

논문 2023-18-39

mmWave 레이더 기반 사람 행동 인식 딥러닝 모델의 경량화와 자원 효율성을 위한 하이퍼파라미터 최적화 기법

(Hyperparameter optimization for Lightweight and Resource-Efficient Deep Learning Model in Human Activity Recognition using Short-range mmWave Radar)

강 지 현*
(Jiheon Kang)

Abstract : In this study, we proposed a method for hyperparameter optimization in the building and training of a deep learning model designed to process point cloud data collected by a millimeter-wave radar system. The primary aim of this study is to facilitate the deployment of a baseline model in resource-constrained IoT devices. We evaluated a RadHAR baseline deep learning model trained on a public dataset composed of point clouds representing five distinct human activities. Additionally, we introduced a coarse-to-fine hyperparameter optimization procedure, showing substantial potential to enhance model efficiency without compromising predictive performance. Experimental results show the feasibility of significantly reducing model size without adversely impacting performance. Specifically, the optimized model demonstrated a 3.3% improvement in classification accuracy despite a 16.8% reduction in number of parameters compared th the baseline model. In conclusion, this research offers valuable insights for the development of deep learning models for resource-constrained IoT devices, underscoring the potential of hyperparameter optimization and model size reduction strategies. This work contributes to enhancing the practicality and usability of deep learning models in real-world environments, where high levels of accuracy and efficiency in data processing and classification tasks are required.

Keywords : mmWave Radar, Human Activity Recognition, Hyper-parameter optimization, Deep Learning

1. 서 론

사람의 행동을 파악하고 상황에 맞는 최적의 상호작용을 위해 사람의 행동을 파악하는 기술에 대한 수요가 증가하고 있으며, 그 중 밀리미터파 (mmWave) 레이더를 통해 실내에서 사람의 행동을 인식하는 방식은 개인 사생활을 침해할 수 있는 기존의 영상 기반 식별 및 감지 시스템을 대신하여 시각적으로 식별하지 않고도 비정상적인 활동이나 움직임 형태를 감지하여 스마트 홈 및 IoT (Internet of Things) 환경에서 사용자 동작 인식을 통한 기기 제어 및 사용자 서비스에 활용될 수 있는 기술로 각광받고 있다 [1, 2]. 또한, 마이크로컨트롤러 및 초고주파 안테나 설계 기술의 발달로 밀리미터파 레이더는 작고, 저렴하며, 비영상 기반으로 개인정보 보호에 유리하다는 점에서 차량 및 산업응용 분야에서 이상적인 센서로 활용되고 있다 [3].

물체 감지를 위한 밀리미터파 레이더 기술은 물체와의 거리, 속도 측정, 심지어 호흡과 심장박동에 의한 흉부의 변위를 통해 심장 박동 수, 호흡 수를 측정할 수 있고, 레이더 센서의 송수신 안테나 배열 조정 및 가상안테나 기술을 통

해 대략적인 물체의 형상을 파악할 수 있는 포인트 클라우드 획득이 가능하다 [4, 5]. 특히, 레이더 센서를 통해 획득된 포인트 클라우드는 물체의 대략적인 형상을 유추할 수 있고 해당 포인트 클라우드 형상의 시간적인 변화를 딥러닝으로 대표되는 인공지능 모델을 통해 물체의 행동 변화를 추정하여 앉기, 걷기, 뛰기, 쓰러짐 등 실내 환경에 발생할 수 있는 사람의 행동 패턴을 인식할 수 있다 [6, 7].

밀리미터파 레이더 센서와 같이 주변 상태를 측정하는 IoT 장치의 경우 일반적으로 임베디드 환경에서 동작하고, 이러한 임베디드 환경에서 사용되는 MCU (Micro Controller Unit)는 상대적으로 적은 메모리 공간의 제약, 낮은 클럭 속도에 따른 연산속도에 제약이 있어 범용 PC와 같은 곳에 적용된 인공지능 모델을 직접 적용하기에는 한계가 존재한다. 밀리미터파 레이더 센서에서 사람의 행동 인식을 위해 측정된 포인트 클라우드의 경우 3차원 복셀 (voxel) 형태로 구성되어 CNN (Convolutional Neural Network), RNN (Recurrent Neural Network) 계열의 딥러닝 모델에 적용되기 위한 입력 및 모델 파라미터 처리 과정에서 상당한 큰 메모리 공간과 빠른 처리속도가 요구된다 [8, 9].

본 연구에서는 밀리미터파 레이더에서 측정된 포인트 클라우드를 기반으로 사람의 행동을 분류할 수 있는 딥러닝 모델을 제한된 컴퓨팅 능력을 갖는 임베디드 시스템에 적용

*Corresponding Author (jhkang@duksung.ac.kr)

Received: Jun. 28, 2023, Revised: Aug. 7, 2023, Accepted: Nov. 14, 2023.
J. Kang: Duksung Women's University (Assoc. Prof.)

