https://doi.org/10.7840/kics.2025.50.5.725

인간 활동 인식에서 RF 기반 데이터를 ViT에 적용하기 위한 Resizing 방법

박 정 준◆, 박 세 웅°

Resizing Method for Applying RF-based Data to ViT in Human Activity Recognition

Jeongjun Park◆, Saewoong Bahk°

요 약

본 논문에서는 인간 활동 인식에서 주로 사용되는 Radio Frequency (RF) 방식을 통해 얻은 RF 기반 데이터를 최신 이미지 분류를 위한 머신러닝 기법인 Vision Transformer (ViT)에 적용하였다. 이 과정에서 이미지 크기와 다른 RF 기반 데이터의 크기를 ViT에 적용할 때 발생하는 문제점을 분석하고, 이를 해결하기 위해 고려해야 할 입력 사이즈resizing 방법들을 제시하였다. 또한, 다양한 resizing 방법들과의 비교를 통해 RF 기반 데이터에 가장 효과적인 resizing 방식을 제안하였으며, 이를 통해 평균 9.57%의 성능 개선을 달성하였다.

키워드 : Vision Transformer, RF 기반 데이터
Key Words : Vision Transformer, RF-based data

ABSTRACT

This paper applies RF-based data, obtained through the commonly used Radio Frequency (RF) approach in human activity recognition (HAR), to the Vision Transformer (ViT), a state-of-the-art machine learning method for image classification. Through this process, we analyze the challenges arising from applying RF-based data, which have different sizes compared to standard image dimensions, to ViT. To address these challenges, we propose various input resizing methods. Furthermore, through a comparison of these resizing methods, we identify the most effective resizing approach for RF-based data, achieving an average accuracy improvement of 9.57%.

Ⅰ. 서 론

최근 인공지능 기술의 발전으로 인간 활동 인식 (Human Activity Recognition, HAR) 기술이 헬스케어, VR/AR, 감시 시스템 등 다양한 분야에서 활용되고 있다[1-3]. 이러한 기술들은 주로 카메라를 활용하여 인간의 움직임과 행동을 파악해왔으나, 이러한 데이터 수집 방식은 개인의 사생활 침해 우려를 야기하여 사용에 제약이 따를 수 있다.

이러한 문제를 해결하기 위해 최근에는 Radio Frequency (RF) 신호를 활용한 인간 활동 인식 기술이 주목받고 있다[4]. RF 신호 기반 HAR 기술은 주로 Ultra-wideband (UWB), Frequency-Modulated Continuous Wave (FMCW), WiFi를 이용하며, 개인의 프라이버시를 보다 효과적으로 보호할 수 있다. 또한, RF 기반 데이터는 2D 이미지 형태로 변환이 가능하여, 최적화된 이미지 분류 기법(예: Convolutional Neural Networks, Vision Transformers)의 도움을 받을 수 있다. 이러한 이유로 RF 데이터를 수집하고 이를 머신러닝 기법으로 분류하려는 연구가 점차 증가하고 있다[5,6].

이미지 분류를 위한 최신 머신러닝 기법 중 하나로 Vision Transformer (ViT)가 있다[7]. ViT는 주로 224×224 크기의 이미지를 입력으로 받으며, 이를 16×16 크기의 패치로 나누어 총 14×14(196)개의 패치로 조각낸다. 각 패치는 특정 차원의 벡터(patch embedding vector)로 표현되며, 이 벡터는 패치의 위치 정보를 포함한 벡터(positional embedding vector)와 결합되어 Transformer Encoder에 입력된다. 이후 연산 과정을 통해 이미지를 분류하게 된다.

그러나 대부분의 RF 기반 데이터는 입력 크기가 224×224가 아니기 때문에 ViT에 직접 적용하기 어려운 문제가 있다. 특히, 패치의 크기가 정사각형일 경우 데이터를 완벽하게 조각내기 어렵고, 패치의 개수가 달라지면서 미리 학습된 positional embedding vector를 각 patch embedding vector에 적용할 수 없게 된다.

이를 해결하기 위해 본 논문에서는 RF 기반 데이터를

◆ First Author : Seoul National University Department of ECE and INMC, jjpark@netlab.snu.ac.kr, 학생회원
° Corresponding Author : Seoul National University Department of ECE and INMC, sbahk@snu.ac.kr, 정교수, 종신회원
논문번호 : 202412-308-A-LU, Received December 2, 2024; Revised December 27, 2024; Accepted December 27, 2024

ViT에 적용하기 위한 새로운 resizing 방법을 제안한다.

