

가속도 센서 데이터 기반의 행동 인식 모델 성능 향상 기법

Improving Performance of Human Action Recognition on Accelerometer Data

남 정 우*, 김 진 현*

Jung-Woo Nam*, Jin-Heon Kim*

Abstract

With a widespread of sensor-rich mobile devices, the analysis of human activities becomes more general and simpler than ever before. In this paper, we propose two deep neural networks that efficiently and accurately perform human activity recognition (HAR) using tri-axial accelerometers. In combination with powerful modern deep learning techniques like batch normalization and LSTM networks, our model outperforms baseline approaches and establishes state-of-the-art results on WISDM dataset.

요 약

스마트 모바일 장치의 확산은 인간의 일상 행동 분석을 보다 일반적이고 간단하게 만들었다. 행동 분석은 이미 본인 인증, 감시, 건강 관리 등 많은 분야에서 사용 중이고 그 유용성이 증명되었다. 본 논문에서는 스마트폰의 가속도 센서 신호를 사용하여 효율적이고 정확하게 행동 인식을 수행하는 합성곱 신경망(모델 A)과 순환 신경망까지 적용한(모델 B) 심층 신경망 모델을 제시한다. 모델 A는 batch normalization과 같은 단순한 기법만 적용해도 이전의 결과보다 더 작은 모델로 더 높은 성능을 달성할 수 있다는 것을 보인다. 모델 B는 시계열 데이터 모델링에 주로 사용되는 LSTM 레이어를 추가하여 예측 정확도를 더욱 높일 수 있음을 보인다. 이 모델은 29명의 피실험자를 대상으로 수집한 벤치마크 데이터 세트에서 종합 예측 정확도 97.16%(모델 A), 99.50%(모델 B)를 달성했다.

Key words : Human activity recognition, Deep learning, Convolutional Neural Networks, Recurrent Neural Networks, Time-series classification

1. 서론

인간 행동 인식(Human Activity Recognition, HAR) 분야는 스마트 장치가 확산되고 스마트 장치의 가속도 센서, 자이로스코프 등의 사용 및 활용이 쉬

워지면서 많은 주목을 받는 연구 주제 중 하나이다 [1]. 센서의 저렴한 가격, 적은 전력 소비, 센서 데이터의 실시간 전송과 컴퓨터 비전, 기계 학습, 인공지능 기술의 발전도 인기 상승에 한몫하고 있다. 인간 행동 인식은 걷기, 뛰기, 앉기, 서기, 운전,

* Dept. of Electronic & Computer Engineering, Seokyeong University

★ Corresponding author

E-mail : jinheon@skuniv.ac.kr, Tel : +82-2-940-7747

※ Acknowledgment

This research is supported by Ministry of Culture, Sports and Tourism(MCST) and Korea Creative Content Agency(KOCCA) in the Culture Technology(CT) Research & Development Program 2019.

Manuscript received Jun. 1, 2020; revised Jun. 14, 2020; accepted Jun. 17, 2020.

This is an Open-Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/3.0>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

문 열기, 이상 행동과 같은, 사용자가 수행하고 있는 물리적 행동을 인식하는 작업이다. 데이터는 스마트폰이나 웨어러블 장치의 센서, 가속도계, 혹은 동영상 프레임이나 사진을 통해서 수집되기도 한다[2]. 이런 행동 인식 기술은 인증, 감시, 제스처 기반의 컨트롤 등의 다양한 응용을 통해 이미 그 유용성이 증명되었다. 대부분의 행동 인식 시스템은 다양한 종류의 카메라를 통해 얻은 영상 정보를 사용하지만, 최근 연구들은 가속도계 역시 동일한 기능을 제공할 수 있다는 것을 보여주었다[3]. 또한, 카메라의 사생활 침해 논란이 불거짐에 따라 가속도계는 더욱 주목을 받고 있다. 영상보다 데이터가 작아서 처리하기 쉽고 인식하려는 대상으로부터 직접 측정된 신호라는 장점도 있다.