할 수 있는 가능성을 확인하기 위한 딥러닝 모델의 하이퍼파라미터 최적화 전략을 소개하고, 성능평가를 통해 행동 분류 성능을 유지 또는 개선 시키면서, baseline 딥러닝 모델을 경량화할 수 있는 방안을 제안한다. 일반적으로 딥러닝 모델에서 하이퍼파라미터 최적화는 학습률 개선하고, 성능 수치를 향상시키는데 목적이 있다. 본 논문에서는 성능을 개선한 것과 더불어 일정 수준의 정확도를 유지하면서 자원제약적인 장치에 딥러닝 모델을 효율적으로 적용하기 위한 경량화 관점에서의 하이퍼파라미터 최적화 기법을 2단계에서 걸쳐 적용하는 방식을 제안한다.

2단계에 걸친 coarse-to-fine 방식을 통해 일정 수준의 성능과 모델 크기에 제한을 두어 탐색 공간을 특정하여 학습 속도를 높이고, 선별된 탐색공간 내 모든 조합에 대하여 성능과 모델 크기를 평가하여 적합한 하이퍼파라미터 조합을 선택할 수 있다. 제안된 방식을 통해 baseline 모델과 유사한 성능 수치를 갖지만 학습 파라미터 수는 2.6%로 경량화할 수 있고, 16.8%로 학습 파라미터를 줄였음에도 인식 성능은 3.3% 개선할 수 있음을 실험을 통해 검증하였다.

본 논문은 다음과 같이 구성되어 있다. 먼저 II 장에서는 밀리미터파 레이더를 이용한 사람 행동 인식과 딥러닝 모델의 하이퍼파라미터 최적화 방법에 대한 최신 연구 내용을 소개하고, III 장에서는 본 연구에서 활용한 데이터셋과 baseline 모델 소개 및 모델을 경량화 시키면서 성능을 유지하거나 향상시킬 수 있는 전략과 실험 결과에 대하여 소개한다. IV장에서 연구 결과에 대한 해석과 이를 바탕으로 한 향후 연구 방향에 대하여 논의한다.

II. 관련연구

1. mmWave 레이더 포인트 클라우드 기반 행동 인식 기법

레이더는 전파의 송수신 방식에 따라 펄스 레이더와 FMCW (Frequency Modulation Continuous Wave) 구분되며, 근거리 내 물체와의 거리, 이동 속도를 측정하기 위해 연속적인 chirp 신호를 사용하는 FMCW 방식이 활용된다. 시간에 따라 주파수가 선형적으로 변하는 chirp 신호와 물체의 의해 반사된 신호의 변화를 분석하여 물체와의 거리와 속도 측정이 가능하며, 다중 안테나 배열과 가상 안테나 기술을 이용하여 안테나 채널별 측정된 포인트들의 집합을 형성할 수 있다. 레이더를 이용하여 비영상기반으로 사람의 손동작 인식 및 행동인식을 위해서는 측정된 포인트별 거리해상도 세분화가 필요하여 초광대역 대역폭을 사용할 수 있는 60GHz 이상 밀리미터파 대역의 소형 레이더가 활용되고 있다 [10].

사람의 행동 변화에 대하여 레이더 센서를 통해 측정된 포인트 클라우드는 레이더 센서와의 상대적 거리, 움직임 속도를 나타내며, 일정 시간 내 발생하는 포인트 클라우드의 형상 분석을 통해 행동 패턴 분류가 가능하다. 최근 3차원 CNN을 통한 포인트 클라우드 특징 추출, 포인트 클라우드의 시간적 변화를 반영하여 특징을 추출하는 RNN 계열인 LSTM (Long-Short Term Memory), GRU (Gated

Recurrent Unit) 적용한 연구사례와 [11, 12], 기존 RNN 계열보다 시계열 특징을 잘 반영할 수 있는 TCN (Temporal Convolutional Network) [13] 적용하여 레이더에서 측정된 포인트 클라우드 기반 손동작 인식 및 행동 인식에 성능 개선을 유도하고 있다. 포인트 클라우드의 정해진 포맷이 없기 때문에, 대부분의 연구가 3차원 복셀로 변경하거나, 2차원 이미지들의 집합으로 변경 후 딥러닝 모델의 입력으로 사용하여 합성곱 계층으로 인한 연산량의 증가와 변환과정에서 데이터의 본래 특성이 일부 소실되는 가능성이 있는데 PointNet [14]은 포인트 클라우드 자체를 딥러닝 모델의 직접 입력하는 방식과 유사한 성능 수치를 만족하면서 모델의 크기를 줄일 수 있는 방식을 제안하였다.

2. 딥러닝 모델의 하이퍼 파라미터 최적화 기법

하이퍼파라미터 최적화는 딥러닝 모델 크기, 성능, 실행 속도에 큰 영향을 미치는 중요한 과정이며, 모델의 학습 속도와 일반화 성능을 크게 향상 시키기 위한 필수 과정이다. 딥러닝 모델 설계자의 직관에 의해 여러 하이퍼파라미터 조합을 설정하여 성능을 개선할 수 있지만, 모든 조합 중 가장 우수한 하이퍼파라미터 조합을 수작업으로 찾아내는 것은 현실적으로 불가능하다. 다양한 하이퍼파라미터 튜닝 전략이 있지만 본 논문에서는 가장 널리 사용되는 몇 가지 방법을 소개한다.