Ⅱ. 실험 방법

본 실험의 목적은 RF 기반 데이터를 resizing하여 ViT에 적용하는 것이다. RF 기반 데이터는 이미지와 다른 크기를 가지며, UWB, WiFi, FMCW 등 다양한 기술로부터 얻을 수 있다. 본 실험에서는 거리 분해능이 높고 시간에 따른 펄스의 다중 경로 정보를 포함하여 2D 형태의 데이터로 변환할 수 있는 특성을 지닌 UWB 데이터를 사용한다. UWB 데이터는 이러한 특성으로 인해 최근 다양한 연구에서 활발히 활용되고 있다. UWB 데이터의 신호를 간단히 표현하면 다음과 같다.

$$r_s(t) = \sum_{f=1}^{F} \alpha_f \delta(t-\tau_f) * m_s(t) + n_s(t) \quad (1)$$

r과 m은 수신과 송신된 신호이고, s와 f는 각각 slow time (pulse index)와 fast time (다중 경로 index)를 나타낸다. n과 α는 각각 노이즈와 감쇠 인자를 나타낸다. 수집된 UWB 데이터[8]를 이용하여 그림 1과 같이 51x500 사이즈의 데이터를 실험에 이용한다. 해당 데이터세트는 9명으로부터 7가지의 행동을 수집한 데이터이며 각 피실험자의 데이터로 테스트하여 classification 정확도의 평균을 구한다. Training data와 testing data는 철저히 분리하여 training에 참여하지 않은 data를 testing에 사용하였다.

본 논문에서는 다음의 네 가지 resizing 방식을 고안 및 비교하였다: 1) simple resizing, 2) square patching, 3) rectangular patching, 4) rectangular patching with interpolation이다.

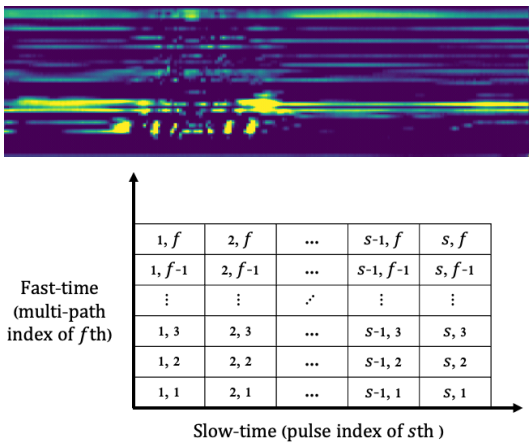

그림 1. UWB 신호의 2D 데이터화 및 데이터 예시
Fig. 1. Example of 2D manipulation of UWB signal and data.

Simple resizing 방법은 RF기반 데이터 사이즈에 상관없이 이를 224 × 224로 up-sampling 또는 down-sampling을 적용하여 이미지 사이즈와 동일하게 변환한 후 ViT에 적용하는 방식이다.

Square patching 방법은 입력 데이터의 사이즈를 변경하지 않고, 기존의 16x16패치를 이용하해 데이터를 조각낸 뒤 ViT에 적용하는 방법이다. 이때 입력 데이터의 사이즈가 16으로 나누어 떨어지지 않는다면 필요한 사이즈만큼 zero padding을 사용한다.

Rectangular patching은 정사각형 패치 대신 입력 데이터의 크기에 따라 직사각형 모양의 패치를 생성하여 patch embedding을 수행하는 방식이다. 본 실험에서는 51×500 크기의 데이터를 14×14개의 패치로 나누기 위해 패치 크기를 4×36으로 설정하였다.

마지막 방법인 rectangular patching with interpolation은 입력 데이터의 짧은 변이 224보다 작을 경우 이를 보간(interpolation)을 통해 224로 확장하고, 긴 변은 크기를 유지한 채 패치 크기를 계산하는 방식이다. 이 방법은 긴 변의 정보를 보존하고, 짧은 변의 정보를 보간하여 이미지 크기와 유사하게 맞춤으로써 resizing 과정에서 정보 손실을 최소화한다. 이 방식에서는 16×36 크기의 패치가 적용된다.

동일한 실험 환경을 유지하기 위해 다음과 같은 실험 설정을 적용하였다. ImageNet으로 사전 학습된 ViT를 활용하여 UWB 데이터를 30 epoch 동안 파인튜닝(fine-tuning) 하였다. Optimizer는 Adam을 사용하였고, learning rate은 0.00001로 설정하였으며, 손실 함수로는 cross entropy를 사용하였다.

