

WiFi 신호를 활용한 CNN 기반 사람 행동 인식 시스템 설계 및 구현

Design and Implementation of CNN-Based Human Activity Recognition System using WiFi Signals

정유신 · 정윤호*

한국항공대학교 항공전자정보공학부

You-shin Chung · Yunho Jung*

School of Electronics and Information Engineering, Korea Aerospace University, Gyeonggi-do, 10540, Korea

[요 약]

기존의 사람 행동 인식 시스템은 웨어러블 센서, 카메라와 같은 장치를 통해 행동을 탐지하였다. 그러나, 이와 같은 방법들은 추가적인 장치와 비용이 요구되고, 특히 카메라 장치의 경우 사생활 침해 문제가 발생한다. 이미 설치되어 있는 WiFi 신호를 사용하면 해당 문제를 해결할 수 있다는 장점이 있다. 본 논문에서는 WiFi 신호의 채널 상태 정보를 활용한 CNN 기반 사람 행동 인식 시스템을 제안하고, 가속 하드웨어 구조 설계 및 구현 결과를 제시한다. 해당 시스템은 실내 공간에서 학습 중 나타날 수 있는 네 가지 행동에 대해 정의하였고, 그에 대한 WiFi의 채널 상태 정보를 CNN으로 분류하여 평균 정확도는 91.86%를 보였다. 또한, 가속화를 위해 CNN 분류기에서 연산량이 가장 많은 완전 연결 계층에 대한 가속 하드웨어 구조 설계 결과를 제시하였다. FPGA 디바이스 상에서 성능 평가 결과, 기존 software 기반 시스템 대비 4.28배 빠른 연산 시간을 보임을 확인하였다.

[Abstract]

Existing human activity recognition systems detect activities through devices such as wearable sensors and cameras. However, these methods require additional devices and costs, especially for cameras, which cause privacy issue. Using WiFi signals that are already installed can solve this problem. In this paper, we propose a CNN-based human activity recognition system using channel state information of WiFi signals, and present results of designing and implementing accelerated hardware structures. The system defined four possible behaviors during studying in indoor environments, and classified the channel state information of WiFi using convolutional neural network (CNN), showing and average accuracy of 91.86%. In addition, for acceleration, we present the results of an accelerated hardware structure design for fully connected layer with the highest computation volume on CNN classifiers. As a result of performance evaluation on FPGA device, it showed 4.28 times faster calculation time than software-based system.

Key word : Accelerator, CNN, FPGA, Human Activity Recognition, WiFi signal.

<https://doi.org/10.12673/jant.2021.25.4.299>



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.

Received 30 September 2014; Revised 21 October 2014
Accepted (Publication) 7 October 2014 (30 October 2014)

*Corresponding Author; Yunho Jung

Tel: +82-2-300-0133

E-mail: yjung@kau.ac.kr

I. 서론

최근 HCI (human-computer interaction), 제스처 인식, 감시 등과 같은 다양한 응용 분야에 적용할 수 있는 컴퓨팅 및 감지 기술 분야의 급속한 발전으로 인해 사람 행동 인식 (HAR; human activity recognition)이 주목받고 있다 [1]. 특히, 실내 공간에서 사람의 움직임과 행동을 모니터링하는 것을 목표로 하는 수많은 애플리케이션으로 인해 실내 공간에서 사용되는 HAR 시스템의 관심이 높아졌다.

일반적으로 사람의 움직임을 탐지하는 방법으로 웨어러블 센서와 카메라를 많이 사용한다 [2]. 센서 기반 행동 인식은 높은 정확도를 보인다는 장점이 있지만 [3], 추가적인 장치와 구매를 위한 비용이 필요하다는 단점이 있다. 카메라 기반 행동 인식은 미세한 움직임까지 포착한다는 장점이 있지만, 이 또한 추가적인 장치를 위한 비용이 발생하고 빛이 없는 곳에서는 사용할 수가 없다. 무엇보다도 카메라 기반 행동 인식은 사생활 침해 문제가 발생한다는 큰 단점이 있다.

이러한 문제를 해결하기 위해 최근 무선 통신을 위해 이미 설치되어 있는 WiFi 신호를 행동 인식에 이용하는 연구가 진행되고 있다 [4], [5]. WiFi 신호를 사용한다면 공공장소는 물론 가정에도 WiFi가 설치되어 있어서 추가적인 설비와 비용이 필요하지 않고 사생활 침해 문제도 발생하지 않는다는 장점이 있다.

