

특집논문 (Special Paper)

방송공학회논문지 제31권 제2호, 2026년 3월 (JBE Vol.31, No.2, March 2026)

<https://doi.org/10.5909/JBE.2026.31.2.251>

ISSN 2287-9137 (Online) ISSN 1226-7953 (Print)

표면 근전도 신호 기반 동작 인식을 위한 TDS-FFT 융합 경량화 네트워크

조 경 민^{a)}, 강 현 석^{b)}, 최 형 석^{b)}, 최 우 석^{c)}, 박 구 만^{d)†}

A Lightweight TDS-FFT Fusion Network for sEMG-based Human Action Recognition

Kyungmin Cho^{a)}, Hyunseok Kang^{b)}, Hyungseok Choi^{b)}, Wooseok Choi^{c)}, and Gooman Park^{d)†}

요 약

본 논문은 표면 근전도 신호를 이용한 전신 자세 분류 및 시계열 분석 기법을 융합한 경량화 모델을 제안한다. 기존 연구가 신체 일부에 국한되었던 한계를 극복하고자, 전신에서 수집된 8채널의 표면 근전도 데이터를 활용해 전신의 복잡한 동작을 분류하는 데 초점을 맞추었다. 이를 위해 Conv2D, TDS, FFT 세 가지의 특징 추출 기법을 비교·분석하고, 이들을 융합하여 고비용의 Transformer 모델을 대체할 수 있는 경량 모델을 구성하였다. 15개 동작으로 구성된 ‘Muscles In Action’ 데이터셋으로 실험한 결과, 최대 95.947%의 정확도를 달성하였다. 결론적으로 표면 근전도 신호만으로 전신 자세 분류가 가능함을 보였으며, 시계열 분석 기법의 융합을 통해 연산 효율성과 분류 성능을 동시에 확보한 모델을 제안하였다.

Abstract

This paper proposes a lightweight model that fuses surface electromyography (sEMG) signals with time-series analysis techniques for full-body posture classification. To overcome the limitations of existing studies restricted to specific body parts, this study focuses on classifying complex full-body movements using 8-channel sEMG data collected from the entire body. To achieve this, we compared and analyzed three feature extraction techniques - Conv2D, Time-Depth Separable (TDS), and Fast Fourier Transform (FFT) - and fused them to construct a lightweight model capable of replacing computationally expensive Transformer models. Experimental results using the ‘Muscles In Action’ dataset, comprising 15 movement postures, demonstrated a maximum classification accuracy of 95.947%. In conclusion, this study demonstrates the feasibility of full-body posture classification using only sEMG signals and presents a model that secures both computational efficiency and classification performance through the fusion of time-series analysis techniques.

Keyword : sEMG, Human Action Recognition, Lightweight, Multi-channel Temporal Data

I. 서론

표면 근전도(sEMG)^[1] 신호는 신체 근육의 활동을 분석할 수 있는 중요한 생체 정보이다. 근육의 활성화 유무와 피로도 같은 생리학적 정보를 비침습적으로 제공하여 재활 의학, 스포츠공학, 로봇공학 등 다양한 분야에서 핵심적인 데이터로 활용된다. 특히 인간의 의도를 파악하여 외골격 로봇이나 전자기기를 제어하는 데 있어 sEMG 신호는 중요한 지표로 자리 잡고 있다.

최근 딥러닝 기술의 발전으로 손, 팔, 허반신 등 신체 국소 부위의 근전도 신호를 분석하여 자세를 추정하는 연구들이 활발히 진행되고 있다. 그러나 이러한 연구들은 신체의 제한적인 부분만을 대상으로 진행되고 있다는 한계가 있다. 사람의 자세는 많은 근육의 유기적인 상호작용의 결과이다. 자세를 취하거나 유지할 때 여러 근육은 순차적 또는 동시에 움직이며, 비슷한 자세라고 하더라도 주변 환경의 변화와 같은 변칙적인 이유로 상호작용을 하는 근육들이 바뀔 수도 있다. 이러한 복잡한 상호작용을 통해 만들어지는 움직임은 신체의 제한적인 부분의 신호만으로는 제대로 파악하기 어렵다. 그렇기에 근육 간의 정교한 패턴을 전체적으로 파악하기 위해서는 전신의 신호를 동시에 분석하는 연구가 필요하다.

이에 본 논문에서는 전신 근전도 신호 처리에 있어, 선행 연구의 결과를 분석하여 발견된 문제점을 개선하고 최적화한 모델을 제안한다. 선행 연구 과정에서 기존 모델 구조가

갖는 한계점을 파악하여, 전신 근전도 신호의 특성을 더 효율적으로 반영할 수 있는 구조적 개선점을 찾는다.

