

스마트폰 사용 보행자의 포즈 인식을 위한 인공 신경망 모델링

권혜민*, 손가은*, 황지윤*, 조동섭**

*이화여자대학교 컴퓨터공학과

e-mail : hmkwon1025@ewhain.net, kategood@ewhain.net, gpffjwldbs@ewhain.net

**이화여자대학교 컴퓨터공학과

e-mail : dscho@ewha.ac.kr

An Artificial Neural Network Modeling to Recognize Pedestrian's Pose using Smartphone

Hyemin Kwon*, Ga-Eun Son*, Jiyun Hwang*, Dongsub Cho**

*Dept of Computer Science and Engineering, Ewha Womans University

**Dept of Computer Science and Engineering, Ewha Womans University

요 약

보행자의 특이 포즈 인식을 위해 최적화된 인공 신경망 모델은 보행자의 스마트폰 사용 여부를 판단하여 안전한 보행환경을 조성하고 효율적인 보행자 및 포즈 인식 소프트웨어를 제공한다. LeNet-4 를 기반으로 모델의 파라미터를 수정하거나 Layer 구조를 변형시키며 모델링 과정을 거치고, 학습된 모델의 성능 분석을 통해 포즈 인식 모델을 최적화하였다. 그 결과 LeNet-5 구조를 수정한 모델이 가장 좋은 성능을 보였고 이를 보완하여 보행자의 스마트폰 사용 방지를 위한 알림 시스템에 활용할 수 있을 것으로 기대된다.

1. 서론

최근 스마트폰 사용 보행자의 교통사고율이 증가함에 따라 보행자 포즈 인식을 통한 경고 시스템의 필요성이 대두되었다[1]. 해외에서는 보행 중 스마트폰 사용에 법적 제재를 가하는 등의 조치를 취하고 있으며 국내에서는 횡단보도에 스마트폰 정지선을 표시하거나 바닥 신호등을 시범운영하는 등 교통·통신 기술적으로 안전한 보행 환경 조성을 위한 노력이 이어지고 있다. 또한 IoT 와 통신 기술을 활용하여 보행 중 스마트폰 사용을 방지하는 앱이 출시되고, 스마트폰에 ‘보행 중 모드’를 도입하는 방안이 제시되었다. 이처럼 전세계적으로 보행자의 스마트폰 사용 규제가 강화되고 있는 추세이다.

이러한 흐름에 발맞춰 언제 어디에서 나타날지 모르는 스마트폰 사용 보행자를 감지하기 위해 보행자의 스마트폰 사용 포즈 인식에 최적화된 LeNet 기반의 인공 신경망 모델을 구축하고자 한다. 스마트폰 사용 보행자가 감지되었을 때 이들에게 스마트폰 사용을 경고하는 알림을 준다면 보다 효율적이고 안전한 보행환경을 조성할 수 있다. 더 나아가 스마트폰 사용 보행자 포즈 인식 모델은 효율적인 보행자 포즈 인식 및 분석을 위한 소프트웨어를 제공할 수 있다.

2. 관련 연구

Convolution Neural Network (CNN)는 큰 크기의 이미지 데이터를 다룰 수 있게 하고 인식률을 높이는 등 컴퓨터 비전과 인공지능 분야에서 큰 성과를

거두었다[2][3]. 하지만 아직 이미지 학습을 기반으로 포즈 인식 문제를 해결하기에는 한계가 있다[4].

LeNet 은 최초의 CNN 모델로, 우표와 수표의 필기체 인식을 위해 탄생하였다. 4 개의 layer 로 이루어진 LeNet-1 은 1.7%의 오차율을 보이며 이미지 인식에 성공적인 결과를 보였다[5].

이후 그림 1 과 같이 더 큰 크기의 이미지 데이터를 input 으로 사용하고, 기존 LeNet-1 의 마지막에 fully connected layer 를 추가로 도입한 LeNet-4 가 고안되었다.

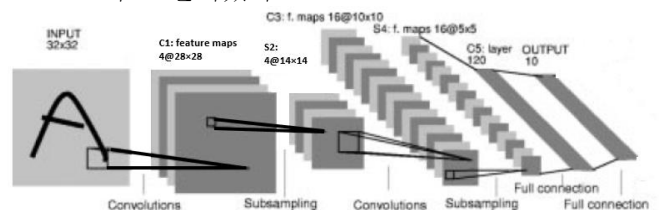


그림 1. LeNet-4 모델의 구조

LeNet-4 모델은 1.1%의 오차율을 보이며 더욱 효과적인 모델로 자리 잡았다[6]. 이후에도 LeNet-4 모델에 Hidden Layer 를 추가하거나(LeNet-5), 세 가지 LeNet-4 모델의 결과를 조합한(Boosted LeNet-4) 모델의 기반이 되어 현재까지 유용하게 사용되고 있다 [7].