WiFi 신호를 이용한 HAR 시스템에서 인식과 분류를 위해 SVM (support vector machine), kNN (k-nearest neighbor)과 같은 여러 기계학습 (machine learning) 기법을 사용한다 [6], [7]. 그러나 이와 같은 얇은 학습(shallow learning) 기법들은 효율적으로 입력 데이터의 특징을 추출하고 선택하는 데 한계가 있다. 이러한 한계를 극복함으로써 인식의 정확성을 높이기 위해 딥러닝(deep learning)을 이용한다.

딥러닝 방식 중 CNN (convolutional neural network)은 기존의 신경망과 비교하여 각 계층의 입력력 형상을 유지하고 입력 데이터의 공간 정보를 유지한 채 학습을 할 수 있다는 장점을 갖고 있어 다양한 분야에서 사용되고 있다 [8]. 그러나, CNN은 연산량이 많아 속도가 느린 단점이 있다 [9]. 이에, 본 논문은 CNN의 속도 개선을 위해 CNN 분류기에서 연산량이 가장 많은 완전 연결 계층에 대한 가속 하드웨어 구조를 포함한 WiFi 신호 기반 HAR 시스템을 제안한다.

본 논문의 II장에서는 제안된 HAR 시스템에 사용되는 CSI (channel state information) 및 CNN에 관해 서술하며, III장에서는 CSI 추출 결과와 제안된 HAR 시스템의 성능 평가 결과, 가속 하드웨어 구조 설계 결과를 제시한다. 끝으로 IV장에서 본 논문의 결론을 맺는다.

II. Backgrounds

2-1 CSI

CSI란 송수신 과정에서 직교주파수분할다중화 (OFDM; orthogonal frequency division multiplexing)의 부 반송파별 채널 상태를 나타내며, 진폭과 위상으로 표현된다. WiFi 채널은 OFDM 변조에 의해 여러 개의 부 반송파로 나누어진다 [10]. OFDM은 한 채널의 송신 신호를 여러 개의 직교성을 갖는 부 반송파로 나누어 동시에 다중화 전송을 하는 기술로, IEEE 802.11a/g/n/ac 표준에서 변조를 위해 사용된다. 다중 안테나를 사용하는 MIMO-OFDM 시스템에서 각 부 반송파에 대해 WiFi 채널은 다음과 같이 나타낼 수 있다.

$$Y_i = H_i X_i + N_i \quad i \in \{1, \dots, S\} \quad (1)$$

식 (1) 에서 X_i 와 Y_i 는 각각 S개의 부 반송파 중 i번째 부 반송파에 대한 M_T 차원 송신 벡터와 M_R 차원 수신 벡터이다. 여기서 M_T 은 송신 안테나 수이고 M_R 은 수신 안테나 수이다. 또한, H_i 는 $M_T \times M_R$ 차원 CSI 행렬, N_i 은 M_R 차원 잡음 벡터를 나타낸다. 행렬 H_i 는 다음과 같이 나타낼 수 있다.

$$H_i = \begin{bmatrix} h_{11} & h_{12} & \dots & h_{1M_R} \\ h_{21} & h_{22} & \dots & h_{2M_R} \\ \vdots & \vdots & \ddots & \vdots \\ h_{M_T1} & h_{M_T2} & \dots & h_{M_TM_R} \end{bmatrix} \quad (2)$$

수신기는 송신기에서 보내는 긴 훈련 심볼을 수신하면, 송신 벡터 X_i 와 수신 벡터 Y_i 를 이용하여 CSI 행렬인 H_i 를 추정한다.

$$\hat{H}_i = \frac{Y_i}{X_i} \quad (3)$$

OFDM 시스템에서 하나의 부 반송파에 대한 CSI는 다음과 같이 나타낼 수 있다.

$$h = |h|e^{j\angle h} \quad (4)$$

여기서 $|h|$ 는 진폭이고 $\angle h$ 는 위상을 나타낸다.

