

인공 신경망 모델에 따른 의학적 이미지 분석의 효율 비교

이수영[○] 김민하

성균관대학교

ensoo94@gmail.com, prekmh@naver.com

Comparative Analysis on the efficiency of Medical image processing by Artificial Neural network model

Lee Suyeong[○] Kim Minha

Sungkyunkwan University

요 약

최근 인공지능은 빠르게 진화하여 다양한 분야에 적용되고 있다. 특히 의료 분야는 딥 러닝을 활용한 의학적 이미지 처리를 활용할 경우 수많은 환자의 사례와 질병의 증상을 진단하는 데 있어 인적 오류를 배제할 수 있기 때문에 그 필요성이 강조되고 있는 상황이다. 본 논문에서는 기존에 의학적 이미지 분석에 활용되어 온 인공 신경망 중에서도 Partially Connected Deep Network에 해당하는 콘볼루션 신경망(Convolutional Neural network, CNN)의 두 모델인 Large Scale Memory Storage and Retrieval(LAMSTAR) neural network와 Region-CNN(RCNN)의 일종인 DarkNet model을 통해 MRI와 CT 같은 이미지 형태의 환자 기록을 분석하고 결과를 도출하여 비교한다. 이 과정을 통해 어떤 모델이 의료 의사 결정을 내리는 데에 더 적합한지 판단하고자 한다.

1. 서 론

영상 처리에 있어서 인공 신경망은 놀라운 발전을 이뤄왔다. 특히 Partially Connected Deep Networks의 모델 중 하나인 CNN은 다른 네트워크 모델들에 비해 link 수가 상대적으로 적은 신경망 모델로서 이미지나 텍스트와 같은 순차적 데이터를 처리하는 데에 유용하다.[1] CNN은 convolutional mask와 pooling 기법을 이용한 연결 가중치의 계산을 특징으로 가지며, fully connected deep network 보다 복잡성이 낮음에도 불구하고 이미지 인식 및 분류에 있어 속도와 정확성 면에서 더 나은 효율을 가진다.

그로 인해 다양한 분야에서 널리 활용되고 있으며 특히 급속하게 발달하고 있는 의료 분야에서 크게 기여할 수 있을 것으로 기대된다. 빅데이터의 등장과 네트워크 기술의 발전으로 인해 수많은 환자의 사례와 새로운 질병의 증상 등 진단을 내리는 데 필요한 필수적인 정보들이 인간의 정보 처리 속도로는 전부 확인할 수 없을 정도로 급속히 늘어나고 있다. 진료 과정에서 기존의 사례와 의학 지식을 바탕으로 진단과 치료에 관한 수많은 의사결정을 해야 하는 의사에게 인공 신경망을 활용한 CDSS(Clinical Decision Support System)[2]의 지원은 판독과 진단에 있어 발생 가능한 인적 오류를 배제하고 정확한 판단을 내리도록 직접적 혹은 간접적으로 도움을 줄 수 있을 것이다.

환자의 상태를 진단할 때 사용하는 MRIs, CT scans, X-rays는 이런 CDSS를 통해 분석해야 하는 중요한 의학적 이미지들이다. 의학적 이미지의 분석을 통해 척추 손상부터 심장 질환, 암 등 다양한 질병들을 파악할 수 있는데, 그 과정은 매우 까다로우며 오랜 시간이 걸린다. 이때 의학 이미지 분석에 인공 신경망을 효과적으로 적용시킨다면 보다 효율적인 진단이 가능할 것이다.

따라서 이 논문에서는 다양한 CNN 모델 중에서 여러 개의 Self-Organizing-map(SOM)으로 이루어진 LAMSTAR neural

network와 RCNN의 일종인 DarkNet model을 통해 동일한 이미지 집합을 학습시킨 후 분석 효율을 각각 비교, 분석하여 의학 이미지 진단에 적용하였을 때 보다 효율적인 영상 처리 모델이 무엇인지 탐구하고자 한다.