II. 관련 연구

1. EMG를 통한 자세 복원 연구

표면 근전도 신호는 근육의 생리학적 활성도를 반영하므로, 이를 분석하여 신체의 동작을 복원하거나 의도를 파악하려는 연구는 꾸준히 진행되고 있다. 특히 **Muscles in Action (MIA)**^[2] 연구는 대규모 멀티모달 데이터셋을 구축하여 비디오 데이터와 근전도 신호를 정밀하게 동기화하고, 이를 기반으로 근육의 수축 패턴과 실제 시각적 움직임 사이의 상관관계를 학습하는 프레임워크를 제안하였다. 이 연구에서는 시각적 정보와 근육 활성화 간의 효과적인 결합을 위해, 1차원 표면 근전도 데이터를 2차원 이미지 형태로 변환하여 분석하는 기법을 도입하였다. 이 기법은 본질적으로 시간 축에 따른 전압 변화인 1차원 데이터를 $C \times T$ 크기의 2차원 텐서로 재구성하는 방식이다. 이때 C 는 시계열 데이터의 채널 수, T 는 시계열 데이터의 길이이다. 변환된 행렬에서 각 행은 개별 근육의 시간적 활성도를, 열은 특정 시점에서의 근육 간 공간적 상관관계를 나타낸다. 이를 통해 딥러닝 모델은 시각적 움직임 정보와 근육의 협응 패턴을 1차원 데이터를 위한 Conv1d 연산이 아닌, Conv2d 연산을 통해서 이미지의 텍스처와 유사하게 통합적인 학습이 가능해진다. 이를 통해, 외형적 움직임과 내부 생체 신호 사이의 상호 보완적인 정보를 정밀하게 파악할 수 있다.

또한, 국소 부위의 신호를 이용한 고자유도 자세 추정 연구로서 **EMG2Pose**^[3]가 대표적이다. 이 연구는 전완근과 같은 국소적인 부위에서 수집된 신호만으로 손가락 관절의 복잡하고 정밀한 움직임을 복원하는 딥러닝 모델을 제안하였다. 초기 연구들이 다채널 시계열 데이터를 시간×채널 형태의 2차원 이미지로 간주하여, 2D 커널을 통해 특정 시점의 근육 활성 패턴과 인접 채널 간의 상관관계를 학습하는 방식을 주로 사용했던 것과 달리, **EMG2Pose**는 **TDS (Time-Depth Separable)**^[4] 구조를 통해 복잡한 다채널 시계열 데이터로부터 효율적으로 상관관계를 학습했다. TDS

- a) 서울과학기술대학교 국방융합과학대학원 국방ICT융합공학과(Seoul National University of Science and Technology)
- b) 서울과학기술대학교 일반대학원 스마트ICT융합공학과(Seoul National University of Science and Technology)
- c) 서울과학기술대학교 전자IT미디어공학과(Seoul National University of Science and Technology)
- d) 서울과학기술대학교 스마트ICT융합공학과(Seoul National University of Science and Technology)

‡ Corresponding Author : 박구만(Gooman Park)
 E-mail: gmpark@seoultech.ac.kr
 Tel: +82-2-970-6430
 ORCID: <https://orcid.org/0000-0002-7055-5568>

* 이 논문은 2026년도 정부(과학기술정보통신부)의 재원으로 정보통신기획평가원의 지원을 받아 수행된 연구임(No.RS-2025-02216275, 사실적 움직임 생성·재현이 가능한 디지털 휴먼 기술)

* 이 논문의 연구 결과 중 일부는 한국방송·미디어공학회 2025년 추계학술대회에서 발표한 바 있음.

· Manuscript January 7, 2026; Revised January 29, 2026; Accepted January 30, 2026.

모듈은 연산 효율성을 극대화하기 위해 일반적인 합성곱을 Depth-wise Separable Convolution^[5]을 통해 두 단계로 분리한 구조이다. 먼저 Depth-wise Convolution을 통해 각 채널 내부의 시간적 변화를 정밀하게 추출하고, 이후 Point-wise Convolution을 통해 채널 간의 공간적 상관관계를 효과적으로 결합한다. 이러한 방식을 통해 파라미터 수를 최소화하면서도 수직적 관계(채널 간)와 수평적 관계(시간적)를 동시에 추출하여, 기존 합성곱 신경망 대비 연산량을 획기적으로 줄이면서 시계열 데이터의 장기 의존성을 효율적으로 학습할 수 있다.

2. 다채널 시계열 분석 연구

전신 근전도 신호는 다수의 근육에서 동시에 발생하는 다채널 시계열 데이터의 특성을 갖는다. 이러한 데이터는 복잡한 시각적·시간적 변동성을 가지기 때문에, 단순히 시간 영역을 순차적으로 분석하는 것을 넘어 데이터에 내재된 주기성과 주파수 특징을 활용하는 접근이 필요하다. 이러한 시계열 데이터의 주기성과 주파수 특징을 효과적으로 분석하기 위해 FEDformer^[6]와 같은 연구에서 주파수로 변환하여 분석하는 것이 더 효과적임을 입증해 왔다.

