3.1 연구 데이터 세트 구성

뇌종양의 유무와 종류를 분류하기 위한 CNN 모델 학습에는 대량의 뇌종양 MRI 이미지 데이터세트가 필요하여 Kaggle에서 제공된 Brain Tumor Classification (MRI)[8] 의료영상 데이터를 연구에 이용한다. 데이터 세트에는 정상 소견의 뇌 MRI 의료영상 500개, 뇌하수종(pituitary) 901개, 뇌수막종(meningioma) 997개, 신경교종(glioma) 926개로 총 3264개의 jpeg 형태의 파일로 구성되어 있다. 본 연구에서는 훈련(training) 데이터를 2643개, 검증(validation) 데이터를 294개, 테스트(test) 데이터를 327개로 분류하여 사용한다.



(그림 7) 연구 데이터 세트 샘플
(Figure 7) Sample of research dataset

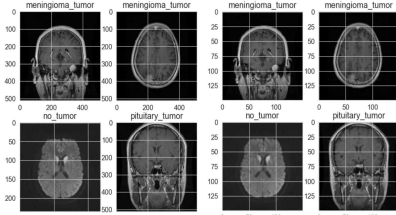


(그림 8) 연구 데이터 세트 구성
(Figure 8) Construction of research dataset

3.2 데이터 전처리

데이터 세트의 각 이미지는 크기가 통일되어있지 않다. 이러한 문제를 해결하기 위해 이미지에 전처리 과정을

진행한다. 전처리를 위해 영상 처리 라이브러리 OpenCV를 통해 받아온 이미지를 `resize` 함수를 이용하여 이미지를 각 모델의 필요 입력 값에 맞게 통일시킨다.



(그림 9) 좌 - `resize` 전, 우 - `resize` 후
(Figure 9) Left - before `resize`, Right - after `resize`

4. 실험

4.1 커스텀 모델 생성

기존 모델과 비교를 위해 커스텀 모델을 다양한 구조로 여러 번 평가하여 높은 성능의 모델로 생성한다.

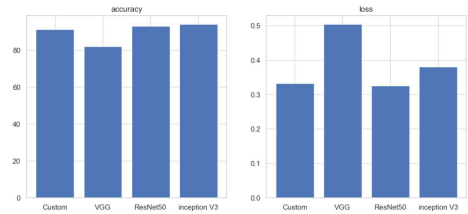
Model: "sequential"		
Layer (type)	Output Shape	Param #
conv2d (Conv2D)	(None, 222, 222, 32)	896
conv2d_1 (Conv2D)	(None, 228, 228, 64)	18496
max_pooling2d (MaxPooling2D)	(None, 110, 110, 64)	0
dropout_1 (Dropout)	(None, 110, 110, 64)	0
conv2d_2 (Conv2D)	(None, 108, 108, 64)	36928
conv2d_3 (Conv2D)	(None, 106, 106, 64)	36928
dropout_2 (Dropout)	(None, 106, 106, 64)	0
max_pooling2d_1 (MaxPooling2D)	(None, 53, 53, 64)	0
dropout_3 (Dropout)	(None, 53, 53, 64)	0
conv2d_4 (Conv2D)	(None, 51, 51, 128)	73856
conv2d_5 (Conv2D)	(None, 49, 49, 128)	147584
conv2d_6 (Conv2D)	(None, 47, 47, 128)	147584
max_pooling2d_2 (MaxPooling2D)	(None, 23, 23, 128)	0
dropout_4 (Dropout)	(None, 23, 23, 128)	0
conv2d_7 (Conv2D)	(None, 21, 21, 128)	147584
conv2d_8 (Conv2D)	(None, 19, 19, 256)	295168
max_pooling2d_3 (MaxPooling2D)	(None, 9, 9, 256)	0
dropout_5 (Dropout)	(None, 9, 9, 256)	0
flatten (Flatten)	(None, 28736)	0
dense_1 (Dense)	(None, 512)	10617344
dense_2 (Dense)	(None, 512)	262656
dropout_6 (Dropout)	(None, 512)	0
dense_3 (Dense)	(None, 4)	2052
Total params: 11,787,076		
Trainable params: 11,787,076		
Non-trainable params: 0		

(그림 10) 커스텀 모델 구조
(Figure 10) Structure of custom model

그림 10은 `sequential` 함수를 이용하여 제작한 커스텀 모델의 구조이다. 3x3 크기의 Convolution 레이어 9개, 2x2 크기의 Pooling 레이어 4개와 Fully-Connected 레이어 3개로 구성되어있고, 활성화 함수로 마지막 출력층의 `softmax` 함수를 제외하고 모두 `ReLU` 함수를 사용한다.

4.2 실험 결과

그림 11은 본 연구의 결과이다. 연구 결과에 따르면 가장 좋은 정확도가 나타난 모델은 94.1%의 정확도를 보인 Inception V3 모델이고, 가장 낮은 정확도가 나타난 모델은 82.2%의 정확도를 보인 VGG-11 모델이었다. Custom 모델과 ResNet50 또한 90% 이상으로 높은 정확도를 보였다.