비교를 위한 신경망 모델을 선정할 때는 기존에 의학적 이미지 분석에 이미 활용되고 있는 모델들 중 다른 모델들과 구분되는 강점이 있는지를 중점으로 조사하였다. LAMSTAR의 경우 Kohonen SOM 모듈을 기반으로 학습이 이뤄지는데, 전체 입력 데이터가 아니라 Subwords로 쪼개 후 기존의 형성된 SOM과의 관계와 가중치가 큰 부분만 저장되어 효율적이다. 또한, 가중치와 함께 상벌에 의한 학습을 특징으로 가정으로써 완전한 학습 단계를 갖추지 못한 불완전한 데이터라도 충분한 정확도가 보장되는 결과를 도출해낼 수 있다.[3]

그림 1을 통해 LAMSTAR의 작동 원리와 구조를 확인할 수 있다. 입력 데이터가 존재할 경우 가중치는 다음과 같이 계산되는데, $F(k)$ 에 해당하는 것이 Forgetting Feature이다. Forgetting Feature가 존재함으로 인해 LAMSTAR는 최신 상태로의 갱신이 다른 신경망 모델에 비해 빠르게 이루어 진다.

$$L(k+1) = L(k) - F(k) \forall k (k \text{는 input word})$$

Forgetting Feature는 다음 수식에 따라 값이 조정된다.

$$F(k) = 0$$

이때 각 p번째 input word마다 $F(k) = bL(k)$ (단, $b < 1$)를 만족한다.

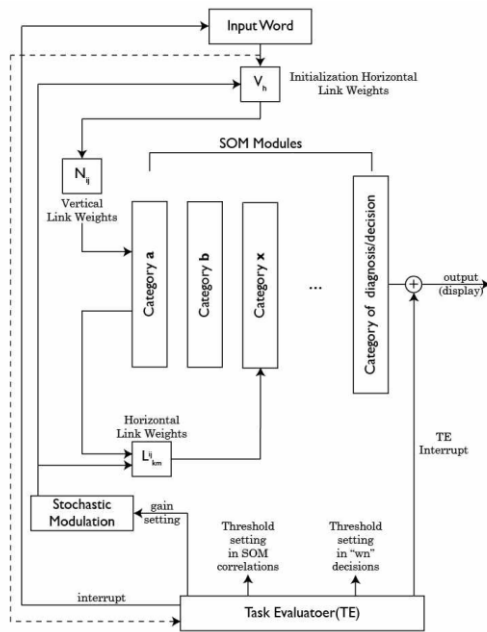


그림 1 General Block Diagram[4] : LAMSTAR network

LAMSTAR와 함께 비교에 쓰인 DarkNet 모델은 GoogLeNet을 기반으로 하는 Pre-trained RCNN이다.[5] DarkNet은 이미지 전체가 아닌 이미지 안의 object를 찾아내는 데에 특화되어 있으며 이런 특징으로 인해 영상에서의 실시간 물체 탐지를 목적으로 개발되었으나, 이미지 분류에도 Fast RCNN[6]등의 타 모델보다 속도와 정확도 면에서 우월한 성능을 보인다. DarkNet은 네 개의 Convolutional 레이어와 두 개의 fully-connected 레이어로 이루어져 있으며 이미지를 입력 받은 후 $S \times S$ 개의 그리드로 쪼갬다. 각각의 그리드는 B 개의

bounding box와 C 개의 class를 통해 예측되며 $S \times S \times (B * 5 + C)$ 의 수식을 통해 얻어지는 결과로 변환된다.[7]

위의 두 모델을 통해 정상인과 뇌출혈 환자의 뇌 CT 이미지를 학습시켜 결과를 얻었으며, 정확한 비교를 통해 동일한 이미지 집합을 이용하였다. 이미지 집합은 정상과 뇌출혈의 두 가지 클래스를 가지며 각 47장으로 이루어져 있다. 학습에는 전문가에 의해 분류된 표본 이미지가 사용되었으며 테스트에는 주어진 이미지 집합을 입력 데이터로 사용하여 올바르게 분류되는지를 판별하였다.

두 모델을 실제로 구현하기 위해 공개되어 있는 오픈소스를 참고하였으며 일부 수정하였다.[8][9]

2. LAMSTAR

LAMSTAR에 학습 이미지가 입력 되면 여러 개의 subword로 분리되어 서로 다른 SOM module에 의해 처리가 된다. 이 때 저장된 연결 가중치에 의해 결정된 일부 SOM 그룹만이 활성화되고 이 과정을 여러 번 반복하면 패턴이 인식되거나 잊혀지게 된다. [10]

뇌출혈 판별을 위한 학습 전에 이미지 상태에 따른 정확도를 알아보고자 이미지를 원본, 수정된 이미지, 완전히 다른 이미지 세 가지로 나누어 학습시켰을 때 이미지와 원본 이미지의 일치율은

다음과 같았다.