최근 딥러닝 기반의 접근 방식은 많은 연구 분야에서 최고 성능을 달성하고 있다. 그중에서도 합성곱 신경망(Convolutional Neural Networks)과 순환 신경망(Recurrent Neural Networks)은 영상과 음성 분야에서 매우 효율적임을 보여줬다. 그러나 영상과 음성에서만 효율적인 것은 아니다. 합성곱 신경망의 강력한 특징 추출 능력과 순환 신경망의 시간 지역성 학습 능력은 인간 행동 인식을 위한 1차원 가속도계 데이터에서도 유효하다[4].

본 연구는 스마트폰의 가속도 센서 데이터에서 자동으로 특징을 추출하여 인간의 행동을 인식하는 인공지능 모델을 제시한다. 이전 연구[5][6]에서도 행동 인식에 합성곱 신경망을 적용하는 시도는 있었지만 심층 신경망에서 사용되는 최신 기법을 적용한 사례는 많지 않았다. 또한, 시계열 데이터가 가진 시간적 관련성을 제대로 다루지 않았다. 따라서 본 연구에서는 최근 심층 신경망에서 자주 사용되는 batch normalization(BN) 기법[7]과 LSTM[8]을 인간 행동 인식 모델에 적용하여 기존에는 분류가 어렵다고 여겨졌던 행동을 높은 정확도로 분류하는 모델을 제시하고, 이 모델의 성능을 실험을 통해 증명한다.

II. 본론

여기서는 합성곱 신경망과 순환 신경망을 이용해 3축 가속도 센서 데이터에서 사람 행동 인식 방법에 대해 기술한다. 먼저 본 연구에서 사용한 데이터셋과 데이터 전처리 방법에 대해 설명할 것이다.

그리고 가속도 센서 데이터를 행동으로 분류하는 두 가지 신경망 구조(모델 A, 모델 B)를 소개할 것이다.

1. WISDM Activity Prediction Dataset

모델의 성능을 실험하고 평가할 데이터 세트는 WISDM Activity Prediction Dataset(WISDM AR)[9]을 사용한다. WISDM 프로젝트의 목표는 스마트폰과 같은 모바일 장치에서 센서 데이터를 수집하며 연구 주제를 탐색하고, 유용한 애플리케이션을 구축하는 것이다[10]. 프로젝트에서 다루는 연구 주제는 가속도계 데이터 기반으로 사용자를 식별하고 사용자의 ID를 검증하는 행동 기반 생체 인증과 사용자가 수행하고 있는 행동을 식별하는 행동 인식 등이 있다. WISDM AR은 이 프로젝트를 통해 수집된 통제된 실험실 환경의 데이터 세트이다. 29명의 피실험자가 스마트폰을 휴대하면서 걷기, 조깅, 계단 올라가기, 계단 내려가기, 앉기, 서기와 같은 일상적인 활동을 수행하며 스마트폰의 3축(x, y, z) 가속도계 데이터를 수집했으며 모든 샘플은 6가지 행동 중 하나의 라벨이 붙어있다. WISDM AR의 3축 가속도계의 값은 (-20, 20)의 실숫값을 가지고 10이라는 수치는 $1G = 9.81m/s^2$ 이며 0은 가속도가 없다는 것을 나타낸다. 모든 경우에서 가속도계 데이터는 50ms마다 수집되었으므로 초당 20개의 샘플을 갖는다. 총 샘플 개수는 1,098,207개이며 각 활동이 데이터 세트에서 차지하는 개수와 비율은 표 1에 나와 있다.

Table 1. Data distribution of WISDM dataset.