Grid search는 탐색 범위 내 모든 하이퍼파라미터 조합을 학습하여 결과를 비교하므로 완전하지만, 많은 수의 조합을 탐색하는데 시간과 컴퓨팅 자원이 풍부할 때만 적합하다. 반면 random search는 grid search와 비슷하지만 각 하이퍼파라미터에 대한 값을 랜덤하게 선택하여 grid search보다 계산적으로 효율적이고 높은 차원의 하이퍼파라미터 공간에서 효율적일 수 있다. 베이지안 최적화는 확률적 접근 방식을 사용하여 과거의 평가 결과를 사용하여 어떤 하이퍼파라미터를 다음에 테스트할지 결정하며, 이는 하이퍼파라미터 공간을 효율적으로 탐색할 수 있게 지원하여 grid search나 random search보다 적은 수의 반복을 필요로 하나 구현이 다소 복잡한 단점이 존재한다. Hyperband는 random search를 사용하여 하이퍼파라미터 조합을 랜덤하게 선택하고, 초기 학습 과정에서 비효율적인 하이퍼파라미터를 신속하게 제거하여 컴퓨팅 자원을 절약하는 방식을 채택한다 [13].

최적의 하이퍼파라미터 튜닝 전략을 선택하는 것은 딥러닝 모델의 종류, 데이터 형식, 사용 가능한 컴퓨팅 자원에 따라 달라지며, 여러 전략을 실험적으로 평가하고, 주어진 상황에 적합한 방법을 선택하는 것이 중요하다. 본 논문에서는 IoT 환경에 적용될 수 있는 학습 모델의 크기, 파라미터 수를 고려하여 baseline 모델의 파라미터 수 보다 작은 크기를 갖도록 하이퍼파라미터 별 탐색 공간을 설정하고 hyperband와 grid search를 통해 탐색 공간을 2단계로 세분화하여 최적 하이퍼파라미터 조합을 모델 크기와 성능 수치를 기준으로 효과적인 하이퍼파라미터를 선택할 수 있는 방법을 실험적 결과를 통해 제시한다.

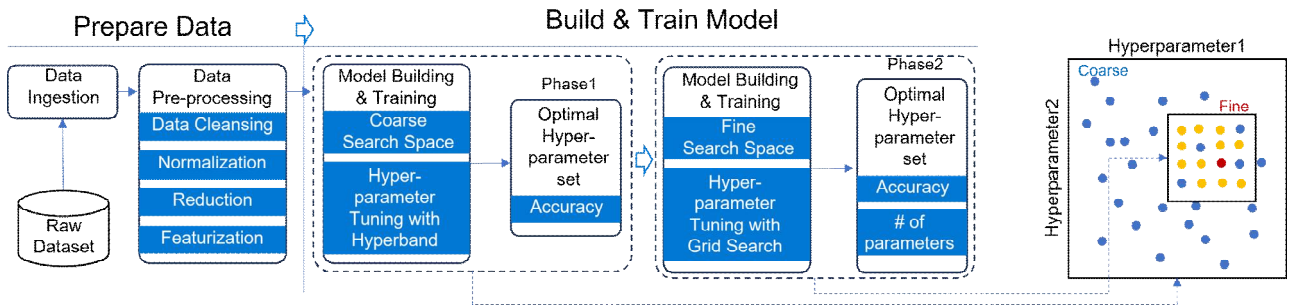


그림 1. Coarse-to-fine 하이퍼파라미터 최적화 절차
 Fig. 1. Procedure of coarse-to-fine hyperparameter optimization

III. 본 론

1. Coarse-to-fine 하이퍼파라미터 탐색 공간 설정 방식

본 논문에서는 2단계 coarse-to-fine 탐색 공간 설정 및 학습 방식을 통해 baseline 딥러닝 모델 대비 모델 크기는 줄이면서 성능을 향상 시킬 수 있는 하이퍼파라미터 탐색 전략을 제안한다.

그림 1은 제안 방식인 단계별 하이퍼파라미터 최적화 절차를 나타낸다. raw 데이터셋은 딥러닝 모델의 입력으로 사용하기 위해 전처리과정을 거치며, 모델 학습과정에서는 두 단계를 거쳐 baseline 모델 대비 파라미터 수는 적으면서 성능을 향상 시킬 수 있는 최적의 하이퍼파라미터 조합을 선정하게 된다. 1단계 (Phase 1)에서는 성능 관점에서 일정 수준 이상의 성능 보장과 학습률의 지속적인 증가를 기준으로 적용 가능한 하이퍼파라미터들에 대하여 baseline 모델 대비 모델 파라미터를 줄일 수 있는 탐색공간을 설정하고 hyperband 최적화 기법을 통해 하이퍼파라미터 조합을 랜덤하게 선택하고, 선택된 하이퍼파라미터 적용 후 최종 학습 단계까지 수행하지 않고 학습률과 성능을 고려하여 초기 학습단계에서 비효율적인 하이퍼파라미터를 신속하게 제거하여 컴퓨팅 자원을 절약하는 방식을 채택하였다.