이러한 주파수 기반 접근을 확장한 대표적인 방법으로 TimesNet^[7]은 1차원 데이터의 다중 주기성 분석에 집중한 혁신적인 방법론을 제시하였다. TimesNet은 시간 영역의 신호를 주파수 영역으로 변환하여, 시계열 데이터 내에 존재하는 여러 주기적 패턴과 비주기적 변화를 동시에 포착하는 방법을 제안했다. 이 과정의 핵심은 FFT (Fast Fourier Transform)^[8] 모듈의 활용이다. FFT 모듈은 시간 도메인의 정보를 푸리에 변환을 통해 주파수 스펙트럼으로 변환하여 분석하는 기법으로, 신호 전체의 주기적 특성과 흐름을 한 번에 파악할 수 있다는 장점이 있다. TimesNet은 이를 통해 신호 내의 주요 주기 성분을 빠르게 파악하고, 노이즈가 많은 고주파 성분을 걸러내거나 주요 주파수 성분만을 선택적으로 학습한다. 이후 변환된 정보를 기반으로 1차원 데이터를 2차원 텐서로 재구성하고 2차원 커널을 활용해 주기 내 변화와 주기 간 변화를 동시에 학습함으로써 시계열 예측 및 분류 과제에서 성능을 비약적으로 향상시켰다.

3. Feature Embedding 기법에 따른 자세 분류 실험

선행 연구^[9]에서는 표면 근전도 신호를 이용한 전신 자세 분류의 가능성이 제시된 바 있다. 해당 연구에서는 전체 모델 구조를 Feature Embedding, Feature Extractor, Classification Head의 3단계로 모듈화하여 실험을 수행했다. Feature Embedding 단계에서 세 가지 시계열 데이터 분석 기법을 각각 적용하여 구성하였고, 각 모듈 이후에 Transformer^[10] 기반의 Feature Extractor와 결합하였을 때의 성능 차이를 비교 분석하였다.

선행 연구에서 비교 분석한 세 가지 데이터 분석 기법은 다음과 같다. 첫 번째는 다채널 시계열 데이터를 2차원 이미지로 간주하여 Conv2D를 통해 데이터를 처리하는 2D Image 변환 방식이다. 두 번째는 EMG2Pose 연구에서 검증된, 연산 효율성과 시공간 특징 추출에 특화된 TDS 모듈이다. 세 번째는 TimesNet의 핵심 아이디어로, 주파수 도메인 변환을 통해 전역적 주기성을 포착하는 FFT 모듈이다.

실험 결과, 이러한 모듈형 구조를 통해 자세 분류가 가능함이 입증되었으며, 각 모듈별 특성에 따른 성능지표가 도출되었다. 본 연구는 이러한 선행 연구의 실험 프레임워크와 결과 데이터를 기초로 하여, 선행 실험에서 확인한 모델의 구조적 특성을 재분석하여 개선 내용을 연구하였다.

III. 본 론

본 연구 이전에 선행된 연구를 통해 근전도 신호만으로 자세 분류가 가능한지를 탐색함과 동시에, 근전도 신호 처리에 최적화된 모델 구조를 탐색하고자 했다. 해당 연구의 결과 8채널의 표면 근전도 데이터만으로 자세 분류가 가능함을 보였고, TDS 모듈을 Feature Embedding 단계에 적용했을 때 타 모듈 대비 우수한 분류 정확도를 기록하였다.

그러나 이러한 성능의 이점에도 불구하고, 선행 연구의 모델 구조에서 공통적으로 파라미터가 과도하게 증가하는 한계를 드러냈다. 선행 모델들은 Feature Extractor로 Transformer Encoder를 공통으로 채택하였는데, 이는 Self Attention 메커니즘에 의존하는 모델이다. 이로 인해 앞단의 Feature Embedding에서 효율적으로 초기 특징을 추출했음에도 불구하고, 뒷단의 무거운 연산 구조로 인해 전체 모

델의 파라미터가 폭발적으로 증가하였다.

따라서 본 연구에서는 Transformer를 제거하고, 이를 대체하기 위해 시계열 데이터의 다중 주기성을 활용한 새로운 경량화 모델을 제안한다. 이 모델은 FFT를 통해 시계열 데이터의 구조를 변환하고, TDS 연산을 적용하여 다중 주기성을 확보함과 동시에 수직적 관계와 수평적 관계를 동시에 확보하도록 설계하였다.