CSI의 진폭과 위상은 반사, 산란 및 회절과 같은 다중 경로 영향으로 인해 영향을 받게 된다. 따라서 송신기와 수신기 사이에 사람의 움직임이 발생하면 채널 상태에 영향을 미치게 되며 이로 인해 CSI의 진폭과 위상이 변하게 된다.

2-2 CNN

CNN은 합성곱 신경망으로 불리며, 이미지/영상 인식 [11], [12], 음성 인식 등 다양한 분야에서 사용되고 있는 신경망 구조

이다. 1998년 Lecun의 LeNet-5 알고리즘 [13]에 의해 최초로 제시되었다. CNN은 입력층, 합성곱 계층 (convolution layer), 풀링 계층 (pooling layer), 완전 연결 계층 (FCL; fully connected layer), 출력층으로 구성된다. 합성곱 계층과 풀링 계층은 입력 데이터의 주요 특성 벡터를 추출하므로 특징추출계층이라 불리며, 완전 연결 계층에서는 클래스를 분류하므로 분류계층이라고 불린다.

합성곱 계층에서는 합성곱 필터 (convolutional filter)가 입력 데이터의 모든 영역에 대해 컨볼루션 연산을 진행한다. 컨볼루션 연산을 통해 추출한 이미지의 특성 맵 (feature map)을 다음 층으로 전달한다. 풀링 계층은 합성곱 계층에서 추출된 여러 특징 중 가장 중요한 특징을 강조하고 특성 맵의 차원을 다운 샘플링한다. 풀링 연산은 대상 영역에서 최댓값을 추출하는 최대 풀링 (max pooling)과 대상 영역에서 평균을 반환하는 평균 풀링 (average pooling)이 있다. 마지막으로 완전 연결 계층은 특징들을 기반으로 class를 분류한다.

III. 제안된 HAR 시스템 설계

3-1 시스템 구조도

그림 1과 같이 제안된 HAR 시스템의 CNN 분류기는 Linux 802.11n CSI Tool [14]을 통해 추출한 raw CSI 데이터의 진폭 값을 입력으로 사용한다. 진폭 값은 raw CSI 데이터에서 절댓값을 취하여 얻을 수 있다. 구해진 진폭 값을 스펙트로그램 형태로 변환 후 CNN 분류기에 입력한다.

스펙트로그램 형태의 진폭 데이터는 제안된 CNN 분류기를 통해 학습되며, 분류기의 모델 구조는 그림 2와 같이 특징을 추출하고 강조하기 위한 3개의 합성곱 계층과 풀링 계층, 분류를 위한 1개의 완전 연결 계층으로 구성된다. 합성곱 계층을 통해 추출된 특징 데이터는 최종적으로 완전 연결 계층으로 입력되고, 가장 큰 값을 갖는 class를 출력함으로써 분류 과정이 완료된다.

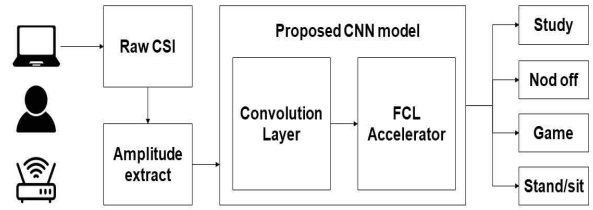


그림 1. 제안된 HAR 시스템 구조도
Fig. 1. The system architecture of proposed HAR system.

3-2 CSI 추출

제안된 HAR 시스템은 그림 3과 같이 학습 중 나타날 수 있는 네 종류의 행동을 정의하여 실험하였다. 실험이 진행된 실내 환경은 그림 4와 같다. WiFi AP와 40cm 떨어진 부분에 앉아 정의한 네 가지 행동을 10초 동안 실험하였다. 조는 경우는 10초 동안 3번 상체를 숙였다 일어났고, 일어났다 앉는 경우는 10초 동안 3번 의자에서 일어났다 앉았다. 게임을 하는 경우는 스마트폰으로 격한 터치를 하는 게임을 한다는 가정하에 책상에 팔꿈치를 댄 상태에서 10초 동안 10번 팔을 앞뒤로 움직였다.