2단계 (Phase 2)에서는 랜덤하게 선택된 하이퍼파라미터 조합 중 성능이 가장 우수한 하이퍼파라미터 조합을 기준으로 각 하이퍼파라미터 탐색 범위를 세분화하여 grid search 방식을 통해 세분화된 모든 하이퍼파라미터 조합에 대하여 성능 및 파라미터 수를 기준으로 적절한 모델 크기와 성능 수치를 기반으로 최종 하이퍼파라미터 조합을 선택하는 방식을 적용하였다

그림 1은 2개의 하이퍼파라미터에 조합 선정에 대한 예시를 나타낸다. 하이퍼파라미터 1, 2를 나타내는 사각형이 baseline 보다 작은 모델 크기를 갖는 영역이고, 1단계 hyperband를 적용하여 일정 수준 이상의 학습률과 성능을 만족하는 조합을 파란색 점으로 표시하였다. 1단계 결과를 통해 도출된 하이퍼파라미터 조합 중 일정 수준 이상의 정확도와 모델 크기를 갖는 조합 영역을 작은 사각형으로 표시하였고, 해당 영역을 보다 세분화하여 해당 영역에서 조합 가능한 모든 하이퍼파라미터 (노란색 점) 별 성능과 모

표 1. 데이터셋 구성

Table 1. Dataset Configuration

Activity	Total duration (seconds)	# of samples	
		Train	Test
Boxing	1,115	2,395	706
Jumping Jacks	1,062	2,313	668
Jumping	1,045	2,213	729
Squats	1,090	2,390	670
Walking	1,269	2,786	764

델 크기를 비교하여 최종적으로 하나의 조합 (빨간색 점)을 선별할 수 있고, 선별 기준은 경량화 관점에서 모델 크기 또는 성능 관점에서 선정할 수 있다.

2. 데이터셋 및 Baseline 딥러닝 모델

본 논문에서는 RadHAR [6]에서 적용된 5가지 행동에 대한 포인트 클라우드 데이터셋과 time-distributed 3D CNN 계층과 bi-directional LSTM이 적용된 딥러닝 모델을 baseline 모델로 지정하여 행동 분류 성능을 90.47% 이상으로 유지하며 RadHAR의 결과인 모델 파라미터 수보다 파라미터 수가 적은 모델로 경량화할 수 있도록 coarse-to-fine 하이퍼파라미터 최적화 기법을 적용하였다.

RadHAR 데이터셋은 5가지 행동 (walking, jumping, jumping jacks, squats, boxing)을 2명의 피시험자가 각 행동에 대해 20초 동안 반복한 것을 TI사 IWR1443BOOST 레이더 센서를 이용해 30 fps로 수집한 포인트 클라우드를 저장 후 복셀로 변환하는 전처리를 거쳐, 한 샘플당 60 프레임 × 10 × 32 × 32 (시간×깊이×가로×세로)의 텐서로 구성되어있고 총 93분 분량의 데이터로 71.6분은 학습용 데이터, 21.4분은 테스트용 데이터로 사용하였다. 세부 데이터셋 구성은 표 1과 같다.

본 논문에서 사용된 baseline 모델 [6]은 RadHAR 데이터셋을 검증하기 위해 제시된 모델이며 표 2와 같이 6개의 time-distributed 3D CNN 레이어와 1개의 bi-directional LSTM 레이어 (총 207,713개 파라미터)로 구성되어있고, 본 논문에서 사용하고 있는 데이터셋을 검증하기 위한 모델로 제시되어 특징적인 부분은 두 개의 time-distributed 3D

표 2. Baseline 모델 구조

Table 2. Baseline model architecture

Layer	# of units	Size	Stride	Activation
3D Conv 1a	32	(3,3,3)	(1,1,1)	Relu
3D Conv 1b	32	(3,3,3)	(1,1,1)	Relu
3D Max Pooling		(2,2,2)	(2,2,2)	
3D Conv 2a	32	(3,3,3)	(1,1,1)	Relu
3D Conv 2b	32	(3,3,3)	(1,1,1)	Relu
3D Max Pooling		(2,2,2)	(2,2,2)	
3D Conv 3a	32	(3,3,3)	(1,1,1)	Relu
3D Conv 3b	32	(3,3,3)	(1,1,1)	Relu
3D Max Pooling		(2,2,2)	(2,2,2)	
Dropout 1				0.5
LSTM	16			
Dropout 2				0.3
Dense	5			

opt_fun : Adam, lr : 0.001, beta_1 : 0.9, beta_2 : 0.999

표 3. 1단계 하이퍼파라미터 최적화에 적용된 탐색 범위

Table 3. Search space of phase1 hyperparameter optimization

Hyperparameters	Description	search space
conv_filter_size	Size of 3D convolutional layer filter	{3, 5}
num_conv_filter	# of 3D convolutional layer filter	{12, 24, 32}
conv_stride	stride value of 3D convolutional layer	{1, 2}
pool_size	pooling value of 3D max pooling	{1, 2}
strides	stride value of 3D max pooling	{1, 2}
num_block	# of block composed of 3D conv and 3D max pooling	{1, 2, 3}
num_conv_layers	# of 3D conv layer per block	{1, 2}
dropout1	Dropout rate	{0.1, 0.3, 0.5, 0.7, 0.9}
dropout2	Dropout rate	{0.1, 0.3, 0.5, 0.7, 0.9}
num_lstm_cell	# of LSTM cell	{8, 12, 16}
opt_func	optimization function	{adam, sgd}
beta_1	Adam parameter	{0.5, 0.7, 0.9}
beta_2	Adam parameter	{0.5, 0.7, 0.99}
lr	learning rate	{0.00001, 0.0001, 0.001, 0.01}

CNN 레이어와 1개의 max pooling 레이어가 하나의 블럭을 만들어, 총 3개의 CNN 블럭과 한 개의 LSTM 레이어로 구성되어 있다.