Ⅲ. 실험 결과

각 방법의 성능을 평가하기 위해 정확도를 측정하였으며, 결과는 표 1에 요약되어 있다.

Simple resizing은 가장 간단한 방식으로, 최근 연구에서 널리 사용되고 있다[9]. 그러나 RF 기반 데이터는

표 1. 방법 별 평균 정확도
Table 1. Average accuracy for each method

Methods	Average accuracy
Simple resizing	52.20%
Square patching	14.28%
Rectangular patching	43.55%
Rectangular patching with interpolation	**61.77%**

224보다 큰 사이즈를 가질 수 있기 때문에 단순히 이를 down-sampling할 경우 정보 손실이 크게 발생한다. 특히, 입력 데이터 사이즈가 224보다 클 경우 필연적으로 정보 손실이 발생하게 된다.

Square patching의 경우, 14.28%로 가장 낮은 정확도를 보였다. 이는 입력 데이터의 사이즈에 따라 패치의 개수가 달라지기 때문이다. 패치 개수가 14×14(196개)에 미치지 못하면 사전 학습된 positional embedding vector와의 결합이 문제를 일으키게 되고, 이로 인해 학습이 제대로 이루어지지 않아 낮은 정확도가 나타난다.

Rectangular patching은 입력 데이터 사이즈에 따라 패치 면적이 작아질 수 있다. 패치 내의 정보량이 줄어들면 patch embedding 과정에서 충분한 정보를 제공하지 못해 성능 저하가 발생할 수 있다.

Rectangular patching with interpolation은 정보 손실 문제가 없고, 입력 데이터 사이즈가 224보다 작더라도 보간을 통해 데이터를 확장하여 패치 크기의 정보 손실을 보완한다. 또한, 패치 개수를 유지할 수 있어 사전 학습된 positional embedding vector의 적용에 문제가 없다. 이러한 특성 덕분에 가장 높은 정확도인 61.77%를 기록하였다.

본 논문은 정확도의 절대적인 수치에 초점을 맞추기보다, 현재 널리 사용되는 방법에 비해 정보 손실을 줄이는 resizing방식을 통해 성능을 개선할 수 있음을 주장한다.

Ⅳ. 결 론

본 논문에서는 이미지 크기와 다른 크기를 가지는 RF 기반 데이터를 ViT에 적용하기 위한input resizing 방법을 제안하였다. 최근 연구에서 가장 널리 사용되는 simple resizing 방식과 비교하여, rectangular patching과 보간을 사용한 방식은 평균적으로 9.57%의 성능 개선을 보였다. 제안된 기법은 UWB 데이터뿐만 아니라 다양한 RF 기반 데이터에도 적용 가능하여 성능 개선뿐만 아니라 확장성 측면에서도 장점을 제공한다.

References

[1] M. Karim, S. Khalid, A. Aleryani, J. Khan, I. Ullah, and Z. Ali, “Human action recognition systems: A review of the trends and state-of-the-art,” *IEEE Access*, vol. 12, pp. 36372-36390, 2024.

[2] G. Ogbuabor and R. La, “Human activity recognition for healthcare using smartphones,” in *Proc. 10th ICMLC 2018*, pp. 41-46, 2018.

[3] W. Lin, M.-T. Sun, R. Poovandran, and Z. Zhang, “Human activity recognition for video surveillance,” in *Proc. IEEE ISCAS 2008*, pp. 2737-2740, 2008.

[4] J. Park, J. Park, and S. Bahk, “UWB two-way ranging based feature extraction for human activity recognition,” *JCCI*, 2023.

[5] M. Piriyajitakonkij, et al., “Sleepposenet: Multi-view learning for sleep postural transition recognition using uwb,” *IEEE J. Biomedical and Health Inf.*, vol. 25, no. 4, pp. 1305-1314, 2020.

[6] T. Han, et al., “IR-UWB sensor based fall detection method using CNN algorithm,” *Sensors*, vol. 20, no. 20, p. 5948, 2020.

[7] A. Dosovitskiy, “An image is worth 16x16 words: Transformers for image recognition at scale,” in *Proc. ICLR*, 2021.

[8] J. Park, *ALERT dataset*, https://github.com/jeongjun-park/ALERTdataset. (Accessed 15 Nov. 2024).

[9] I. Brishtel, et al., “Driving activity recognition using UWB radar and deep neural networks,” *Sensors*, vol. 23, no. 2, p. 818, 2023.