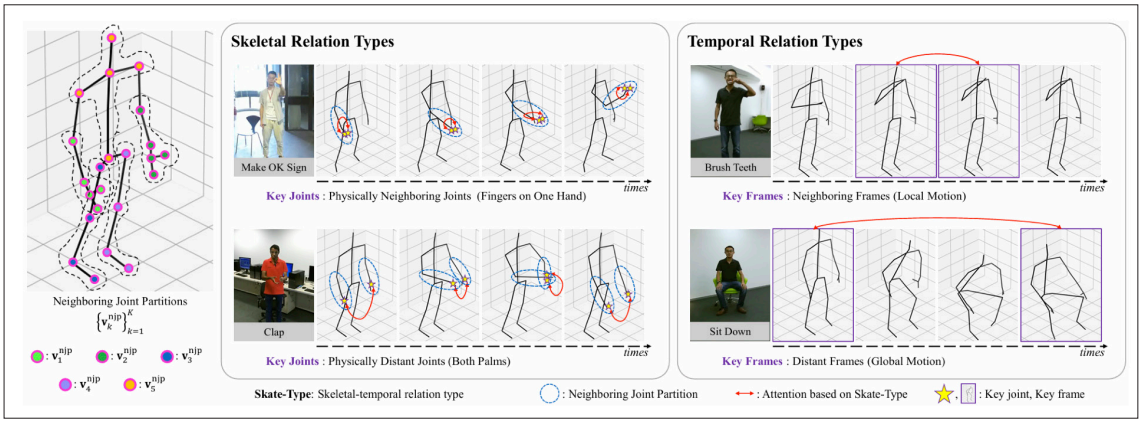
졸업논문
소개

골격 기반 행동 인식을 위한 효율적인 트랜스포머 모델 연구

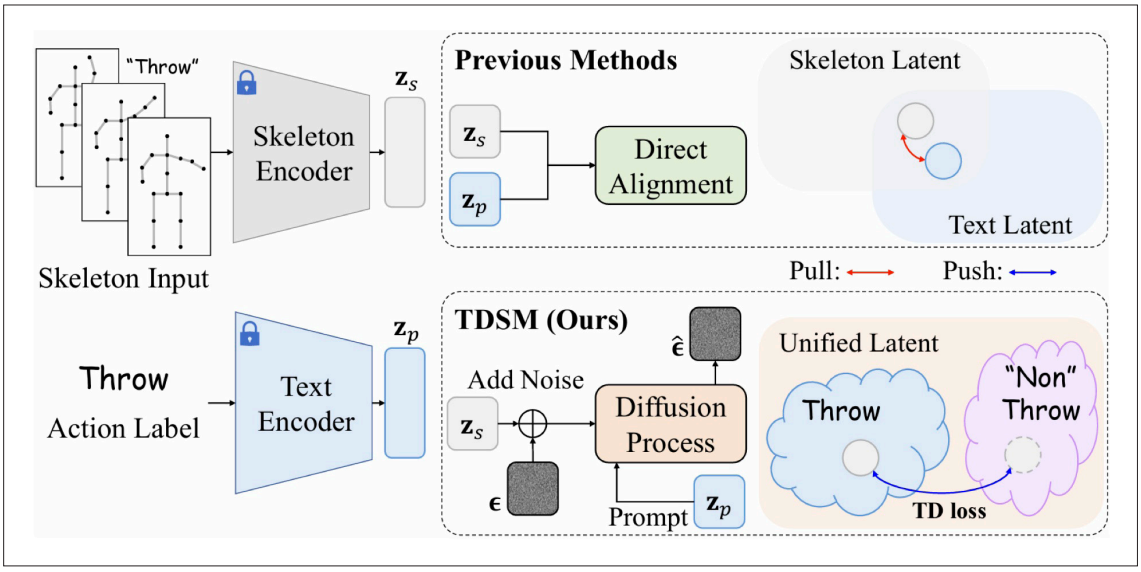
도정혁 / 한국과학기술원 Video and Image Computing Lab.