각 행동의 CSI를 추출하기 위해 송신기와 수신기를 실내 환경에 설치하였다. 실험에 사용된 WiFi AP는 ipTIME 사의 N704BCM으로 주파수 채널은 2.4GHz, 채널 대역폭은 20MHz를 사용하였다. 수신기는 Intel Ultimate N WiFi 5300 NIC를 탑재하고, 리눅스 우분투 16.04 LTS 버전이 설치된 노트북을 활용하였다. WiFi 신호를 수신하기 위해, 수신기는 WiFi AP에 초당 33개의 802.11n 패킷을 ping하고 각 패킷의 CSI를 추출한다. CSI 추출은 Linux 802.11n CSI tool을 사용하여 진행되었다. 해당 툴은 각 패킷에 대해 20MHz WiFi 채널의 56개의 부 반송파 중 30개의 부 반송파에 대한 CSI 데이터만을 추출한다.

툴을 통해 추출한 CSI의 raw 데이터에서 MATLAB을 이용하여 CSI 진폭 값을 얻은 뒤, CNN 분류기 입력을 위해 30x30 크기의 스펙트로그램을 생성한다. 그림 5는 10초 (30패킷) 동안 각 행동에 대한 30개의 부 반송파의 CSI 진폭 값을 표현한 스펙트로그램이다. 행동별로 다른 패턴이 나타나며 행동 횟수에 따라 주기성이 관찰되었다.

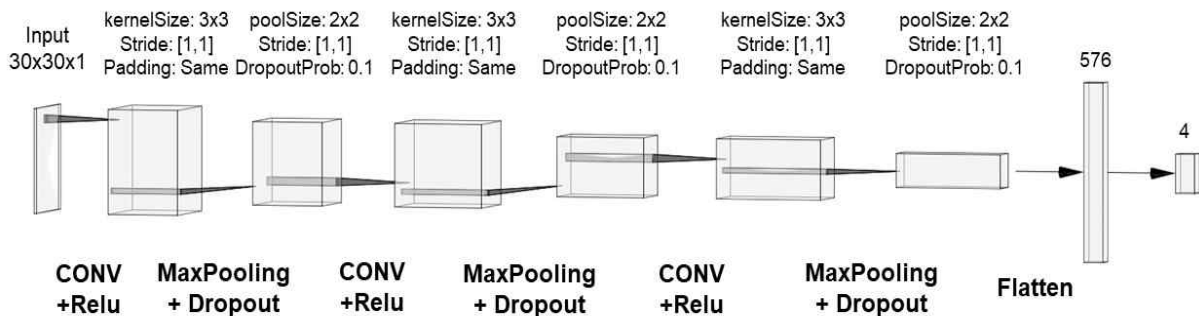


그림 2. 제안된 CNN 분류기 모델 구조
Fig. 2. Model structure of proposed CNN classifier.

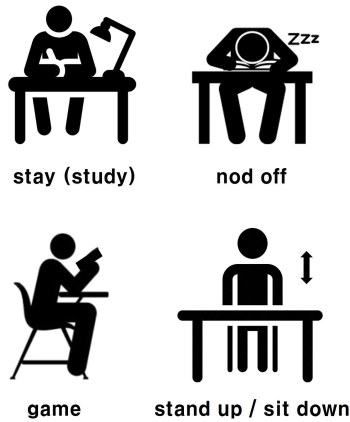
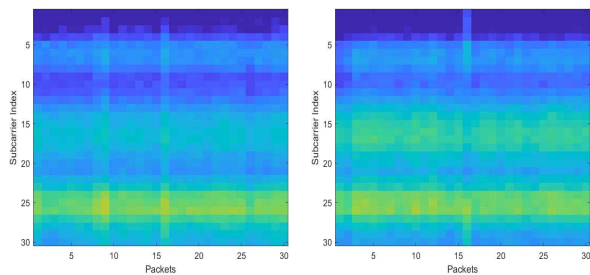


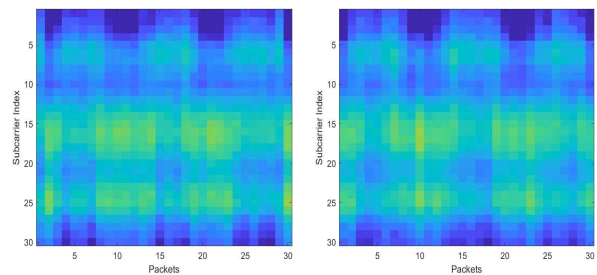
그림 3. 학습 중 나타날 수 있는 4가지 행동
 Fig. 3. Four human activities during studying.