3. 하이퍼파라미터 최적화 실험 결과

Coarse-to-fine 방식으로 2단계에 걸쳐 실험을 수행하였으며, 1단계는 baseline 모델에 적용된 하이퍼파라미터를 바탕으로 표 3과 같이 모델 파라미터 수를 줄일 수 있도록 탐색 범위를 설정하였다. 전체 탐색 범위에 따른 하이퍼파라미터 조합은 466만개로 확보한 컴퓨팅 자원을 사용 시 모든 조합에 대하여 평가할 수 없어 Hyperband를 적용하였다.

그림 2는 Hyperband 알고리즘으로 탐색 공간 내 하이퍼파라미터 조합에 의해 학습되는 과정의 경향성을 나타내며, 그래프의 가로축은 하이퍼파라미터 종류, 세로축은 각 하이퍼파라미터의 설정 값을 나타낸다. Iteration이 진행될수록

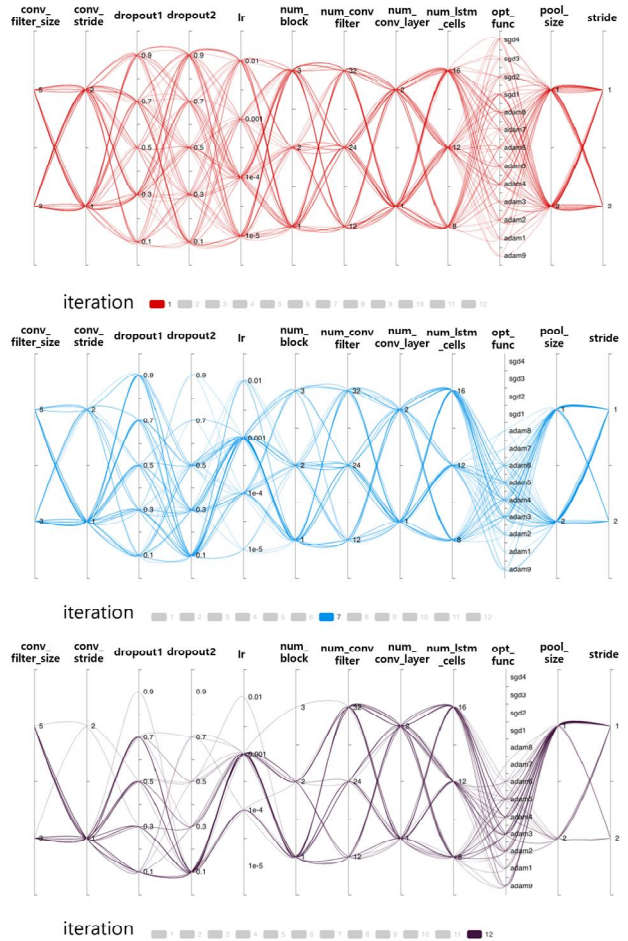


그림 2 Iteration에 따른 최적화 경향성

Fig. 2. Optimization tendency according to iteration

표 4. 1단계 실험에서 최대 성능을 만족하는 하이퍼 파라미터 구성
Table 4. hyperparameter configuration that satisfy maximum performance in Phase 1

Hyperparameters	value
conv_filter_size	3
num_conv_filter	12
conv_stride	1
pool_size	2
strides	2
num_block	3
num_conv_layers	1
dropout1	0.1
dropout2	0.3
num_lstm_cell	16
opt_func	adam
beta_1	0.9
beta_2	0.99
lr	0.001

성능을 기준으로 적합한 하이퍼파라미터 조합 선정되는 것을 확인할 수 있다. 컨볼루션 계층에서 사용되는 커널의 stride를 나타내는 conv_stride의 경우 1과 2의 조합이 가능하며, iteration 초기에는 1과 2 모두 동일한 비율로 나타나

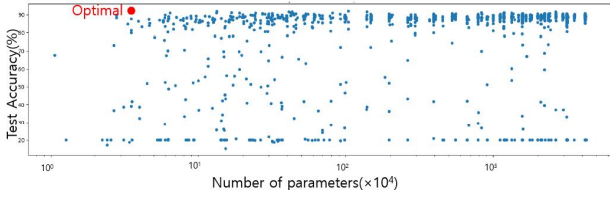


그림 3. 1단계 실험에서의 파라미터 수에 따른 성능 분포
Fig. 3. Performance distribution according to number of parameters in Phase 1

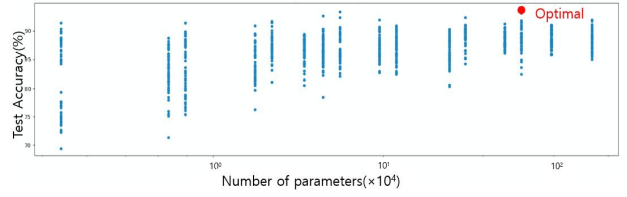


그림 4. 2단계 실험에서의 파라미터 수에 따른 성능 분포
Fig. 4. Performance distribution according to number of parameters in Phase 2