이후 AlexNet 과 GoogLeNet 이 등장해 더욱 실용적인 이미지 인식 모델이 제시되었고, 최근에는 물체 감지에 뛰어난 강점을 지닌 ResNet 이 보편적으로 사용되고 있다[8].

3. 보행자 포즈 인식을 위한 모델링

보행자의 특이 포즈 중, 스마트폰을 사용하는 포즈를 인식하는 모델의 최적화를 위해 스마트폰을 사용하는 보행자와 사용하지 않는 보행자를 이원적으로 구분하는 LeNet-4 기반 모델을 설계한다.

이후 LeNet-4 기반 모델의 다양한 파라미터를 표 1과 같이 수정하여 각 모델의 성능을 분석한다.

표 1. 성능 평가를 위해 수정된 LeNet 기반 모델의 파라미터

실험	고정 파라미터	수정 파라미터	수정 값
1	Epochs 활성화 함수 모델 구조	Node size*	500
			1000
			2000
			5000
2	Node size 활성화 함수 모델 구조	Epochs	50
			100
			200
			500
			1000
3	Epochs Node size 모델 구조	활성화 함수	ReLU
			LeakyReLU
			ELU
			tanh
4	Epochs Node size 활성화 함수	모델 구조	LeNet-1
			LeNet-4
			LeNet-5**
			LeNet-5 수정***

* Dense Layer에서의 Node size를 의미

** 기존의 LeNet-5 모델(그림 2)을 그림 3으로 수정한 모델 구조를 사용

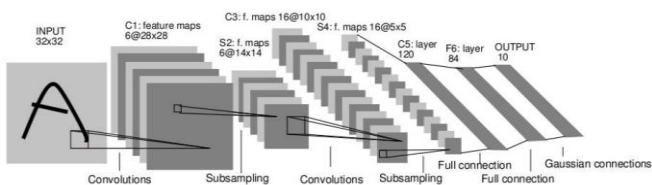


그림 2. LeNet-5 모델 구조[6]

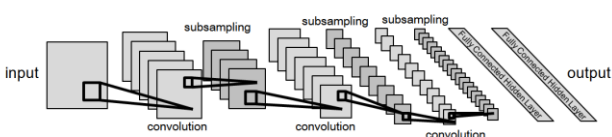


그림 3. LeNet-5 수정 모델

4. 보행자 포즈 인식 모델 실험을 위한 데이터

스마트폰을 사용하는 보행자와 사용하지 않는 보행자의 포즈를 구분하기 위해 그림 4와 같이 스마트폰을 사용하는 보행자의 좌측 방향 사진 1,100장과 스마트폰 미사용 보행자 사진 1,100장을 수집한 후, 모두 224*224 pixel 크기로 동일하게 처리하여 학습하였다.



그림 4. 스마트폰 사용 보행자(좌)와 스마트폰 미사용 보행자(우)의 사진 데이터

앞서 제시한 LeNet 모델을 기반으로 Epochs 값, 활성화 함수의 종류, Dense Layer의 Node size 값, 모델 구조를 수정하여 학습을 위한 인공 신경망으로 사용하였다.

학습에 사용되지 않은 스마트폰 사용 보행자 사진 150장과 미사용 보행자 사진 150장을 테스트 데이터로 사용하여 네 가지 지표(Precision, Recall, Accuracy, F1 score)로 모델의 정확도를 평가하였다[9].

5. 실험 결과 분석 및 검토

5-1. Dense Layer의 Node size 수정

표 2. Node size 수정에 따른 모델 성능 분석

Node size	Precision	Recall	Accuracy	F1 score
500	0.985	0.433	0.713	0.602
1000	1	0.313	0.657	0.477
2000	0.958	0.307	0.647	0.465
5000	0.957	0.447	0.714	0.609

Dense Layer의 Node size 값 수정 결과, Node size는 모델 성능에 크게 영향을 끼치지 않는 것으로 분석되었다. 따라서 가장 보편적으로 사용되는 500을 앞으로 학습시킬 모델의 Node size로 설정하였다.