최근 인공지능(AI) 기술의 발전은 인간 행동을 이해하고 인식하는 다양한 응용 분야로 확장되고 있다. 특히, 사람의 움직임을 추적하는 '골격 기반 행동 인식'은 건강 모니터링, 스포츠 분석, 인간-로봇 상호작용 등에서 중요한 기술로 주목받고 있다. 본 연구는 이 골격 기반 행동 인식을 보다 정확하고 효율적으로 수행할 수 있는 새로운 딥러닝 모델을 제안하고, 그 성능을 다양한 데이터셋에서 검증한 결과를 담고 있다.

기존에는 사람의 움직임을 RGB 영상이나 깊이 정보 등을 이용해 인식했지만, 이러한 방식은 조명, 배경, 카메라 각도 등에 민감해 성능 저하가 발생할 수 있다. 반면, 사람의 주요 관절 좌표만으로 구성된 골격 데이터는 배경이나 외형 정보 없이 순수한 움직임 정보만을 담고 있어, 보다 경량화되고 개인정보 보호 측면에서도 강점을 가진다.



<그림 1>



<그림 2>

하지만 골격 데이터를 효과적으로 처리하기 위해 기존에는 그래프 신경망(Graph Convolutional Network, GCN)이 주로 사용되었다. GCN은 관절 간 연결 정보를 기반으로 행동을 분류하지만, 서로 멀리 떨어진 관절 간의 상관관계를 파악하는 데 한계가 있다. 이를 해결하고자 본 연구에서는 트랜스포머 기반의 새로운 모델, **SkateFormer** (Skeletal-Temporal Transformer)를 제안한다.

SkateFormer는 골격 데이터를 네 가지 관계 유형(가까운 관절, 먼 관절, 짧은 시간 간격, 긴 시간 간격)으로 분할하여 각각에 맞는 맞춤형 어텐션(attention)을 수행한다. 예를 들어, '손뼉치기' 동작은 양 손의 먼 관절 사이의 관계가 중요하고, '양치질'은 짧은 시간 간격에서의 손 움직임이 핵심이다. SkateFormer는 이러한 동작별 특징을 잘 포착할 수 있도록 설계되어 효율성과 정확성을 모두 갖추었다.

더 나아가, 본 연구는 보지 못한 행동도 인식할 수 있는 제로샷 행동 인식(Zero-Shot Action Recognition) 문제도 함께 다루었다. 제로샷 인식이란, 학습 중 보지 못한 행동을 텍스트 설명만으로 인식하는 기술로, 실제 응용에서는 매우 유용하다. 이를 위해, 본 논문은 **TDSM** (Triplet Diffusion for Skeleton-Text Matching)이라는 새로운 방법을 제안했다. 이는 최근 주목받는 확산 모델을 활용하여 골격과 텍스트 정보를 효과적으로 연결하는 방식이다.

이 두 가지 모델(SkateFormer와 TDSM)은 각각 감독 학습 기반 행동 인식과 제로샷 행동 인식 분야에서 새로운 성능 기록을 세웠으며, NTU RGB+D, Kinetics Skeleton 등 다양한 벤치마크 데이터셋에서 기존 최신 방법들을 능가하는 결과를 보여주었다.

이번 연구는 인간의 행동을 보다 효율적이고 정확하게 인식함으로써, 실시간 감시 시스템, 개인 건강 관리, 재활 치료, 스마트 홈 등 다양한 분야에 적용될 수 있는 기술적 가능성을 제시한다. 골격 기반 행동 인식 기술의 미래를 앞당길 수 있는 의미 있는 시도라고 할 수 있다.

졸업논문
소개

도 정 혁

- 2019년 2월 : 한국과학기술원 전기및전자공학부 학사
- 2021년 2월 : 한국과학기술원 전기및전자공학부 석사
- 2025년 8월 : 한국과학기술원 전기및전자공학부 박사
- 주관심분야 : 골격 기반 행동인식, 위성영상처리, 확산모델