1. 모델 경량화

본 연구에서 제안하는 모델의 핵심은 무겁고 연산량이 많은 Transformer를 FFT를 통한 차원 변환과 TDS 기반의 합성곱을 결합하여 더 효율적으로 특징을 추출할 수 있는 모델로 대체하는 것이다. 이 융합 모델은 1차원의 다채널 시계열 데이터를 다중 주기 분석을 통해 2차원 텐서로 재구성하고, 이를 경량화된 연산 과정을 통해 처리하여 수직적 특징과 수평적 특징, 그리고 주기 내 특징과 주기 간 특징을 동시에 학습한다. 이는 단순한 1차원 합성곱 연산에서는 놓치기 쉬운 장기적인 시간 의존성을 주파수 도메인을 통해 전역적으로 포착하면서도 연산량을 줄일 수 있다는 장점이 있다.

구체적인 동작은 다음과 같다. 먼저, 입력된 시계열 신호 $x \in R^{B \times C \times T}$ 에 대해 FFT를 수행하여 주파수 영역으로 변환하고, 진폭이 가장 높은 Top-k개의 주요 주파수 성분을 식별한다. 이때 x 는 입력 텐서, R 은 실수, B 는 배치 크기, C 는 채널 수, T 는 입력 시퀀스의 길이를 뜻한다. 식별된

각 주기 P 를 기준으로 1차원 데이터를 2차원 텐서 $x_{2d} \in R^{B \times C \times T/P \times P}$ 로 재배열한다. 이때 x 는 2차원으로 변환된 텐서, B 는 배치 크기, C 는 채널 수, P 는 FFT를 통해 찾아낸 주기의 길이, T/P 는 전체 길이에서 주기가 반복된 횟수를 뜻한다. 이 과정을 통해 시계열 데이터 내의 주기 내 변화와 주기 간 흐름을 2차원 공간상에 구조화하여 모델이 분석할 수 있도록 변환한다.

이렇게 변환된 2차원 텐서에는 TDS를 적용시킨다. 기존의 합성곱 연산은 공간적 특징과 채널 정보를 동시에 연산하여 파라미터 수가 방대해지는 반면, TDS 구조는 이를 분리하여 연산함으로써 더 적은 파라미터로 효율적인 연산이 가능해진다. TDS 구조에서는 채널별로 독립적인 연산을 수행하는 Depth-wise Convolution과 채널 간의 정보를 혼합하는 Point-wise Convolution을 순서대로 진행한다. Depth-wise 단계에서는 채널 간 간섭 없이 각 채널 고유의 2D 시각적 패턴을 추출하며, 이어지는 Point-wise 단계에서는 채널 간의 관계를 학습한다. 최종적으로 서로 다른 주기에 대해 추출된 특징맵들은 주파수 진폭 기반의 가중치를 통해 적응적으로 합산되어 다시 1차원 시계열 형태로 복원된다. 이러한 구조를 통해 적은 연산량으로 Transformer의 역할을 대체할 수 있도록 한다.

2. 전체 시스템 파이프라인

제안하는 방법을 통해 구축된 전체 자세 분류 모델의 아

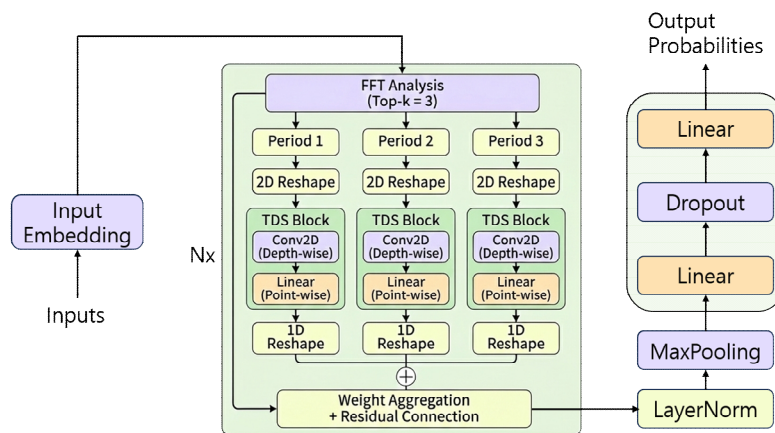


그림 1. 전체 모델 파이프라인
Fig. 1. Overall Architecture of Proposed Model

키텍처는 그림 1에서 표현한 것과 같이 3단계로 구성된다.

첫 번째 단계에서 입력된 다채널 시계열 데이터는 1×1 합성곱 층을 통과하여 모델이 처리하기 적합한 고차원 특징 공간으로 투영된다. 이는 원본 신호의 정보를 보존하면서 학습 가능한 잠재적 표현을 생성하는 과정이다.

두 번째 단계에서는 앞서 설계한 TDS-FFT 융합 레이어를 여러 번 적층하여 설계됐다. 실험을 위해 N 개의 레이어를 적층하도록 설계되었으며, 각 레이어 사이에는 깊은 구조에서도 안정적으로 학습이 되도록 잔차 연결^[11]을 적용한다. 레이어를 통과할수록 모델은 단순한 신호 변화에서 복합적인 자세 패턴으로 정보를 구체화한다.