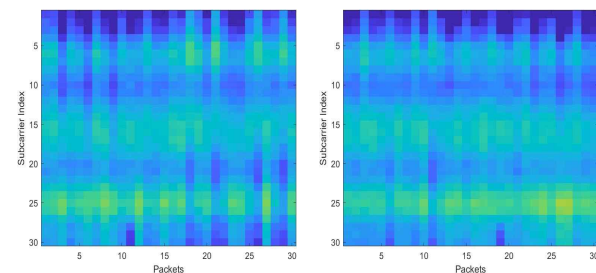
그림 4. 실험 환경
 Fig. 4. Experimental setup.



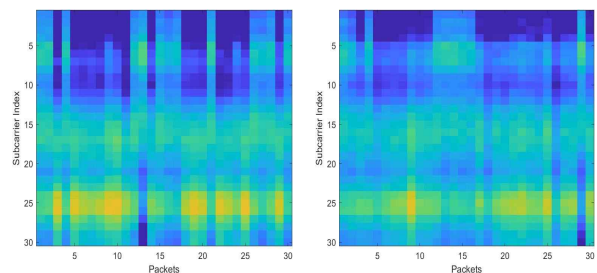
(a) Stay (Study).



(b) Nod off.



(c) Game.



(d) Stand up/Sit down.

그림 5. CSI 진폭 스펙트로그램
 Fig. 5. CSI amplitude spectrogram.

3-3 성능 평가 결과

본 논문에서 사용한 CNN 분류기의 모델 구조는 그림 2와 같다. 해당 모델을 선정하기 위해 여러 모델을 정의하여 성능과 복잡도를 비교하였다. 성능은 정확도로, 복잡도는 전체 파라미터 수를 바탕으로 비교하였다. 성능 평가를 위한 데이터 셋은 가만히 있는 경우 (공부하는 경우) 441개, 조는 경우 604개, 게임을 하는 경우 520개, 일어났다 앉는 경우 391개로 총 1,956개를 사용하였고, 학습을 위해 사용된 데이터는 1,562개, 검증을 위한 데이터는 394개를 사용하였다. 학습은 텐서플로우 (tensorflow)를 통해 이루어졌고, categorical crossentropy loss function과 NAdam optimizer를 사용하였으며, learning rate는

0.001, batch size는 25, epoch는 100으로 진행하였다.

여러 모델을 학습한 결과, 표 1과 같이 정확도와 파라미터 수가 관찰되었다. 이 중, 정확도가 90% 이상이면서 복잡도가 가장 낮은 3개의 합성곱 층과 분류를 위한 1개의 완전 연결 계층으로 구성된 모델을 최종 모델로 선정하였다. 해당 모델의 입력 값은 30x30으로 합성곱 계층에서 크기가 3x3인 합성곱 필터를 적용하여 30x30x16 특성 맵을 출력한다. 합성곱 결과에 크기가 2x2인 최대 풀링을 적용하여 15x15x16 특성 맵을 출력한 뒤 오버피팅 (overfitting) 방지를 위해 확률이 0.1인 드롭아웃을 적용한다. 이와 같은 과정을 세 번 반복한다. 마지막 특징 맵은 3x3x64가 되며 이는 완전 연결 계층을 거쳐 분류된다. 학습 결과, 평균 정확도 91.86%를 보였다.

표 1. CNN 모델 구조별 정확도와 파라미터 수

Table. 1. Accuracy and number of parameters by CNN architecture.

	2 CONV		3 CONV	
	Accuracy	Parameter	Accuracy	Parameter
1 FCL	87.28	11,024	91.86	25,488
2 FCL	88.04	205,968	92.37	97,424
3 FCL	89.06	213,904	92.11	105,360
4 FCL	90.08	215,824	91.35	107,620
	4 CONV		5 CONV	
	Accuracy	Parameter	Accuracy	Parameter
1 FCL	90.59	97,424	92.62	392,848
2 FCL	92.88	113,808	90.59	425,104
3 FCL	91.86	121,744	91.35	433,040
4 FCL	91.35	123,664	90.33	434,960

3-4 가속 하드웨어 설계

제안된 HAR 시스템을 하드웨어로 구현 시, 연산 시간을 줄이기 위해 CNN 분류기에서 연산량이 가장 많은 완전 연결 계층에 대한 가속 하드웨어 구조 설계를 진행하였다. 그림 6은 설계된 CNN 가속 하드웨어와 검증에 위한 Xilinx PYNQ-Z1 FPGA 디바이스 기반 SoC 플랫폼의 구조도 [15]로, H/W IP와 ZYNQ Processor 간 AMBA 버스 통신을 위한 AXI 인터페이스와 설계한 완전 연결 계층 연산 가속기로 구성된다.