표 1 이미지 상태에 따른 정확도 차이

이미지 상태	사용한 이미지	이미지 일치도
원본		107.15
수정된 이미지		58.00
새로운 이미지		13.67

램스타는 연결 가중치에 의해 계산된 SOM module 값을 결과로 경향성에 따른 점수를 도출한다. 명확한 클래스를 갖고 입력된 이미지를 분류하는 것이 아니라, 입력된 사진 중 비슷한 형태를 가진 것들끼리 비슷한 점수를 갖도록 한다. 따라서 원본 이미지에 근접할수록 결과값이 높게 출력된다. 뇌출혈 환자마다 비슷한 양상을 띠고 있는 CT 사진을 사용하면 LAMSTAR를 이용한 뇌출혈 진단이 가능하다고 판단했다.

3. DarkNet

DarkNet 모델을 통해 이미지들을 원하는 클래스로 나누어 학습할 수 있다. 뇌출혈 판별을 위한 학습을 하기 전 이미지의 상태에 따라 이미지 분류의 정확도에 어떤 차이가 생기는지 알아보기 위해 LAMSTAR와 동일하게 세 가지 경우로 나누어 각 경우의 클래스 일치도를 백분율의 형태로 얻었다. 이때는 VOC Challenge에서 이미 범주화가 완료된 충분히 많은 양의 이미지[11]를 이용해 학습된 가중치 파일[12]을 이용해 검사하였으며 별도로 자체적인 학습 과정은 거치지 않았다.

표 2 이미지 상태에 따른 정확도 차이

이미지의 상태	사용한 이미지의 예	클래스 일치도(%)
동일한 이미지		99.47
손상된 이미지		27.91
새로운 이미지		92.52

실험 결과 근접치의 수치는 동일한 이미지, 새로운 이미지, 그리고 손상된 이미지 순으로 높았다. 이를 통해 정확한 학습 데이터를 통해 가중치 결과를 얻은 DarkNet 모델이라면 사용하지 않은 새로운 이미지를 검사했을 때도 높은 정확도를 가지므로 충분한 선학습을 통해 새로운 환자의 뇌출혈 판단이 가능하다는 결론을 내릴 수 있었다.

위의 결론을 바탕으로 뇌출혈 여부의 판단을 위해 정상인과 뇌출혈 환자의 실제 뇌 CT 촬영 이미지를 이용해 DarkNet 모델을 학습시키고 가중치 파일을 얻었다. 이때 클래스는 정상과 뇌출혈의 두 가지로 나누었으며 output은 588 개의 텐서로 인코딩된다. 얻은 가중치 파일에 의해 이미지를 검사할 경우 해당하는 클래스 상자가 출력된 이미지와 클래스 일치도, 그리고 결과를 도출하는 데 걸린 시간을 결과값으로 얻는다. DarkNet의 경우 학습에 걸리는 시간은 반복 횟수를 직접 설정함으로써 이미지의 수와 관계없이 속도를 조절할 수 있었다. 이미지의 수가 적고 반복 횟수가 클수록 높은 정확도를 보였으며 본 연구에서는 수행 환경의 한계로 200으로 설정하였다. 이때 얻은 결과를 LAMSTAR 모델을 통해 얻은 결과값과 비교하여 의학적 이미지 판별의 효율을 비교하였다.

4. 비교

LAMSTAR와 DarkNet을 통해 얻을 수 있는 결과값 중 학습에 소요되는 시간과 클래스 분류의 정확도를 이용해 두 모델의 효율을 비교하였다. 비교 기준의 경우 기존 연구 중 인공 신경망을 통한 신장 암 진단의 효율 비교에 사용된 기준[13]을 일부 차용하였다. 결과는 표 2와 같다.

표 3 LAMSTAR network와 DarkNet의 뇌출혈 진단 성능 비교

	뇌출혈	
	LAMSTAR	DarkNet
Training Time	< 6mins	< 3hours
Test Accuracy	83.2%	89.4%
Negative Specificity	0.82	0.9
Positive Specificity	0.9	0.87

Comments:

1. DarkNet의 경우, Training Time과 Test Time의 차이가 크게 벌어졌으며, Test에는 평균 19s의 시간이 소요되었다.
2. Positive/Negative Specificity는 정상과 뇌출혈 각각에 대해 올바르게 진단한 비율을 의미한다.