표 1. WISDM 데이터 세트의 클래스별 데이터 분포

Activity	Examples	Percentage
Walking	424,000	38.6%
Jogging	342,177	31.2%
Upstairs	122,869	11.2%
Downstairs	100,427	9.1%
Sitting	59,939	5.5%
Standing	48,395	4.4%

2. 데이터 전처리 & 학습 데이터 생성

가공되지 않은 시계열 가속도계 데이터는 곧바로 신경망 분류 알고리즘으로 적용될 수 없다. 신경망의 입력으로 사용되기 위해선 미가공된 데이터를 신경망이 학습하기 편한 레코드 단위로 변환해야

한다. 그러기 위해서 먼저 (-20, 20) 값을 가지는 가속도 센서값들을 20으로 나눠서 (-1, 1)로 정규화했다. 그리고 2초씩 건너뛰며 4초 길이의 데이터를 추출하는 슬라이딩 윈도우 방식을 적용했다. (50% 겹침) 그러므로 하나의 레코드는 4초 * 20Hz = 80개의 데이터 샘플을 가진다. 이러한 방식으로 미가공된 시계열 가속도계 데이터로부터 26,520개의 레코드를 생성했다.

Table 2. Data distribution after preprocessing.

표 2 전처리 후의 데이터 세트의 클래스별 분포

Activity	Examples	Percentage
Walking	10,506	39.6%
Jogging	8,174	30.8%
Upstairs	2,881	10.9%
Downstairs	2,323	8.8%
Sitting	1,463	5.5%
Standing	1,173	4.4%

표 2를 보면 활동별 레코드 개수에 불균형이 있음을 알 수 있다. 이것은 미가공된 데이터의 클래스별 불균형에서 기인한 것으로 데이터 세트의 이러한 불균형은 모델을 평가할 때 결과를 왜곡할 가능성이 있다. 클래스 불균형을 다루는 방법은 클래스 가중치 조절, 오버/언더 샘플링 등의 방법이 있지만 본 연구에서는 단순히 전체 레코드에서 가장 적은 수를 차지하는 클래스(Standing)와 그 레코드 개수(1173개)를 찾고 각 클래스에서 의도적으로 그 개수만큼만 레코드를 추출하는 방법을 택했다.

위에 기술한 데이터 전처리를 거친 후의 레코드 수는 7,038개(1173*6)이다. 이것을 학습 데이터 80%,

검증 데이터 20%로 나눠서 신경망 알고리즘의 학습, 검증에 사용했다.

3. 모델 A (1D ConvNet)

모델 A의 구조는 같은 연구 주제의 A. Ignatov [6]의 모델에서 영감을 받았다. 그림 1은 모델 A의 구조를 보여준다. A. Ignatov의 모델과 다른 점은 1) 더 적은 필터 개수와 더 작은 필터 크기, 2) 풀링 레이어를 제거하고 최근 심층 신경망에서 자주 사용되는 BN[7] 레이어를 추가한 것이다. 더 적은 필터 개수와 더 작은 필터 크기는 모델 크기와 연산량을 감소시켰고 BN 레이어 추가는 더 빠르고 안정성 높은 학습을 가능하게 한다.

하나의 합성곱 블록은 1*3 크기의 필터를 가지는 합성곱 레이어, BN 레이어, 그리고 ReLU[11] 활성화 함수로 이루어져 있다. 모델 A는 두 개의 합성곱 블록을 가지는데 첫 번째 블록은 16개의 특징맵, 두 번째 블록은 32개의 특징맵을 생성한다. 이 32개의 특징맵은 평탄화를 거쳐서 6개의 동작으로 분류하는 Fully-connected (FC) 레이어로 연결된다. 마지막으로, FC 레이어의 출력은 Softmax 함수를 거쳐 각 클래스에 대한 확률로 표현된다.