5-2. Epochs 수정

표 3. Epochs 수정에 따른 모델 성능 분석

Epochs	Precision	Recall	Accuracy	F1 score
50	1	0.367	0.683	0.537
100	0.969	0.413	0.7	0.579
200	0.985	0.433	0.713	0.602
500	0.941	0.213	0.6	0.348
1000	0.903	0.433	0.693	0.586

Epochs 수정 결과 Epochs 200으로 학습된 모델이 가장 좋은 성능을 보였다.

5-3. 활성화 함수 수정

표 4. 활성화 함수 수정에 따른 모델 성능 분석

활성화 함수	Precision	Recall	Accuracy	F1 score
ReLU	1	0.313	0.657	0.477
LeakyReLU	0.942	0.327	0.653	0.485
ELU	1	0.353	0.677	0.522
tanh	0.96	0.32	0.653	0.48

4 개의 활성화 함수 중 ELU 의 성능이 가장 뛰어난 것으로 분석되었다.

5-4. 모델 구조 수정

표 5. 모델 구조 수정에 따른 모델 성능 분석

모델 구조	Precision	Recall	Accuracy	F1 score
LeNet-1	1	0.373	0.687	0.544
LeNet-4	1	0.353	0.677	0.522
LeNet-5*	0.927	0.34	0.657	0.498
LeNet-5 수정**	1	0.473	0.737	0.643

*Dense Layer 들의 Node size 가 각각 1000, 500 인 LeNet-5 모델

**LeNet-5 에 filter 개수가 100 인 convolution layer 를 추가한 모델

모델 구조 수정 결과 LeNet-5 수정 모델이 가장 높은 성능 지표를 보였으며, 실험 데이터셋에 대한 인식률은 47.3%이다.

6. 결론

보행자의 특이 포즈 인식을 위한 보다 나은 인공 신경망 모델을 구현하였다. 실험적으로 다양하게 LeNet-4 모델을 수정하여 성능을 평가한 결과, LeNet-5 수정 모델이 스마트폰 사용 보행자 인식에 가장 적합한 모델임을 도출하였다. 하지만 추후 이 모델의 인식률을 더 높이고, 실시간으로 보행자 포즈 인식이 가능하도록 개선하는 연구가 필요하다.

이러한 인공 신경망을 기반으로 한 보행자 포즈 인식 모델을 실시간으로 적용하면 횡단보도, 공공장소 등에서의 사고를 예방하기 위한 시스템을 구현할 수 있을 것으로 기대된다. 이는 안전한 보행 환경 조성을 위한 국제적인 규제 강화 추세에 알맞은 소프트웨어를 마련함과 동시에 자율 주행 자동차 운행에 필요한 보행자 인식 소프트웨어로써 유용하게 활용될 수 있을 것이다.

참고문헌

- [1]어진이, "보행 중 스마트폰 사용에 따른 사고 예방을 위한 사용자 경험 연구", 디자인전공석사학위 논문, 이화여자대학교, 2018.
- [2]Xiabing Liu, et al., "3D head pose estimation with convolutional neural network trained on synthetic images," 2016 IEEE International

- Conference on Image Processing (ICIP), 2016.
- [3]V. Golovko, M. Egor and A. Brich, "A Simple Shallow Convolutional Neural Network for Accurate Handwritten Digit Classification," Pattern Recognition and Information Processing, Vol. 13, pp. 77-85, 2016.
- [4]K. Luberg, "Human Body Poses Recognition Using Neural Network with Class Based Data Augmentation," Master's Thesis of University of TARTU, 2018.
- [5]Y. LeCun, et al., "Learning algorithms for classification: A comparison on handwritten digit recognition," Neural networks: the statistical mechanics perspective, pp. 261-276, 1995.
- [6]Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," Proc. IEEE, Vol. 86, No. 11, pp. 2278-2324, 1998.
- [7]Xu, Bing, et al., "Empirical evaluation of rectified activations in convolutional network," arXiv preprint arXiv:1505.00853, 2015.
- [8]C Szegedy, S Ioffe, et al., "Inception-v4, inception-resnet and the impact of residual connections on learning," Thirty-First AAAI Conference on Artificial Intelligence (AAAI-17), 2017.
- [9]Markus Junker, Rainer Hoch, and Andreas Dengel, "On the evaluation of document analysis components by Recall, Precision, and Accuracy," 99 Proceedings of the Fifth International Conference on Document Analysis and Recognition, 1999.

"본 연구는 과학기술정보통신부 및
정보통신기획평가원의 SW 중심대학사업의
수행결과로 추진되었음"(2019-0-01247)