마지막 단계에서는 자세 분류기를 통해 사전에 지정한 K 개의 분류 벡터로 변환된다. 시간 축에 대해 최댓값을 추출하는 Global Max Pooling을 적용하여 전체 동작 구간 중 가장 특징적인 정보를 대푯값으로 압축한다. 이후 선형계층을 통과하여 최종적으로 각 자세에 대한 확률을 출력한다.

IV. 실험 및 결과 고찰

1. 데이터셋

본 연구에서 제안하는 경량화 모델의 자세 분류 성능을

검증하기 위해서 Muscles in Action^[2] 데이터셋을 사용하였다. Muscles in Action 데이터셋은 시각적 움직임과 근육 활성화도 간의 관계를 분석하기 위해 구축된 대규모 멀티모달 데이터셋으로, 10명의 피험자로부터 수집된 12.5시간 분량의 데이터를 포함한다.

RGB 비디오 데이터는 iPhone 10 카메라를 통해 29.97 fps의 속도로 촬영되었으며, 표면 근전도 데이터는 ANR 사

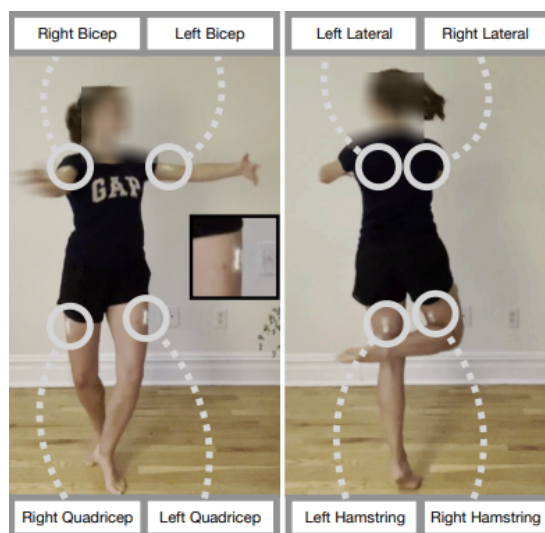


그림 3. 표면 근전도 센서 부착 위치
 Fig. 3. sEMG Sensor Placement



그림 2. Muscles in Action 데이터셋의 15가지 동작 예시
 Fig. 2. Visual Examples of 15 Actions from the Muscles in Action dataset

의 M40 Muscle Sense Bluetooth wireless EMG sensor를 통해 10fps의 속도로 촬영되었다. 수집된 동작은 그림 2에서 표현하는 것과 같이 총 15가지로, 피험자가 각 운동을 5분씩 수행하는 동안 촬영되었다. 표면 근전도 수집 위치는 그림 3에서 표현한 것과 같이 양쪽 이두근, 광배근, 대퇴사두근, 햄스트링 부위에서 수집되었다. 수집된 두 데이터의 샘플률을 맞추기 위해 비디오의 샘플률을 10fps의 속도로 재 샘플링하여 두 데이터를 서로 시간적 동기화하였다. 이후 데이터를 3초 단위로 나누어 30프레임의 시퀀스로 만들어 저장하였다.

2. 실험 환경 및 하이퍼 파라미터 분석

본 연구의 모든 실험을 위한 딥러닝 프레임워크는 pytorch 2.1.2 버전을 사용했으며 CUDA 버전은 12.9 버전을 사용하였다. 실험을 위한 연산장치로 CPU는 Intel사의 Xeon gold 6426Y 중 4개 코어를 사용하였고, GPU는 NVIDIA RTX A6000을 사용하였다. 데이터 처리 및 수치 연산에는 NumPy와 SciPy 라이브러리를 사용하였고, 그밖에 라이브러리는 Docker container를 통해 관리하였다.

모델의 학습을 위한 손실함수는 15개 동작에 대한 다중 분류 문제이므로 Cross Entropy Loss를 사용했다. Cross Entropy Loss는 모델이 예측한 확률분포와 실제 레이블의 분포 차이를 측정하는 지표로, 이 값을 최소화하는 방향으로 모델의 파라미터를 업데이트하였다.

최적화 알고리즘으로는 Adam 옵티마이저를 사용하였고, 별도의 스케줄러는 사용하지 않았다. 대신 모델의 과적합을 방지하고 최적의 일반화 성능을 도출하기 위해, 매 에포크가 종료될 때 검증 데이터셋에 대한 손실값을 기록하였으며, 학습이 완료된 후에는 마지막 에포크의 모델이 아니라, 검증 손실이 가장 낮았던 시점의 모델 가중치를 저장하여 테스트 데이터셋에 대한 성능 평가를 수행하였다.