4. 모델 B (1D ConvNet + LSTM)

비록, 모델 A가 합성곱 신경망을 이용해 사람 행동을 인식하는데 사용될 수 있지만, 여전히 시계열 데이터라는 특성을 적극적으로 활용하지 못하고 있다. 최근 연구에서 순환 신경망(Recurrent Neural Network)은 동영상이나 음성처럼 시간이나 순서에 의존적인 데이터를 다루는 동적 시스템에서 뛰어

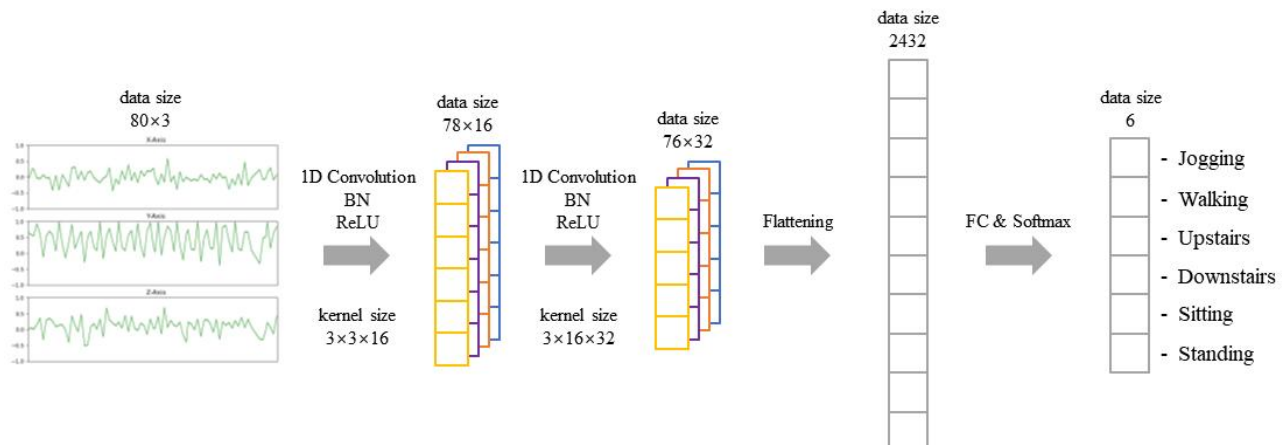


Fig. 1. The architecture of Model A (1D ConvNet).

그림 1. 모델 A의 구조 (1차원 합성곱 신경망)

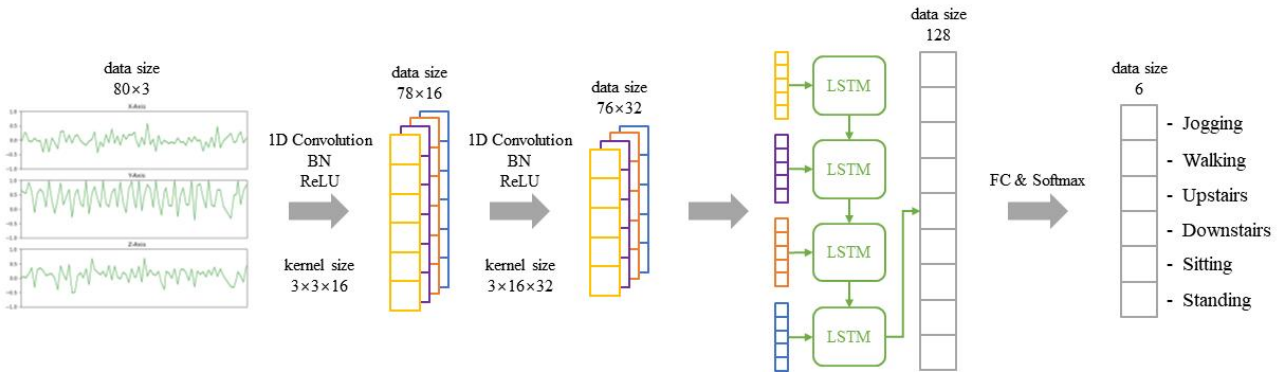


Fig. 2. The architecture of Model B (1D ConvNet + LSTM).
 그림 2. 모델 B의 구조 (1차원 합성곱 신경망 + LSTM 블록)

난 효율을 보여주었다[12]. R. Fernandez-Fernandez et al.은 합성곱 신경망과 순환 신경망 중 하나인 LSTM[8] 구조를 이용해 Quick, Draw! 데이터 세트를 분류하는 모델을 제시했다[13]. 모델 B는 그 모델에서 영감을 받아 가속도계 데이터를 분류하는 합성곱 + LSTM 구조를 제안한다.