표 5. 2단계 하이퍼파라미터 최적화에 적용된 탐색 범위
Table 5. Search space of phase2 hyperparameter optimization

Hyperparameters	Description	search space
conv_filter_size	Size of 3D convolutional layer filter	{3}
num_conv_filter	# of 3D convolutional layer filter	{6, 12}
conv_stride	stride value of 3D convolutional layer	{1}
pool_size	pooling value of 3D max pooling	{2}
strides	stride value of 3D max pooling	{2}
num_block	# of block composed of 3D conv and 3D max pooling	{3}
num_conv_layers	# of 3D conv layer per block	{1, 2}
dropout1	Dropout rate	{0.1, 0.3, 0.7}
dropout2	Dropout rate	{0.3, 0.5}
num_lstm_cell	# of LSTM cell	{4, 8, 12, 16}
opt_func	optimization function	{adam}
beta_1	Adam parameter	{0.5, 0.7, 0.9}
beta_2	Adam parameter	{0.5, 0.99}
lr	learning rate	{0.0001, 0.001}

표 6. 2단계 실험에서 최대 성능을 만족하는 하이퍼 파라미터 구성
Table 6. Search space of phase1 hyperparameter optimization

Hyperparameters	value
conv_filter_size	3
num_conv_filter	12
conv_stride	1
pool_size	2
strides	2
num_block	3
num_conv_layers	1
dropout1	0.1
dropout2	0.3
num_lstm_cell	16
opt_func	adam
beta_1	0.5
beta_2	0.99
lr	0.001

지만, 학습 종류 단계에서 conv_stride는 1로 수렴함을 확인할 수 있다. 학습 과정을 시각화한 것이지만 실험결과 표 4를 통해 stride 값이 2보다는 1이 더 높은 성능을 나타냄을 확인할 수 있다.

하이퍼파라미터가 조합되는 과정에서 컨볼루션 레이어나 풀링 레이어에 의해 차원 축소가 충분히 이루어지지 않은 상태에서 많은 수의 cell을 가진 LSTM 레이어와 만날 경우 파라미터 수가 수천만에 달하는 경우도 발생할 수 있다.

그림 3은 1단계 하이퍼파라미터 최적화를 통한 조합 별 모델의 크기 (학습 파라미터 수)에 따른 테스트셋에서의 정확도를 나타내며, 표 3과 4은 1단계 최적화 과정에서 설정된 하이퍼파라미터 탐색 공간과 가장 우수한 성능을 나타내는 하이퍼파라미터 조합을 나타낸다. 이때 모델 크기는 파라미터 수 35,053개이며 정확도는 92.3%로 baseline 하이퍼파라미터 조합 대비 모델 크기는 약 16.8%로 축소하였고, 분류 성능은 1.83% 높게 개선시키는 효과를 확인할 수 있어 성능을 유지하면서 경량화에 대한 목표를 충족했다고 판단할 수 있다.

2단계 최적화 실험에서는 1MByte 이하의 flash 메모리를 갖는 MCU 적용하는 것을 가정하고, 센싱 및 연산 과정에 필요로 하는 메모리 등을 고려하고, 1단계 실험 결과를 바탕으로 임의로 정한 경량화 조건인 테스트셋에서 baseline 모델 하이퍼파라미터 설정값에 따른 최소 분류 정확도를 기준으로 70% 이상의 정확도와 4만개 이하의 파라미터 수를

충족하는 모델에 사용된 하이퍼파라미터를 바탕으로 표 5와 같이 탐색 공간을 설정하였고 총 1,152개 조합에 대하여 grid search를 통해 모델 크기와 성능을 비교하였다. 학습을 수행하는 머신의 성능 및 적용 대상 장치의 성능을 고려하여 탐색 공간의 축소 및 확대를 통해 최적 조합을 찾을 수 있을 것으로 기대한다.

그림 4는 2단계 실험에서 모델의 크기 (파라미터 수)에 따른 테스트셋에서의 정확도를 나타내며, 표 6는 2단계 최적화 과정에서 최대 성능에 대한 하이퍼파라미터 조합을 나타낸다. 최고 성능을 나타내는 하이퍼파라미터 조합은 1단계 조합 대비 beta_1 값을 제외한 모든 값이 동일한 파라미터 수 35,053개의 모델이며, 정확도는 93.7%로 1단계 조합보다 1.4% 상승했음을 확인할 수 있다. 이러한 성능 향상의 주된 요인은 1단계 조합보다 grid search 방식으로 낮은 학습 속도 (beta_1=0.5)로 30 epoch를 더 학습함에 따라 optimal point에 더 가깝게 수렴한 것으로 판단된다.

표 7는 모델 크기별 성능 변화를 요약한 것이며, 여기서 관측되는 특징은 파라미터 수 약 5,000개 정도로 모델 크기가 baseline 대비 2.6%로 줄어도 분류 정확도 90% 이상의 성능을 만족할 수 있고, 대부분의 조합에서 baseline 조합 대비 모델 크기가 작아졌음에도 분류 정확도를 유사하게 유지할 수 있음을 확인할 수 있다.