본 연구에서 제안하는 경량화 모델의 학습에 사용한 파라미터는 선행 연구와 원활한 비교를 위해 기본적인 파라미터는 표 1에서 설명하는 바와 같이, 선행 연구와 동일한 파라미터를 사용하였다. 선행 연구에서는 여러 번의 실험을 통해 동일한 에포크에서 가장 좋은 성능을 보인 파라미터를 선정하였고, 본 연구의 초기 학습률과 학습 에포크 수

는 선행 연구와 유사하게 각각 0.00005, 150 에포크로 설정했다.

제안하는 경량화 모델의 최적 깊이를 탐색하기 위해 나머지 하이퍼 파라미터를 고정한 채 융합 모듈을 3, 5, 7, 9번 적층하여 학습시키며 각각 성능을 비교하였다. 이후 적층 횟수별 실험 결과에서 가장 뛰어난 성능을 보인 모델과 선행 연구의 결과를 서로 비교하였다.

표 1. 각 실험에서 설정한 하이퍼 파라미터
Table 1. Hyper parameter Settings

| | Learning Rate | Epochs | Optimizer | Scheduler |
|--------|---------------|--------|-----------|-------------|
| Conv2d | 0.00005 | 150 | Adam | None |
| TDS | 0.005 | 150 | Adam | MultistepLR |
| FFT | 0.001 | 150 | RAdam | MultistepLR |
| custom | 0.00005 | 150 | Adam | None |

3. 실험 결과 및 성능 비교

본 연구에서 제안하는 융합 모델의 성능과 선행 연구에서의 성능 비교는 Top_1 Accuracy, Top_3 Accuracy, Test Loss 그리고 파라미터 개수를 비교한다. 성능지표는 Muscles in Action의 Test 데이터셋에서의 결과를 통해 성능지표를 계산한다.

Top_1 Accuracy는 모델이 예측한 가장 높은 확률의 클래스가 실제 정답과 일치하는 비율을 의미한다. 분류 성능을 나타내는 가장 직관적인 지표로, 수치가 높을수록 모델이 정확한 예측을 했음을 나타낸다.

Top_3 Accuracy는 모델이 예측한 상위 3개 확률 클래스 내에 실제 정답이 포함될 비율이다. 각 클래스끼리의 패턴이 유사하여 클래스 간 모호함이 존재하는 데이터에서 모델의 잠재적인 분류 능력을 평가하는 데 유용한 수치이다. 수치가 높을수록 모델이 정답과 유사하게 추정함을 나타낸다.

Test Loss는 학습에 사용되지 않은 테스트 데이터셋에 대한 Cross Entropy 손실값을 나타낸다. 이는 모델이 예측한 확률분포와 실제 정답 분포 사이의 차이를 나타내며, 수치가 낮을수록 모델이 정답을 분명하게 예측함을 나타낸다.

파라미터 수는 모델이 학습 가능한 가중치의 총 개수를

나타낸다. 모델의 학습비용 및 연산량과 직결되는 수치로, 수치가 낮을수록 더 가볍고 효율적임을 나타낸다.

해당 지표들을 바탕으로 융합 모듈을 적층시킨 횟수에 따른 결과를 비교한 뒤, 가장 좋은 결과 값과 기존 연구에서 제안한 Transformer 기반 모델과 본 연구의 경량화 모델을 비교한다.

표 2에서 정리한 모듈의 적층 횟수에 따른 성능을 살펴보면 모듈을 깊게 적층시킬수록 전반적인 성능이 향상됐지만, 9번의 적층부터 감소하기 시작했다. 이에 따라 최적의 적층 횟수는 7번으로 간주하여 선행 연구 결과와 비교한다.

표 3에서 설명하는 선행 연구의 실험 결과와 이번 실험 결과를 비교해 보면, 모델의 복잡함을 나타내는 파라미터 수에서 기존 연구의 모델은 약 530만~1,420만 개의 파라미터를 갖지만, 제안하는 모델은 약 47만 개로 측정되었다. 이는 제안 모델이 기존 대비 약 91% 이상의 파라미터를 절감했음을 보여주며, TDS와 FFT를 융합한 모델이 성공적으로 경량화되었음을 알 수 있다.

분류 성능 측면에서도 Top_1 Accuracy는 경량화 모델이 95.947%를 기록하며, 파라미터가 대폭 감소했음에도 불구하고 기존 연구의 최댓값인 95.312%와 비교해 간소한 차이로 더 높은 수치를 기록하였다. Top_3 Accuracy 역시 경량화 모델이 98.877%를 기록하면서 기존 연구의 최댓값인

98.584%보다 더 높은 수치를 기록했다. 이는 경량화 모델이 극단적으로 파라미터를 줄이면서도 분류 성능을 충분히 유지함을 시사한다.