완전 연결 계층 연산 가속기의 RTL 설계는 Verilog-HDL를 이용하였다. FPGA 구현 결과, 표 2와 같이 해당 디바이스 기준 14.49% LUT를 사용하고 최대 107.44MHz의 동작 주파수에서 연산 가능함을 확인하였다. 144개의 샘플 데이터에 대해 가속 하드웨어 기반 HAR 시스템은 95.46ms로 기존 software 기반 시스템 대비 4.28배 빠른 연산 시간을 보였다.

IV. 결론

본 논문에서는 WiFi 신호의 채널 상태 정보 (CSI)와 CNN을 이용한 사람 행동 인식 시스템을 제안하였다. 책상에서 학습한다는 상황을 가정하여 학습 중 나타날 수 있는 네 가지 행동에 대해 정의하였고, 각 행동에 대해 CSI가 특정한 패턴을 보임을 확인하였다. 이를 CNN을 사용하여 분류하였고, 평균 정확도는 91.86%를 보였다. 또한, 가속화를 위해 CNN 분류기에서 연산량이 가장 많은 완전 연결 계층에 대한 가속 하드웨어 구조 설계 결과를 제시하였고, 기존 software 기반 시스템 대비 4.28배 빠른 연산 시간을 보임을 확인하였다.

제안된 HAR 시스템은 CNN에서 완전 연결 계층만 가속하였으므로 추후 합성곱 계층을 포함한 CNN 가속기 설계와 Quantization을 통한 경량화를 연구할 계획이며, 이를 통해 WiFi 신호를 행동 인식에 이용하는 것을 발전시킬 계획이다.

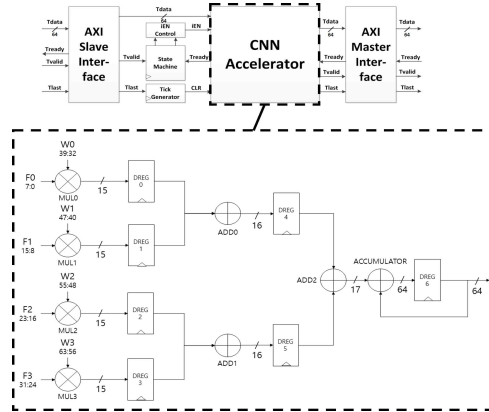


그림 6. 제안된 완전 연결 계층 가속기 블록도
Fig. 6. Block diagram of the proposed fully connected layer accelerator.

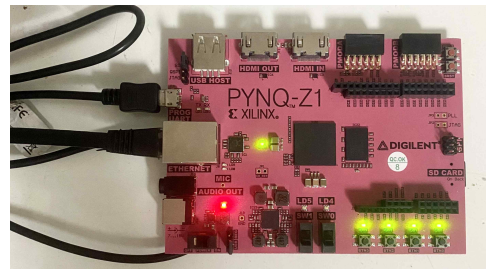


그림 7. FPGA 기반 실험 환경
Fig. 7. FPGA-based experimental setup.

표 2. 제안된 HAR 시스템의 검증 결과

Table. 2. Verification results of the proposed HAR system.

Parameter	Value
Target FPGA	XC7Z020-1CLG400C
Operating frequency	107.44MHz
LUT	14.49%
HW Execution Time	0.09546s
SW Execution Time	0.409s

Acknowledgments

본 논문은 2021년도 정부 (과학기술정보통신부)의 재원으로 정보통신기획평가원의 지원을 받아 수행되었으며 (No. 2019-0-00056, 2020-0-00201), CAD tool은 IDEC에 의해 지원되었음.