그림 2는 모델 B의 구조를 보여준다. 모델 B의 합성곱 블록 부분은 모델 A와 같다. 모델 B는 합성곱 블록의 특징맵에서 평탄화를 거치지 않고 LSTM 레이어로 연결된다. LSTM 레이어의 출력은 1차원이므로 바로 FC 레이어와 Softmax 함수로 연결된다.

두 모델 모두 신경망 모델 구현, 학습 및 검증은 Tensorflow v2[15]으로 구현하였다. 학습 시, 커널 초기화는 Xavier 초기화[16] 기법을 사용하였고 손실 함수는 cross-entropy를 적용하였다. 그라디언트 갱신에는 Adam[17]을 사용하였고 learning rate는 0.0005로 설정하였다. 검증 데이터에서 정확도가 더 개선되지 않을 때까지 학습을 진행하였다.

III. 실험 결과 및 고찰

전처리와 학습 데이터 생성 부분에서 얻은 1,408

개의 검증 데이터로 학습된 모델의 분류 정확도를 평가했다. 본 실험의 행동 인식 실험 요약 결과를 표 3에 나타냈다. 이 표는 모델 A와 모델 B의 예측 정확도를 다른 연구 결과와 함께 클래스별로 나타내고 있다.

그림 3의 오차 행렬을 보면 클래스별 분류 결과를 조금 더 자세하게 알 수 있다. 모델 A의 경우, 계단 올라가기와 계단 내려가기 두 클래스 사이의 혼돈이 가장 많았고, 이것은 다른 연구 결과들과 일치한다. 하지만 여전히 예측 정확도는 각각 94.89%와 92.34%로 다른 연구에 비해서 높은 수치를 확보했다. LSTM을 적용한 모델 B의 경우, 클래스 간 혼돈이 거의 없이 모든 경우에서 99% 이상의 예측 정확도를 보였다.

실험 결과, 본 연구에서 제안한 모델이 가장 높은 수준의 정확도를 달성했음을 확인했다. 제안한 두 모델 중 특히 LSTM 레이어를 추가한 모델 B는 종합 정확도 99.5%를 기록했다. 합성곱 레이어의 추가는 이전엔 분류가 몹시 어렵다고 여겨졌던 계단 올라가기와 계단 내려가기를 99% 이상의 높은 정확도로 분류해냈다[9].

Table 3. Evaluation of WISDM prediction accuracy(%) between different methods (The best method is highlighted in bold).
 표 3. WISDM 데이터 세트의 예측 정확도(%) 비교 평가 (가장 좋은 성능을 보인 알고리즘은 볼드 표시)

	MLP[9]	Dropout[14]	RF[14]	CNN[6]	Ours(Model A)	Ours(Model B)
Walking	91.7	93.03	91.77	98.50	98.72	99.57
Jogging	98.3	94.64	97.54	97.87	100	100
Upstairs	61.5	53.45	38.26	72.22	94.89	99.15
Downstairs	44.3	63.81	45.19	87.00	92.34	99.57
Sitting	95.0	88.96	89.05	82.63	98.30	98.72
Standing	91.9	86.26	84.03	93.33	100	100
Overall	91.7	80.02	74.31	93.32	97.16	99.50