5. 결론 및 향후 연구

이 논문에서는 인공 신경망 모델을 이용한 의학적 이미지 분석의 효율을 비교해 보았다. 4.에서 시간과 정확도를 비교한 결과 LAMSTAR와 DarkNet은 각각 다른 강점과 약점을 지니고 있음을 알 수 있었다. LAMSTAR의 경우 완전히 범주 처리 되지 않은 이미지들을 통해서도 충분한 학습만 이루어진다면 빠르게 원하는 결과를 얻을 수 있고 최신 상태로의 갱신이 쉽다는 장점을 가졌으나 결과값이

명시적이지 않은 숫자로 도출되기 때문에 직관적인 이해가 어렵고 경계상의 데이터들의 경우 오진단이 발생할 확률이 DarkNet에 비해 높았다. 이에 반해 DarkNet은 초기 학습에 필요한 이미지가 범주 처리 되어 있어야 하므로 전문가의 도움이 필요하고 높은 정확도를 위해서는 학습에 많은 시간이 소요된다는 단점이 있었으나 한 번 가중치를 얻고 나면 진단을 내리는 데에 걸리는 시간이 짧고 같은 반복 횟수를 가지는 경우 LAMSTAR에 비해 높은 정확도를 지닌다는 장점이 있었다.

인공 신경망이 의학 분야에서 이용되는 이유는 방대한 양의 데이터 처리와 밀접한 관련이 있다. 따라서 본 연구의 목적에 부합하는 모델은 가중치가 빠르게 갱신되는 LAMSTAR라는 결론을 내렸다.

그러나 두 모델 모두 실제로 활용되기에는 아직 한계가 존재한다. 특히 LAMSTAR의 경우 각 SOM의 분류가 이루어지는 과정에 있어 더욱 깊은 이해를 바탕으로 가중치가 비슷한 데이터들의 처리를 개선하는 것이 앞으로 연구해야 할 과제이다.

참 고 문 헌

- [1] Hoo-Chang Shin, Deep Convolutional Neural Networks for Computer-Aided Detection: CNN Architectures, Dataset Characteristics and Transfer Learning, *IEEE transactions on medical imaging*, Vol.35 No.5, pp. 1285-1298, 2016
- [2] Burroni M, Corona R, Dell'Eva G, et al. Melanoma computer-aided diagnosis: reliability and feasibility Study, *Clinical Cancer Research*, Vol.10 No.6, pp. 1078-0432, 2014
- [3] Daniel Graupe, Large memory storage and retrieval(LAMSTAR) network, US5920852 A, filed April 30, 1997, and issued July 6, 1999
- [4] Kordylewski. H, Graupe. D, Liu. K, *IEEE Journal of Biomedical and Health Informatics*, Vol. 5 No.3, pp. 202-209, 2001
- [5] D. Mishkin. Models accuracy on imagenet 2012 val. <https://github.com/BVLC/caffe/wiki/Models-accuracy-on-ImageNet-2012-val>. Accessed:2015-10-2
- [6] R. B. Girshick. Fast R-CNN. CoRR, abs/1504.08083, 2015
- [7] Joseph Redmon, You Only Look Once: Unified, Real-Time Object Detection, *Allen Institute for AI*, 2016
- [8] mareurs, LAMSTAR(python) open source, [Internet]. Available at: <https://github.com/mareurs/lamstar.git> [Accessed 26 January 2013]
- [9] pjreddie, DarkNet open source, [Internet]. Available at: <https://github.com/pjreddie/darknet.git> [Accessed 4 June 2017]
- [10] Daniel Graupe, (1997) *Principles of artificial neural networks*, World Scientific, Singapore
- [11] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, A. C. Berg, and L. Fei-Fei. ImageNet Large Scale Visual Recognition Challenge. *International Journal of Computer Vision (IJCV)*, 2015
- [12] pjreddie, Trained weight file, [Internet]. Available at: <https://pjreddie.com/darknet/imagenet/> [Accessed 4 June 2017]
- [13] C. S. Niederberger et al., "A neural computational model of stone recurrence after ESWL," in *Int. Eng. Appl. Neural Networks Conf.*, 1996, pp.423-426.