경량화 관점에서 모델 크기에 직접적으로 연관된 변수는 3D CNN 레이어와 LSTM 레이어이고 최대 성능 93.7%에서의 num_lstm_cell이 16이고, conv_size는 72가 아닌 36인 경

표 7. 2단계 하이퍼파라미터 최적화에 적용된 탐색 범위
Table 7. Search space of phase2 hyperparameter optimization

# of parameters	Ratio of model size	Accuracy(%)
5,401	2.6%	90.3
8,335	4.0%	88.2
8,929	4.3%	91.4
11,863	5.7%	90.9
12,713	6.1%	91.6
14,485	6.9%	89.3
15,647	7.5%	92.6
16,753	8.0%	93.3
19,687	9.4%	90.6
21,085	10.1%	90.9
26,185	12.6%	90.3
27,941	13.4%	92.3
32,785	15.7%	91.2
35,053	16.8%	93.7
39,641	19.0%	91.0
46,753	22.5%	91.7

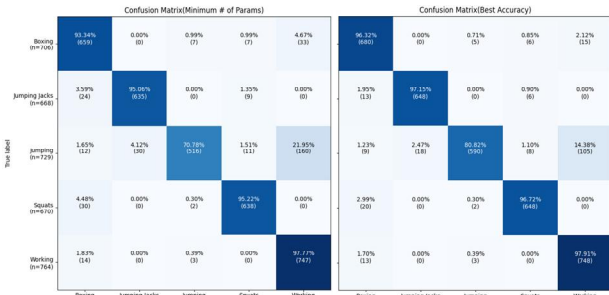


그림 5. 최소 모델 크기 및 최대 성능에 대한 성능 시험 결과 혼동행렬

Fig. 5. Confusion matrix for minimum model size and best performing experiment results

우인 것으로 미루어 봤을 때 시계열 데이터의 특성으로 인해 LSTM 레이어의 capacity가 성능에 가장 큰 영향을 준 것으로 판단된다. 그림 5는 baseline 모델을 경량화하여 최소 학습 파라미터 수를 갖는 경우와 최대 성능을 나타내는 경우의 정확도 성능 수치에 대한 혼동행렬을 나타내며, 하이퍼파라미터 조정이 전체적인 성능 수치의 향상에 기여하였음을 확인할 수 있으며, 상대적으로 낮은 분류 성능을 보이는 jumping의 경우 개별 클래스에 대한 성능 향상에 하이퍼파라미터 조정 영향에는 한계가 있으며, baseline 모델 구조를 변경하거나 학습 데이터셋을 재수집하는 과정을 통해 분류 성능을 높일 수 있을 것으로 판단된다.

IV. 결론

본 논문에서는 mmWave 레이더에서 획득된 사람 행동에 대한 포인트 클라우드 데이터를 통해 행동을 분류하는 기능을 자원 제약적인 IoT 장치에 적용 가능성을 확인하기 위한 baseline 딥러닝 모델의 하이퍼파라미터 최적화 기법 및 절

차에 대하여 제안하고 실험을 통해 가능성을 검증하였다. 경량화된 모델을 실제 IoT 장치에 적용한 실험을 수행하지는 못하여 경량화 모델의 실제적인 효율성을 정량적으로 평가하지는 못하였지만, 사용자가 설정한 baseline 모델을 기준으로 성능을 유지하거나 일정 수준 증가시키면서, 메모리 사용량을 획기적으로 줄일 수 있는 경량화 과정은 본 연구와 유사한 응용연구에 기여할 수 있을 것으로 판단된다. 2단계 coarse-to-fine 최적화 절차를 통해 탐색 공간이 상대적으로 넓어 컴퓨팅 자원을 고려한 hyperband 기법을 적용하여 임의의 수준 (파라미터 수 및 성능)을 만족하는 탐색 범위를 정하고, 1단계에서 정해진 탐색 범위를 축소하고 세분화하여 grid search 방식을 통해 2단계 전체 탐색 공간에 대하여 파라미터 수, 성능 수치 도출을 통해 baseline 모델 및 파라미터 세팅 대비 모델 크기를 줄이면서, 성능을 개선할 수 있음을 확인하여, baseline 모델 수준의 성능을 유지하면서 모델 크기를 경량화할 수 있음을 실험을 통해 확인하였다. 본 연구에서 제안한 coarse-to-fine 방식을 통해 메모리 사용량의 감소, 추론시간의 단축, 학습률 개선 등 일정 수준의 정확도를 유지하면서 baseline 모델을 경량화하기 위한 하이퍼파라미터 탐색 및 설정 기법을 레이더의 시계열 포인트 클라우드 기반 행동인식 및 손동작 인식 응용 시 실제 자원 제약적인 장치에 적용 가능할 것으로 기대된다.