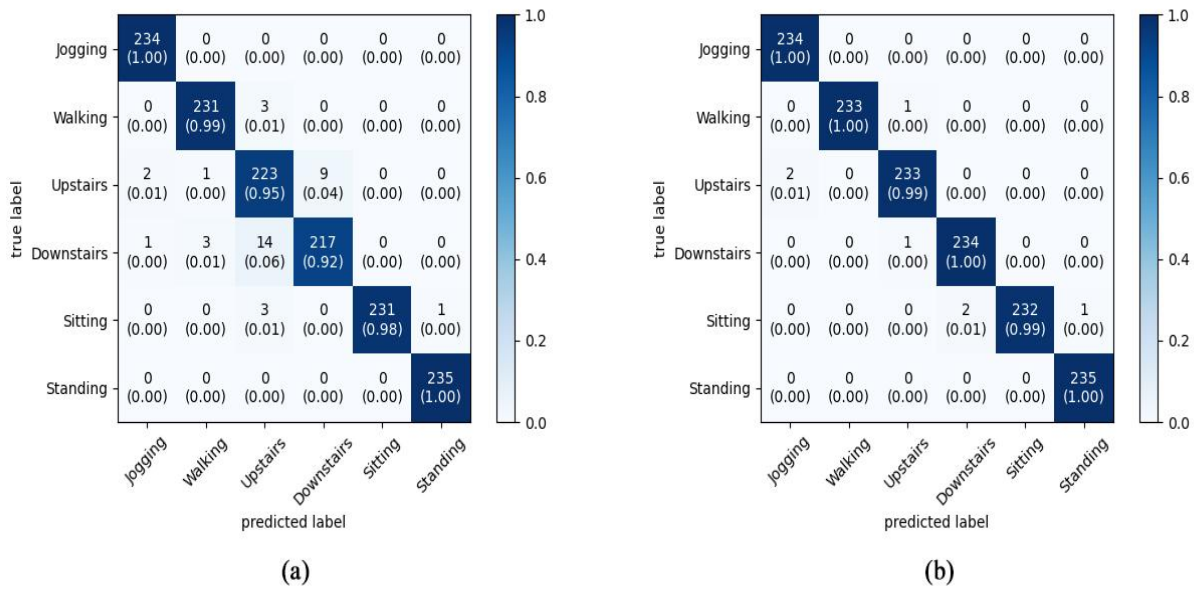


Fig. 3. Confusion matrix on validation dataset (a) Model A (1D ConvNet), (b) Model B (1D ConvNet + LSTM).
 그림 3. 검증 데이터의 오차 행렬 (a) 모델 A (1D ConvNet), (b) 모델 B (1D ConvNet + LSTM)

IV. 결론

본 연구에서는 가속도 센서 데이터를 이용해서 인간의 행동을 인식하는 새로운 AI 모델을 제시하였다. 제안한 방법 중, 모델 A는 기존의 합성곱 신경망에서 파라미터 수를 줄여 연산량을 낮추었고 batch normalization 레이어를 추가하는 것만으로도 계단 오르내리기 행동의 예측 정확도를 개선할 수 있다는 것을 보여줬다. 모델 B는 여기에 LSTM 레이어를 추가하여 다변수 시계열이라는 가속도 센서 데이터의 특성을 적극적으로 반영하였고, 결과적으로 WISDM 데이터 세트에서 통합 예측 정확도 99.50%를 달성하였다.

References

[1] Ó. D. Lara and M. A. Labrador, "A survey on human activity recognition using wearable sensors," *IEEE Commun. Surv. Tutorials*, vol.15, no.3, pp.1192-1209, 2013.
 DOI: 10.1109/SURV.2012.110112.00192

[2] C. Jobanputra, J. Bavishi, and N. Doshi, "Human activity recognition: A survey," in *Procedia Computer Science*, vol.155, pp.698-703, 2019.
 DOI: 10.1016/j.procs.2019.08.100

[3] A. Avci, S. Bosch, M. Marin-Perianu, R. Marin-

Perianu, and P. Havinga, "Activity recognition using inertial sensing for healthcare, wellbeing and sports applications: A survey," in *23th International Conference on Architecture of Computing Systems 2010, ARCS 2010 - Workshop Proceedings*, pp.167-176, 2010. DOI: 10.1.1.604.8265

[4] I. Hwang, G. Cha, and S. Oh, "Multi-modal human action recognition using deep neural networks fusing image and inertial sensor data," in *IEEE International Conference on Multisensor Fusion and Integration for Intelligent Systems*, vol.2017-Novem, pp.278-283, 2017.
 DOI: 10.1109/MFI.2017.8170441

