

## 전환 동작 인식을 위한 깊이별 분리 합성곱 기반 FCN 모델

이유경, 손창식, 강원석\*  
대구경북과학기술원 지능형로봇연구부

yklee@dgist.ac.kr, changsikson@dgist.ac.kr, \*wskang@dgist.ac.kr

## An FCN Model based on Depthwise Seperable Convolution for Transitional Activity Recognition

Lee Yoo Kyung, Son Chang Sik, Kang Won Seok\*  
DGIST, Division of Intelligent Robot

### 요 약

본 논문에서는 센서 기반 인간 동작 인식 분야에서 경량성과 성능을 동시에 고려한 FCN 기반 모델을 제안한다. 제안 모델은 기존 모델의 합성곱 계층을 깊이별 분리 합성곱으로 대체하여 파라미터 수를 78% 감소시켰다. HAPT 공개 데이터셋을 활용한 실험 결과, 제안한 모델은 기존 FCN 대비 정확도 2.46%, 매크로 F1 0.48%, 가중치 평균 F1 2.18% 개선하였다. 특히 전환 동작 인식에서 높은 정확도를 보여주었으며, 이는 전환 인식이 중요한 인간 동작 인식 응용 분야 및 모바일 환경에서의 적용 가능성을 시사한다.

### I. 서 론

최근 웨어러블 센서를 활용한 인간 동작 인식(human activity recognition, HAR) 기술은 재활, 낙상 감지 등 다양한 헬스케어 및 스마트 기기 응용 분야에서 중요성이 높아지고 있다[1]. 특히 다양한 센서로부터 획득한 시계열 데이터에서, 각 시점 단위로 신체 활동을 예측하는 dense 예측 방식이 주목받고 있다. 이 방식은 하나의 시퀀스에 여러 동작 클래스가 포함된 경우나, 짧은 시간 내 급격히 전환되는 동작을 보다 정밀하게 포착하는 데 효과적이다. 최근에는 Fully Convolutional Network(FCN) 기반 모델이 dense 예측에서 좋은 성능을 보여준 바 있다[2]. 이러한 dense 예측 모델은 실시간 처리 및 모바일 환경에서도 활용 가능성이 높지만 이를 위해서는 연산량과 메모리 사용을 줄인 경량 모델이 요구된다[3]. 본 논문에서는 기존 FCN 구조[2]에 깊이별 분리(depthwise seperable) 합성곱[4]을 적용하여 모델 경량화를 수행하고, 짧은 전환 동작과 같은 움직임의 보다 정확하게 포착할 수 있음을 보인다.

### II. 본론

#### 2.1 제안 모델

제안된 모델은 기존 FCN[2] 구조를 기반으로 한다. 기존 FCN 은 6 개의 합성곱 계층과 6 개의 최대 풀링 계층으로 구성되며, 최종 합성곱 계층은 소프트맥스 활성화 함수를 사용하여 시계열 시점별 행동을 예측하는 dense 예측을 수행한다. 본 논문에서는 기존 FCN 모델의 6 개의 합성곱 계층을 모두 깊이별 분리 합성곱으로 대체하였다. 또한 입력과 첫 번째 깊이별

합성곱 계층 사이에 합성곱 계층을 추가하였다. 제안된 모델의 구조 및 처리 과정은 그림 1 과 같다.