또한, 모델의 학습 안정성과 일반화 능력을 나타내는 Test Loss를 비교한 결과, 기존 연구의 최솟값에 비해 경량화 모델이 약 0.0005가량 높게 측정되었다. 이는 전체 규모 고려할 때 매우 미미한 수치이다. 일반적으로 모델의 파라미터가 감소하게 되면 손실값이 증가하는 경향이 있으나, 파라미터를 90% 이상 경량화했음에도 불구하고 기존 모델과 동등한 수준으로 수렴했음을 시사한다. 이는 TDS와 FFT를 융합한 모델이 안정적으로 학습되었음을 알 수 있다.

V. 결론

본 논문에서는 표면 근전도 신호를 기반으로 자세를 분류하면서도, 선행 연구에서 파라미터가 과도하게 늘어나는 문제를 해결하기 위해 TDS와 FFT를 융합한 경량화 모델을 제안한다. 제안 모델은 시간 영역의 특징을 포착하는 TDS 모듈과 주파수 영역의 주기성을 분석하는 FFT 모듈을 결합하여, 기존의 Transformer 기반의 모델을 효과적으로 대체

표 2. 모듈 적층 횟수에 따른 실험 결과 비교

Table 2. Comparison of Results According to Module Stacking Count

| | Top_1 Accuracy(↑) | Top_3 Accuracy(↑) | Test Loss(↓) | Parameter(↓) |
|---------|-------------------|-------------------|--------------|--------------|
| 3_layer | 93.276% | 98.681% | 0.026871 | 212,589 |
| 5_layer | 94.824% | 98.437% | 0.021999 | 344,637 |
| 7_layer | 95.947% | 98.877% | 0.018933 | 476,685 |
| 9_layer | 95.849% | 98.877% | 0.016973 | 608,733 |

표 3. 선행 연구의 실험 결과와 이번 실험 결과 비교

Table 3. Comparative Analysis Between Previous Results

| | Top_1 Accuracy(↑) | Top_3 Accuracy(↑) | Test Loss(↓) | Parameter(↓) |
|--------------|-------------------|-------------------|--------------|--------------|
| Conv2d | 90.087% | 97.607% | 0.038642 | 5,349,612 |
| TDS | 95.312% | 98.584% | 0.018378 | 7,960,397 |
| FFT | 94.464% | 98.788% | 0.282330 | 14,216,271 |
| Custom Model | 95.947% | 98.877% | 0.018933 | 476,685 |

하였다. 제안 모델의 구조적 효율성을 극대화하기 위해, 융합 블록의 적층 횟수에 따른 성능 변화를 실험적으로 검증하였다. 실험 결과 적층 횟수가 많아질수록 파라미터 수와 분류 성능이 선형적으로 증가하였고 7번 적층했을 때 가장 우수한 분류 성능을 보임을 확인했다. 반면에 적층 횟수를 9번 증가시켰을 때는 오히려 분류 성능이 소폭 하락하였다. 이를 통해 7번의 적층 횟수를 가지는 모델을 최종 모델로 확정하고 선행 연구 결과와 비교하였다.

확정된 모델을 통해 **Muscles in Action** 데이터셋에서 선행 연구 결과와 성능을 비교한 결과, 제안 모델은 선행 연구 대비 파라미터 수를 최대 96% 이상 절감한 47만 개로 줄이는 데 성공하였다. 이러한 경량화에도 불구하고 95.947%의 **Top_1 Accuracy**를 달성하여 선행 연구 결과인 95.31%에서 소폭 향상되는 결과를 얻었다. 비록 **Test Loss**에서 미세한 증가가 있었으나, 파라미터의 감소 폭을 고려했을 때 이는 무시할 수 있는 수준의 차이임을 입증하였다.

결론적으로 본 연구에서 제안한 모델은 **Transformer** 기반의 모델이 가지는 문제점을 해결하기 위해 **TDS**와 **FFT**를 융합하여 경량화와 성능 개선을 동시에 달성하였다. 이는 향후 웨어러블 디바이스나 디지털 휴먼과 같은 자세 복원이 필요한 분야에서 효율적으로 쓰일 것이다.