[5] C. A. Ronao and S. B. Cho, "Human activity recognition with smartphone sensors using deep learning neural networks," *Expert Syst. Appl.*, vol.59, pp.235-244, 2016.
 DOI: 10.1016/j.eswa.2016.04.032

[6] A. Ignatov, "Real-time human activity recognition from accelerometer data using Convolutional Neural Networks," *Appl. Soft Comput. J.*, vol.62, pp.915-922, 2018. DOI: 10.1016/j.asoc.2017.09.027

[7] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," in *32nd International Conference on Machine Learning, ICML 2015*,

vol.1, pp.448–456, 2015.

DOI: 10.5555/3045118.3045167

[8] S. Hochreiter and J. Schmidhuber, “Long Short-Term Memory,” *Neural Comput.*, vol.9, no.8, pp.1735–1780, 1997.

[9] [1] J. R. Kwapisz, G. M. Weiss, and S. A. Moore, “Activity recognition using cell phone accelerometers,” *ACM SIGKDD Explorations Newsletter*, vol.12, no.2, pp.74–82, 2011.

DOI: 10.1145/1964897.1964918

[10] G. Weiss, “WISDM Project,” <https://storm.cis.fordham.edu/~gweiss/wisdm/>

[11] V. Nair and G. E. Hinton, “Rectified linear units improve Restricted Boltzmann machines,” in *ICML 2010 – Proceedings, 27th International Conference on Machine Learning*, pp.807–814, 2010. DOI: 10.5555/3104322.3104425

[12] K. Smagulova and A. P. James, “A survey on LSTM memristive neural network architectures and applications,” *Eur. Phys. J. Spec. Top.* 2019 22810, vol.228, no.10, pp.2313–2324, 2019.

DOI: 10.1140/epjst/e2019-900046-x

[13] R. Fernandez-Fernandez, J. G. Victores, D. Estevez, and C. Balaguer, “Quick, Stat!: A Statistical Analysis of the Quick, Draw! Dataset,” *CoRR* 2019, abs/1907.06417, 2019. DOI: 10.11128/arep.58

[14] M. Abadi et al., “TensorFlow: Large-Scale Machine Learning on Heterogeneous Distributed Systems,” Preliminary White Paper, 9, 2015.

[15] X. Glorot and Y. Bengio, “Understanding the difficulty of training deep feedforward neural networks,” in *Journal of Machine Learning Research*, vol.9, pp.249–256, 2010.

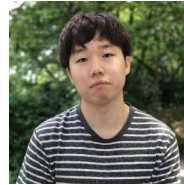
[16] D. P. Kingma and J. L. Ba, “Adam: A method for stochastic optimization,” in *3rd International Conference on Learning Representations, ICLR 2015 – Conference Track Proceedings*, 2015.

[17] B. Kolosnjaji and C. Eckert, “Neural network-based user-independent physical activity recognition for mobile devices,” in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol.9375 LNCS, pp.378–

386, 2015. DOI: 10.1007/978-3-319-24834-9_44

BIOGRAPHY

Jung-Woo Nam (Member)



2019 : BS degree in Computer Engineering, Seokyeong University.
2019~today : MS degree course in Electronic & Computer Engineering, Seokyeong Univ.

Research interests : Computer Vision, Image Signal Processing, Deep Learning

Jin-Heon Kim (Member)



1982 : BS degree in Electrical Engineering, Korea University
1983 : Research Engineer, OPC
1984 : MS degree in Electrical Engineering, Korea University
1984 : Senior Researcher, Samsung Advanced Institute of Technology

1989 : FAE, ZyMOS Korea Branch

1990 : PhD degree in Electrical Engineering, Korea University

1995~today : Professor, Dept. of Computer Engineering, Seokyeong Univ.

Research interests : Computer Vision, Image Signal Processing, Image/Video System