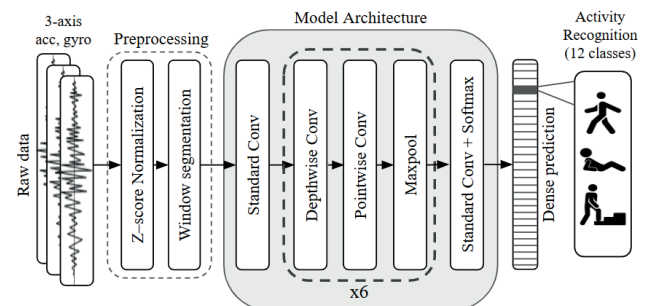


그림 1. 제안 모델 구조 및 처리 과정

#### 2.2 실험 데이터셋

HAPT[5] 데이터셋은 UCI-HAR 데이터셋의 확장 버전으로, 19 에서 48 세의 30 명의 피험자 데이터를 포함하고 있으며, 3 축 가속도계(accelerometer) 및 자이로스코프(gyroscope) 센서를 사용하여 수집되었다. 이 데이터셋은 3 가지 정적 동작, 3 가지 동적 동작, 및 6 가지 전환 동작(stand-to-sit, sit-to-stand 등)을 포함하며, 특히 전환 동작은 정적 및 동적 동작에 비해 측정 시간이 짧은 주기를 가지므로 동작 간 샘플 수의 극심한 불균형 분포를 보인다.

#### 2.3 전처리 및 평가 지표

본 논문에서는 피험자 기반 분류 기준[6]을 참고하여 훈련, 검증 및 실험 데이터로 구분하였다. 학습 데이터의 평균과 표준 편차를 기준으로 Z-점수 정규화를 수행하였으며, 원본 신호를 시퀀스로 분할하기 위해 고정된 슬라이딩 윈도우 기법을 적용하였다. 윈도우 크기는 64 로 설정하고 중첩은 적용하지 않았다. 성능

평가 지표로는 정확도(accuracy), 매크로 F1(macro averaged F1), 가중치 평균 F1(weighted average F1)을 채택하였다.

## 2.4 실험 결과 및 분석

제안 모델과 FCN 의 HAPT 데이터셋에서의 분류 성능을 평가하였으며 표 1 에 요약하였다. 제안 모델은 FCN 대비 파라미터 수를 약 78% 줄이면서도 정확도 0.865, 매크로 F1 0.6691, 가중 평균 F1 0.8639 로 모든 주요 지표에서 FCN 대비 우수한 성능을 나타냈다. 특히 매크로 F1 의 성능 개선은 클래스 불균형 환경에서도 소수 클래스에 대한 예측이 효과적으로 수행되었음을 보여준다.

표 1 . 분류 성능 비교

Model	Accuracy	Macro Averaged F1	Weighted Average F1	Parameter
FCN	0.8416	0.6643	0.8421	48,876
Proposed	0.865	0.6691	0.8639	10,892

그림 2 와 그림 3 은 혼동 행렬(confusion matrix)로, 12 가지 세부 동작에 대한 분류 성능을 나타낸다. 제안된 모델은 짧은 전환 동작, 특히 Stand-to-Sit(86.9%), Sit-to-Lie(64.8%), Stand-to-Lie(87.1%)에서 FCN 모델(81.6%, 55.8%, 72.3%)에 비해 각각 5.3%, 9%, 그리고 14.8% 높은 정확도를 보였다.