References

[1] A. T. Campbell et al., "The rise of people-centric sensing,"

- IEEE Internet Computing*, Vol. 12, No. 4, pp. 12-21, Jul-Aug. 2008.
- [2] L. M. Dang, K. B. Min, H. Wang, M. J. Piran, C. H. Lee, H. J. Moon, "Sensor-based and vision-based human activity recognition: a comprehensive survey," *Pattern Recognition*, Vol. 108, 2020.
- [3] O. Politi, I. Mporas, V. Megalooikonomou, "Human motion detection in daily activity tasks using wearable sensors," *2014 22nd European Signal Processing Conference (EUSIPCO)*, Lisbon, Portugal, pp. 2315-2319, 2014.
- [4] Y. Ma, G. Zhou, S. Wang, "WiFi sensing with channel state information: a survey," *ACM Computing. Surveys*, Vol. 52, No. 3, pp. 1-36, Jul. 2019.
- [5] S. Yousefi, H. Narui, S. Dayal, S. Ermon, S. Valaee, "A survey on behavior recognition using WiFi channel state information," *IEEE Communications Magazine*, Vol. 55, No. 10, pp. 98-104, Oct. 2017.
- [6] Y. Wang, K. Wu, L. M. Ni, "WiFall: device-free fall detection by wireless networks," *IEEE Transactions on Mobile Computing*, Vol. 16, No. 2, pp. 581-594, Feb. 2017.
- [7] J. Yang, H. Zou, H. Jiang, L. Xie, "Device-free occupant activity sensing using WiFi-enabled IoT devices for smart homes," *IEEE Internet of Things Journal*, Vol. 5, No. 5, pp. 3991-4002, Oct. 2018.
- [8] J. Y. Seo, *Deep Learning Tensorflow Textbook*, gilbut, pp. 135-136, 2021.
- [9] A. Khan, A. Sohail, U. Zah hoora, A. S. Qureshi, "A survey of the recent architectures of deep convolutional neural networks," *Artificial Intelligence Review*, Vol. 53, pp. 5455-5516, 2020.
- [10] E. H. Kim, D. D. Kim, Y. J. Suh, "Convolutional neural network based gesture recognition using WiFi channel state information," *Korea Institute of Communications and Information Sciences*, Gangwon, Korea, pp. 989-990, Feb. 2020.
- [11] H. Choi, J. Ko, "Deep learning-based automatic wrinkles segmentation on microscope skin images for skin diagnosis," *Journal of Advanced Navigation Technology*, Vol. 24, No. 2, pp. 148-154, Apr. 2020.
- [12] G. M. Kim, S. J. Pa, S. J. Hwang, H. Y. Kim, J. H. Baek, "Deep learning based pine nut detection in UAV aerial video," *Journal of Advanced Navigation Technology*, Vol. 25, No. 1, pp. 115-123, Feb. 2021
- [13] Y. Lecun, L. Bottou, Y. Bengio, P. Haffner, "Gradient-based learning applied to document recognition," *Proceedings of the IEEE*, Vol. 86, No. 11, pp. 2278-2324, Nov. 1998.
- [14] D. Halperin, W. Hu, A. Sheth, D. Wetherall, "Tool release: gathering 802.11n traces with channel state information," *ACM SIGCOMM Computer Communication Review*, Vol. 41, No. 1, pp. 53, Jan. 2011.
- [15] S. H. Oh, C. H. Bae, S. R. Kim, J. C. Cho, Y. H. Jung, "Design and implementation of CNN-based HMI system using doppler radar and voice sensor," *Journal of IKEEE*, Vol. 24, No. 3, pp. 777-782, Sep. 2020.



정 유 신 (You-Shin Chung)

2017년 3월 ~ 현재 : 한국항공대학교 항공전자정보공학부 학사
※ 관심분야 : 무선 통신 시스템, 딥러닝



정 윤 호 (Yunho Jung)

1998년 2월 : 연세대학교 전자공학과 (공학사)
2000년 2월 : 연세대학교 전기전자공학과 (공학석사)
2005년 2월 : 연세대학교 전기전자공학과 (공학박사)
2005년 ~ 2007년 : 삼성전자 책임연구원
2007년 ~ 2008년 : 연세대학교 연구교수
2008년 ~ 현재 : 한국항공대학교 항공전자정보공학부 교수
※ 관심분야 : 무선 통신 시스템, 항공통신 시스템, 레이더 시스템, SoC 설계