(그림 11) 모델별 정확도, 손실값
(Figure 11) Accuracy and loss of each model

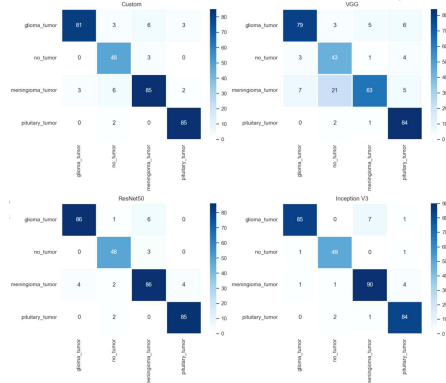
손실량의 경우 ResNet50과 Custom 모델이 비슷한 수치를 찍는 동시에 다른 두 모델에 비해 적은 손실량을 보였으며, VGG-11의 경우 손실량이 가장 높은 결과를 보였다.

		Custom	VGG11	ResNet	Inception V3
accuracy		0.914	0.822	0.932	0.941
f1 Score	glioma	0.92	0.87	0.94	0.94
	no_tumor	0.87	0.72	0.92	0.95
	meningioma	0.89	0.76	0.90	0.93
	pituitary	0.96	0.90	0.97	0.95
precision	glioma	0.96	0.89	0.96	0.98
	no_tumor	0.81	0.62	0.91	0.94
	meningioma	0.90	0.90	0.91	0.92
	pituitary	0.94	0.85	0.96	0.93
recall	glioma	0.87	0.85	0.92	0.91
	no_tumor	0.94	0.84	0.94	0.96
	meningioma	0.89	0.66	0.90	0.94
	pituitary	0.98	0.97	0.98	0.97
loss		0.331	0.503	0.326	0.356

(표 1) 네 가지 모델 성능평가
(Table 1) Performance evaluation of four models

표 1은 네 가지 모델에 대한 성능평가지표에 따른 성능평가를 나타낸 표이다. 전체적인 성능평가를 보더라도 VGG-11 모델은 정확도, f1-score, 정밀도, 재현율 모든 면

에서 다른 세 모델보다 확연히 성능이 떨어지는 것을 볼 수 있었다. 또한, Inception V3 모델은 성능평가 항목에서 다른 모델보다 전체적으로 높은 평가 점수를 보인다.



(그림 12) 모델별 혼동행렬
(Figure 12) Confusion Matrix for each model

그림 12는 각 모델별 혼동행렬을 나타낸 히트맵이다. x축이 예측 값, y축이 실제 값으로, VGG-11의 경우 뇌수막종(meningioma)을 정상조건(no_tumor)으로 예측한 오류가 눈에 띄게 많은 것을 확인할 수 있다.

5. 결론

본 실험에서는 MRI를 통한 뇌종양 진단을 빠르고 정확하게 진단하기 위한 4가지 모델의 자동 분류 시스템을 구현하여 성능을 비교해 보았다. 실험 결과, Inception V3가 94.1%의 정확도를 달성하며 가장 성능이 높게 나왔고, VGG-11모델이 82.2%의 정확도를 달성하며 가장 성능이 낮게 나왔다.

하지만 본 연구에서 VGG-11은 레이어 수가 적은 편에 속하였고, VGG-13, 16, 19 및 그 이상의 모델에서는 어떠한 결과를 확인할 수 있을지 모르며, filter 개수, batch size, epoch, learning rate 등의 하이퍼 파라미터에 따라 다양한 결과를 확인할 수 있을 것이다. 또한, 본 실험에서는 간단한

구조의 커스텀 모델을 사용하여 기존 모델을 뛰어넘지 못하는 성능을 보였지만, 다양하고 복잡한 구조의 모델을 생성함에 따라 더욱 좋은 성능을 보일 수 있다는 커스텀 모델의 발전 가능성을 확인하였다.

참고문헌(Reference)

- [1] Mayo Clinic. Diseases & Conditions : Brain tumor [Internet]. Available: <https://www.mayoclinic.org/diseases-conditions/brain-tumor/symptoms-causes/syc-20350084>
- [2] Jaeyong Kang, Jeonghwan Gwak, "MRI-Based Brain Tumor Classification Using Ensemble of Deep Features and Machine Learning Classifiers", Sensors 2021, Vol. 21, No. 6, March 2021.
- [3] Washington University School of Medicine in Saint Louis(2017, February). Do You Know the Differences Between a CT, MRI and PET Scan?. [Online]. Available: <https://www.mir.wustl.edu/do-you-know-the-differences-between-a-ct-mri-and-pet-scan/>.
- [4] Muhammad Faisal Siddiqui, Ahmed Wasif Reza, Jeevan Kanesan, "An Automated and Intelligent Medical Decision Support System for Brain MRI Scans Classification", PLOS ONE, Vol. 10, No. 8, August 2015.
- [5] Karen Simonyan, Andrew Zisserman, "Very Deep Convolutional Networks for Large-Scale Image Recognition", arXiv:1409.1556 [cs.CV], Sep 2014.
- [6] Kaiming He, Xiangyu Zhang, Shaoqing Ren, Jian Sun , "Deep Residual Learning for Image Recognition", arXiv:1512.03385 [cs.CV], Dec 2015
- [7] Christian Szegedy, Vincent Vanhoucke, Sergey Ioffe, Jon Shlens, Zbigniew Wojna, "Rethinking the Inception Architecture for Computer Vision", arXiv:1512.00567[cs.CV], Dec 2015.

❶ 저 자 소 개 ❶

문 승 재(Seung-jae Moon)

2019년 ~ 현재 한신대학교 IT콘텐츠학과 학사

관심분야 : 딥러닝, 컴퓨터 비전, 자연어처리, 생성형 인공지능, etc.

E-mail : msj0243@gmail.com