참 고 문 헌 (References)

- [1] C. J. De Luca, "The Use of Surface Electromyography in Biomechanics," *Journal of Applied Biomechanics*, Vol.13, No.2, pp.135-163, May 1997.
doi: <https://doi.org/10.1123/jab.13.2.135>
- [2] Chiquier, M., & Vondrick, C., "Muscles in Action," *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 22093-22103, October, 2023.
doi: <https://doi.org/10.1109/iccv51070.2023.02019>
- [3] Salter, S., Warren, R., Schlager, C., Spurr, A., et al., "emg2pose: A Large and Diverse Benchmark for Surface Electromyographic Hand Pose Estimation," *Advances in Neural Information Processing Systems (NeurIPS)*, Vol. 37, December, 2024.
doi: <https://doi.org/10.52202/079017-1770>
- [4] A. Hannun, A. Lee, Q. Xu, and R. Collobert. "Sequence-to-sequence speech recognition with time-depth separable convolutions." *Interspeech*, pp. 3785-3789, April, 2019.
doi: <https://doi.org/10.21437/Interspeech.2019-2460>
- [5] A. G. Howard, M. Zhu, B. Chen, D. Kalenichenko, W. Wang, T. Weyand, M. Andreetto, and H. Adam, "MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications." *arXiv preprint arXiv:1704.04861*, April, 2017.
doi: <https://doi.org/10.48550/arXiv.1704.04861>
- [6] Q. Zhou, J. Chen, J. Wang, H. Luo, T. Zhang, and L. Wen, "FEDformer: Frequency Enhanced Decomposed Transformer for Ling-Term Series Forecasting" *International conference on machine learning (ICML)*. pp. 27268-27286, June, 2022.
doi: <https://doi.org/10.1109/access.2023.3287893>
- [7] Wu, H., Hu, T., Liu, Y., Zhou, H., Wang, J., & Long, M., "TimesNet: Temporal 2D-Variation Modeling for General Time Series Analysis," *International Conference on Learning Representations (ICLR)*, May, 2023.
doi: <https://doi.org/10.48550/arXiv.2210.02186>
- [8] Cooley, J. W., and Tukey, J. W., "An algorithm for the machine calculation of complex Fourier series," *Mathematics of Computation*, Vol. 19, No. 90, pp. 297-301, April, 1965.
doi: <https://doi.org/10.1090/S0025-5718-1965-0178586-1>
- [9] K. Cho, H. Kang, G. Park, et al. "Human Motion Classification using surface EMG signal." *The Korean Society of Broadcast and Media Engineers*, Seoul, Korea. pp. 371-374, 2025.
- [10] A. Vaswani et al., "Attention Is All You Need," *Advances in Neural Information Processing Systems (NeurIPS)*, Long Beach, CA, USA, pp. 5998-6008, 2017.
doi: <https://doi.org/DOI: 10.1201/9781003561460-19>
- [11] K. He et al., "Deep Residual Learning for Image Recognition," *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Las Vegas, NV, USA, pp. 770-778, 2016.
doi: <https://doi.org/10.1109/CVPR.2016.90>

저 자 소 개



조 경 민

- 2019년 3월 ~ 2025년 2월 : 경상국립대학교 메카트로닉스공학부 메카트로닉스공학 학사
- 2025년 3월 ~ 현재 : 서울과학기술대학교 국방융합과학대학원 국방ICT융합공학과 석사과정
- ORCID : <https://orcid.org/0009-0006-5977-7190>
- 주관심분야 : 딥러닝



강 현 석

- 2015년 3월 ~ 2021년 8월 : 강원대학교 삼척캠퍼스 전자공학과 학사
- 2021년 9월 ~ 2023년 8월 : 서울과학기술대학교 일반대학원 스마트ICT융합공학과 석사
- 2023년 9월 ~ 현재 : 서울과학기술대학교 일반대학원 스마트ICT융합공학과 박사과정
- ORCID : <https://orcid.org/0000-0003-0783-3841>
- 주관심분야 : 3D 컴퓨터 비전, 딥러닝



최 형 석

- 2017년 3월 ~ 2024년 8월 : 광운대학교 컴퓨터정보공학과 학사
- 2024년 9월 ~ 현재 : 서울과학기술대학교 일반대학원 스마트ICT융합공학과 석사과정
- ORCID : <https://orcid.org/0009-0006-3665-2512>
- 주관심분야 : 3D 컴퓨터 비전, 딥러닝



최 우 석

- 2021년 3월 ~ 현재 : 서울과학기술대학교 전자T미디어공학과 학사
- ORCID : <https://orcid.org/0009-0000-8764-8130>
- 주관심분야 : 로봇 비전, 강화학습



박 구 만

- 1984년 : 한국항공대학교 전자공학과 공학사
- 1986년 : 연세대학교 대학원 전자공학과 석사
- 1991년 : 연세대학교 대학원 전자공학과 박사
- 1984년 ~ 1996년 : 삼성전자 신호처리연구소 선임연구원
- 1991년 ~ 현재 : 서울과학기술대학교 스마트ICT융합공학과 교수
- 2006년 ~ 2007년 : Georgia Institute of Technology, Dept. of ECE. Visiting Scholar
- 2016년 ~ 2017년 : 서울과학기술대학교 나노IT디자인융합대학원 원장
- ORCID : <https://orcid.org/0000-0002-7055-5568>
- 주관심분야 : 컴퓨터 비전, 실감미디어