본 연구에 적용한 mmWave 레이더 기반 사람행동인식용 딥러닝 모델의 경량화 성공의 주요 요인은 3D CNN 레이어를 적절한 수준까지 차원 축소한 후에 충분한 capacity를 갖는 LSTM 레이어로 처리한 것으로 판단되며, 실험 결과는 성공적인 경량화를 의미하지만, 실험에 사용된 데이터는 실험실 환경에서 수집된 공개 데이터셋이므로 실제 서비스 환경에서 행동 데이터 수집 후 이에 대한 추가 검증 및 실제 자원 제약적인 타겟 MCU에 적용 가능성에 대해서는 추가적인 검증이 필요할 것으로 판단된다. 또한, 딥러닝 모델의 입력 데이터 및 시계열 모델 연산량 관점에서 3차원 공간 데이터인 포인트 클라우드를 2차원 평면으로 투영하는 전처리 기법과 타겟 MCU 선정 후 해당 환경에서 구동하기 위해 지원되는 연산자의 종류 및 실행 시 발생하는 메모리 사용량 및 처리 시간 등 경량화 수행 후 실제 적용 가능성에 대한 검증이 필요할 것으로 판단되어 향후 연구로 진행 예정이다.

References

- [1] C. Yu, Z. Xu, K. Yan, Y. R. Chien, S. H. Fang, H. C. Wu, "Noninvasive Human Activity Recognition Using Millimeter-Wave Radar," *IEEE Systems Journal*, Vol. 16, No 2, pp. 3036-3047, 2022.
- [2] A. Shastri, N. Valecha, E. Bashirov, H. Trtaria, M. Lentmaier, F. Tufvesson, M. Rossi, P. Casari, "A Review of Millimeter Wave Device-Based Localization and Device-Free Sensing Technologies and Applications," *IEEE Communications Surveys &*

- Tutorials, Vol. 24, No. 3, pp. 1708-1749, 2022.
- [3] B. Liu, K. Ma, H. Fu, K. Wang, F. Meng, "Recent Progress of Silicon-Based Millimeter-Wave SoCs for Short-Range Radar Imaging and Sensing," *IEEE Transactions on Circuits and Systems II: Express Briefs*, Vol. 69, No. 6, pp. 2667-2671, 2022.
- [4] Y. Rong, I. Lenz, D. W. Bliss, "Vital Signs Detection Based on High-Resolution 3-D mmWave Radar Imaging," *2022 IEEE International Symposium on Phased Array Systems & Technology*.
- [5] S. Lyer, L. Zhao, M. P. Mohan, J. Jimenno, M. Y. Siyal, A. Alphones, M. F. Karim, "mm-Wave Radar-Based Vital Signs Monitoring and Arrhythmia Detection Using Machine Learning," *Sensors*, Vol. 22, No. 9, 2022.
- [6] A. D. Singh, S. S. Sandha, L. Garcia, M. Srivastava, "RadHAR: Human Activity Recognition from Point Clouds Generated through a Millimeter-wave Radar," *Proceedings of the 3rd ACM Workshop on Millimeter-wave Networks and Sensing Systems*, pp. 51-56, 2019.
- [7] H. Arab, I. Ghaffari, L. Chioukh, S. O. Tatu, StevenDufour, "A Convolutional Neural Network for Human Motion Recognition and Classification Using a dxMillimeter-Wave Doppler Radar," *IEEE Sensors Journal*, Vol. 22, No. 5, pp. 4494-4502, 2022.
- [8] P. P. Ray, "A review on TinyML: State-of-the-art and Prospects," *Journal of King Saud University - Computer and Information Sciences*, Vol. 34, No. 4, pp. 1595-1623, 2022.
- [9] F. J. Adbu, Y. Zhang, M. Fu, Y. Li, Z. Deng, "Application of Deep Learning on Millimeter-Wave Radar Signals: A Review," *Sensors*, Vol. 21, No. 6, 2021.
- [10] T. Stadelmayer, M. Stadelmayer, A. Santra, R. Weigel, F. Lurz, "Human Activity Classification Using mm-Wave FMCW Radar by Improved Representation Learning," *Proceedings of the 4th ACM Workshop on Millimeter-Wave Networks and Sensing Systems*, Vol. 1, No. 1, pp. 1-6, 2020.
- [11] H. Zhou, Y. Zhao, Y. Liu, S. Lu, X. An, Q. Liu, "Multi-Sensor Data Fusion and CNN-LSTM Model for Human Activity Recognition System," *Sensors*, Vol. 23, No. 10, 2023.
- [12] X. Huang, N. Patel, K. P. Tsoi, "Application of mmWave Radar Sensor for People Identification and Classification," *Sensors*, Vol. 23, No. 8, 2023.
- [13] S. Bai, J. Z. Kolter, V. Koltun, "An Empirical Evaluation of Generic Convolutional and Recurrent Networks for Sequence Modeling," *arXiv:1803.01271*, 2018.
- [14] C. R. Qi, H. Su, K. Mo, L. J. Guibas, "PointNet: Deep Learning on Point Sets for 3D Classification and Segmentation," *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 652-660, 2017.

Jiheon Kang (강지현)



2007 Electrical Communication Engineering
from Kangwon National University
(B.S.)

2012 Electronics and Computer Engineering
from Korea University (M.S.)

2012 Department of Electrical Engineering
from Korea University (Ph.D)

2021 ~ Department of Software from Duksung Women's
University. (Assit.. Prof.)

Career:

2007 Korea Institute for Defense Analyses, Research Engineer

2016 Sensorway Co., Ltd, Senior Engineer

2021 SK Telecom., Manager

Field of Interests: Artificial IoT, Embedded System,
Machine/Deep Learning

Email: jhkang@duksung.ac.kr