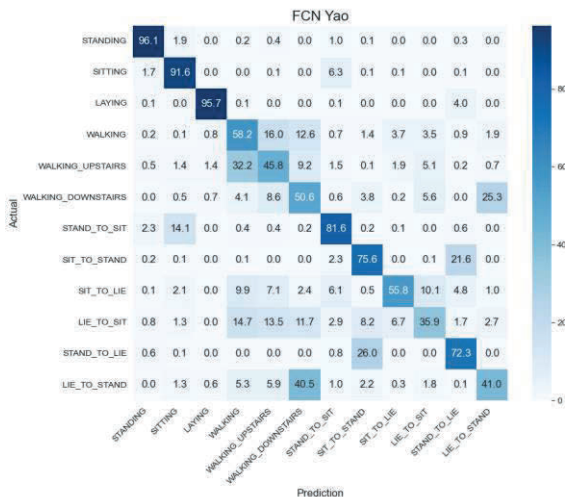


그림 2 . FCN 모델의 혼동 행렬

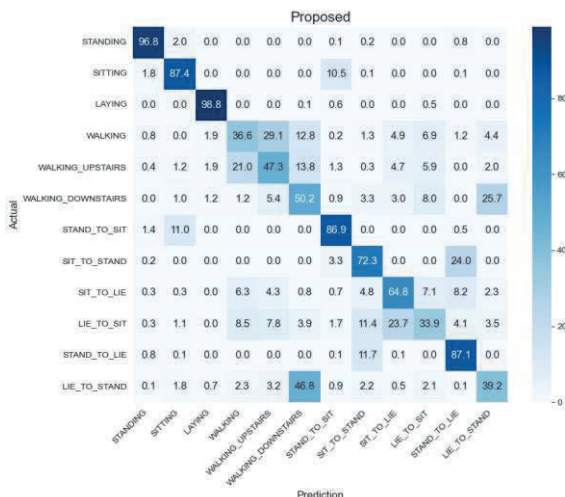


그림 3 . 제안 모델의 혼동 행렬

반면 Walking 동작의 경우 FCN(58.2%)에 비해 제안 모델은 비교적 낮은 성능(41.3%)을 보였지만, 그 외 Standing, Laying, Walking Upstairs 와 같은 정적, 동적 동작에서 우수한 분류 성능을 나타냈다.

## III. 결론

본 논문에서는 기존 FCN 기반 인간 동작 인식 모델에 깊이별 분리 합성곱을 적용한 경량 모델을 제안하였으며, HAPT 데이터셋을 활용하여 그 성능을 평가하였다. 제안된 모델은 기존 FCN 모델에 비해 약 78%의 파라미터 수를 줄이면서도 모든 주요 지표에서 우수한 성능을 나타냈다. 특히 Stand-to-Lie, Sit-to-Lie 등 짧은 전환 동작에서 FCN 보다 높은 분류 정확도를 기록하여 전환 동작 인식에서의 성능 향상을 확인하였다. 향후 연구에서는 다양한 행동을 포함한 복잡한 HAR 응용문제에 적용하고, 제안된 모델의 일반화 성능을 평가하고자 한다.

## ACKNOWLEDGMENT

본 논문은 과학기술정보통신부에서 지원하는 DGIST 기관고유사업에 의해 수행되었습니다(25-IT-02).

## 참 고 문 헌

- [1] Ó. D. Lara and M. A. Labrador, "A survey on human activity recognition using wearable sensors," *IEEE Communications Surveys and Tutorials*, vol. 15, no. 3, pp. 1192–1209, 2013, doi: 10.1109/SURV.2012.110112.00192.
- [2] R. Yao, G. Lin, Q. Shi, and D. C. Ranasinghe, "Efficient dense labelling of human activity sequences from wearables using fully convolutional networks," *Pattern Recognit*, vol. 78, pp. 252–266, Jun. 2018, doi: 10.1016/j.patcog.2017.12.024.
- [3] J. Wang, Y. Chen, S. Hao, X. Peng, and L. Hu, "Deep Learning for Sensor-based Activity Recognition: A Survey," Jul. 2017, doi: 10.1016/j.patrec.2018.02.010.
- [4] A. G. Howard *et al.*, "MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications," Apr. 2017, [Online]. Available: <http://arxiv.org/abs/1704.04861>
- [5] J. L. Reyes-Ortiz, L. Oneto, A. Samà, X. Parra, and D. Anguita, "Transition-Aware Human Activity Recognition Using Smartphones," *Neurocomputing*, vol. 171, pp. 754–767, Jan. 2016, doi: 10.1016/j.neucom.2015.07.085.
- [6] M. Ronald, A. Poulouse, and D. S. Han, "ISPLInception: An Inception-ResNet Deep Learning Architecture for Human Activity Recognition," *IEEE Access*, vol. 9, pp. 68985–69001, 2021, doi: 10.1109/ACCESS.2021.3078184.